GAZİ ÜNİVERSİTESİ

Gazi

1926

TOPLAM PUAN VE ALT PUANLARIN BİRLİKTE KESTİRİLDİĞİ MODELLERİN KESTİRİM DOĞRULUĞU, GÜVENİRLİK VE SINIFLAMA DOĞRULUĞUNA GÖRE KARŞILAŞTIRILMASI

Ayşenur Erdemir

DOKTORA TEZİ

EĞİTİM BİLİMLERİ ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Temmuz, 2021

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren tezden fotokopi çekilebilir.

YAZARIN

Adı : Ayşenur

Soyadı : Erdemir

Bölümü : Eğitim Bilimleri

İmza :

Teslim Tarihi :

TEZİN

Türkçe Adı : Toplam Puan ve Alt Puanların Birlikte Kestirildiği Modellerin Kestirim Doğruluğu, Güvenirlik ve Sınıflama Doğruluğuna Göre Karşılaştırılması

İngilizce Adı : Comparison of Models Estimating Overall Score and Subscore Simultaneously in Terms of Precision, Reliability and Classification Accuracy

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazar Adı Soyadı : Ayşenur Erdemir

İmza :

JÜRİ ONAY SAYFASI

Ayşenur Erdemir tarafından hazırlanan "Toplam Puan ve Alt Puanların Birlikte Kestirildiği Modellerin Kestirim Doğruluğu, Güvenirlik ve Sınıflama Doğruluğuna Göre Karşılaştırılması" adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Anabilim Dalı'nda Doktora tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Hakan Yavuz Atar _____

Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi

Başkan: Prof. Dr. Şeref Tan _____

Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi

Üye: Prof. Dr. Selahattin Gelbal _____

Eğitimde Ölçme ve Değerlendirme, Hacettepe Üniversitesi

Üye: Prof. Dr. İsmail Karakaya _____

Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi

Üye: Doç. Dr. Celal Deha Doğan _____

Eğitimde Ölçme ve Değerlendirme, Ankara Üniversitesi

Tez Savunma Tarihi: 22/06/2021

Bu tezin Eğitim Bilimleri Anabilim Dalı'nda Doktora tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Yücel Gelişli

Eğitim Bilimleri Enstitüsü Müdürü ……

Şehit Öğretmen
Şenay Aybüke Yalçın'a

TEŞEKKÜR

Doktora eğitimim boyunca birlikte çalışmaktan onur duyduğum, her alanda çalışmalarımı destekleyen, her konuda kendisinden çok şey öğrendiğim ve örnek aldığım değerli hocam ve danışmanım Prof. Dr. Hakan Yavuz Atar'a saygı ve minnetle şükranlarımı sunarım.

Tez izleme komitemde ve jürimde yer alan, süreç boyunca akademik gelişimimi destekleyen ve tezimin daha iyi olması için değerli görüşlerini paylaşan kıymetli hocalarım Prof. Dr. Selahattin Gelbal ve Prof. Dr. Şeref Tan'a teşekkürlerimi sunarım. Tez savunma jürimde yer alarak görüş ve önerileri ile önemli katkılar sunan değerli hocalarım Prof. Dr. İsmail Karakaya ve Doç. Dr. Deha Doğan'a çok teşekkür ederim.

Lisansüstü eğitimim boyunca birlikte çalıştığım, birlikte öğrendiğim, birlikte dinlendiğim, eğlendiğim ve desteklerini her zaman hissettiğim kıymetli arkadaşlarımın varlıkları için minnettarım; sevgili Arş. Gör. Dr. Derya Akbaş, Arş. Gör. Esra Oyar, Arş. Gör. Ergün Cihat Çorbacı, Dr. Öğr. Üyesi Fuat Elkonca, Dr. Öğr. Üyesi Görkem Ceyhan, Dr. Öğr. Üyesi Mahmut Sami Koyuncu, Dr. Öğr. Üyesi Mehmet Şata, Arş. Gör. Oya Erdinç-Akan, Arş. Gör. Dr. Serpil Çelikten-Demirel, Dr. Sibel Ada, Arş. Gör. Dr. Zafer Ertürk'e her şey için ve ayrı ayrı teşekkürlerimi sunarım.

Esra, Görkem, Mahmut Sami ve Serpil, tez çalışmam sırasında tezimin daha nitelikli ve anlaşılır hale gelmesinde önemli katkılar sağlayan dönütleriniz ve emekleriniz için ayrıca, özellikle ve defaatle teşekkürlerimi iletmek istiyorum.

Birlikte çalıştığım için çok mutlu olduğum Arş. Gör. Dr. Özge Ceren Çelik, Arş. Gör. Sabire Kılıç ve Arş. Gör. Dr. Ümre Kaynak'a dostlukları için teşekkür ederim.

Bu araştırmada yer alan analizlerin büyük bölümü TÜBİTAK ULAKBİM, Yüksek Başarım ve Grid Hesaplama Merkezi'nde (TRUBA kaynaklarında) gerçekleştirilmiştir. Desteklerinden ötürü TRUBA ekibine ve doktora eğitimim boyunca bana yurt içi doktora burs imkânı sağlayan TÜBİTAK'a teşekkür ederim.

TRUBA kaynaklarının kullanımı konusunda desteğini esirgemeyen TÜBİTAK ULAKBİM mühendislerinden kuzenim Tuncay Çolak'a kodların Linux sistemine uyumlu hale getirilmesindeki emeği ve yardımları için teşekkür ederim.

Her zaman ve her alanda desteklerini hissettiğim güç kaynağım ailem, geniş ölçekli kuzenlerim ve tüm sevdiklerime hayatıma kattıkları değer, mutluluk ve bütün güzellikler için en kalbi teşekkürlerimi sunarım.

Ayşenur

TOPLAM PUAN VE ALT PUANLARIN BİRLİKTE KESTİRİLDİĞİ MODELLERİN KESTİRİM DOĞRULUĞU, GÜVENİRLİK VE SINIFLAMA DOĞRULUĞUNA GÖRE KARŞILAŞTIRILMASI

(Doktora Tezi)

Ayşenur Erdemir

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Temmuz, 2021

ÖZ

Bu araştırmanın ana amacı, toplam test ve alt testlere ilişkin yetenek kestirim modellerinin karşılaştırmalı incelenmesidir. Çalışma kapsamında ele alınan ÇBMTK, ÜDMTK ve Bifaktör modelleri, kestirim doğruluğu, güvenirlik ve sınıflama doğruluğuna dayalı olarak karşılaştırılmıştır. Çalışmada simülasyon veri ve Türkiye'de bir devlet üniversitesi Yabancı Diller Yüksek Okulu tarafından geliştirilip uygulanan İngilizce yeterlilik sınavına ait gerçek veri kullanılmıştır. Simülasyon çalışmasında örneklem büyüklüğü 5000, madde sayısı 30 ve boyut sayısı dört olarak belirlenmiştir. Manipüle edilen değişkenler toplam testte yer alan çoklu puanlanan madde yüzdesi (%5, %10, %25, %50), test güçlüğü (çok zor, zor, orta, kolay, çok kolay) ve boyutlar arası korelasyondur (0,2; 0,5; 0,8). 60 çapraz koşulun (3 korelasyon x 4 testteki çoklu puanlanan madde yüzdesi x 5 test güçlüğü) her biri için 100 replikasyon yapılmış ve toplamda 6000 veri üretilmiştir. Madde yanıt örüntüleri SimuMIRT programı ile elde edilmiştir. ÇBMTK, ÜDMTK ve Bifaktör modelleri için yetenek kestirimleri BMIRT programı kullanılarak gerçekleştirilmiştir. Kestirim modellerinin ve simülasyon koşullarının, yetenek kestirimlerine göre elde edilen RMSE, güvenirlik ve sınıflama doğruluğu değerlerine olan etkisi faktöriyel karma ANOVA ile test edilmiştir. Son olarak gerçek veri her üç kestirim modeli ile analiz edilmiş, yetenek kestiriminin standart hatası, marjinal güvenirliği ve geçti-kaldı kararına göre sınıflama doğruluğu değerleri

incelenmiştir. Genel olarak, simülasyon çalışması sonuçları hem toplam puan hem de alt puan açısından ÇBMTK modelinin bütün koşullarda ÜDMTK ve Bifaktör modelden daha iyi performans sergilediğini göstermektedir. Korelasyon yüksek olduğunda modellere göre toplam puan için elde edilen güvenirlikler arasındaki fark düşüktür. Alt puanlarda ÇBMTK ve ÜDMTK birbirine yakın sonuçlara sahiptir. Toplam puanda her üç model için de korelasyon arttıkça model performansı iyileşmiştir. Alt puanlarda ÇBMTK ve ÜDMTK modelleri için korelasyon arttıkça model performansı iyileşirken Bifaktör model performansı kötüleşmiştir. Çoklu puanlanan madde yüzdesi artışının da olumlu bir etkiye sahip olduğu görülmüştür. Test güçlüğü açısından ise test orta güçlükte olduğunda modellerin daha iyi performans gösterdiği, test güçlüğü çok zor olduğunda en yüksek hata, en düşük güvenirlik ve sınıflama doğruluğu değerlerinin elde edildiği sonucuna ulaşılmıştır. En yüksek sınıflama doğruluğu değerleri test kolay ya da çok kolay olduğunda elde edilmiştir. Değişkenlerin düzeylerine göre sonuçlarda bazı farklılıklar görülmüş olup hepsi detaylı olarak raporlanmıştır. Gerçek veri analizi ile elde edilen bulgular da simülasyon çalışmasını destekler niteliktedir.

Anahtar Kelimeler : Alt puan, toplam puan, Çok Boyutlu Madde Tepki Kuramı, Üst Düzey Madde Tepki Kuramı, Bifaktör model

Sayfa Adedi : xxiii + 264

Danışman : Prof. Dr. Hakan Yavuz Atar

COMPARISON OF MODELS ESTIMATING OVERALL SCORE AND SUBSCORE SIMULTANEOUSLY IN TERMS OF PRECISION, RELIABILITY AND CLASSIFICATION ACCURACY

(Ph.D Thesis)

Ayşenur Erdemir

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

July, 2021

ABSTRACT

This research aims to evaluate the ability estimation models for the overall score and subscores simultaneously. MIRT, HOIRT, and Bifactor models discussed within the scope of the study were compared based on precision, reliability, and classification accuracy. In the study, both simulation data and real data of an English proficiency exam developed and administered by the School of Foreign Languages of a state university in Turkey were used. In the simulation study, the sample size was determined as 5000, the number of items as 30, and the number of dimensions as four. The manipulated variables are the percentage of polytomously-scored items in the total test (5%, 10%, 25%, 50%), test difficulty (very difficult, difficult, medium, easy, very easy), and correlation between dimensions (0.2, 0.5, 0.8). The number of replication was 100 for each of the 60 cross-conditions (3 correlations x 4 levels of percentage of polytomously-scored items x 5 levels of test difficulties), and 6000 data were generated. Overall scores and subscores for MIRT, HOIRT, and Bifactor models were estimated using the BMIRT program. A factorial mixed ANOVA was performed to test the effects of estimation models and simulation conditions on RMSE, reliability, and classification accuracy values of ability estimations. Lastly, the real data were analyzed with all three estimation models, and the standard error, marginal reliability, and classification accuracy values of the ability estimation were examined. In general, the

simulation study results show that the MIRT model outperforms the HOIRT and Bifactor models in all conditions, both in terms of overall score and subscores. When the correlation is high, the difference between the reliability obtained from the estimation models for the overall score is low. For subscores, MIRT and HOIRT have similar results. For overall scores, as the correlation increased, the model performance improved for all three models. For subscores, as the correlation increased, the model performance improved for the MIRT and HOIRT models, while the Bifactor model performance declined. In terms of test difficulty, it was concluded that the models performed better when the test was of medium difficulty, and the highest error, lowest reliability and classification accuracy values were obtained when the test difficulty was very difficult. The highest classification accuracy values were obtained when the test was easy or very easy. There were some differences in the results depending on the levels of the variables, and all of them were reported in detail. Findings obtained with real data analysis also support the simulation study.

Key Words : Subscore, overall score, MIRT, HOIRT, Bifactor model

Page Number : xxiii + 264

Supervisor : Prof. Dr. Hakan Yavuz Atar

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU i

ETİK İLKELERE UYGUNLUK BEYANI ii

JÜRİ ONAY SAYFASI iii

TEŞEKKÜR v

ÖZ vii

ABSTRACT ix

İÇİNDEKİLER xi

TABLOLAR LİSTESİ xvi

ŞEKİLLER LİSTESİ xx

BÖLÜM 1 1

GİRİŞ 1

1.1. Problem Durumu 1

1.2. Araştırmanın Amacı 7

1.3. Araştırmanın Önemi 7

1.4. Sınırlılıklar 9

BÖLÜM 2 11

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR 11

2.1. Çok Boyutlu Örtük Özellik 11

2.2. Çok Boyutlu Madde Tepki Kuramı 14

2.2.1. Veri Setinin Yapısına Göre: Basit Yapı – Karmaşık Yapı (Between- and within-item multidimensionality) 16

2.2.2. Boyutlar Arasındaki Telafisel (Compensatory) ve Telafisel Olmayan (Non-compensatory) İlişki 17

2.3. Çok Boyutlu MTK Modelleri ve Kestirimi 18

2.3.1. Çok Boyutlu MTK İçin Toplam Puanın Hesaplanması – Maksimum Bilgi Yöntemi 20

2.4. Üst Düzey (Higher-Order) Madde Tepki Kuramı 22

2.5. Bifaktör Model 24

2.6. Elde Edilen Puanlara Ait Sınıflama Performansı 26

2.7. İlgili Araştırmalar 29

BÖLÜM 3 35

YÖNTEM 35

3.1. Araştırmanın Türü 35

3.2. Simülasyon Çalışması 35

3.2.1. Simülasyon Koşulları 36

3.2.2. Verilerin Üretilmesi 39

3.2.3. Madde ve Yetenek Parametrelerinin Kestirilmesi – Model Kestirimleri ... 41

3.2.4. Simülasyon Çalışması İçin Değerlendirme Kriterleri 43

3.2.5. Veri Analizi 44

3.2.6. Simülasyon Aşamaları – Özet 45

3.3. Gerçek Veri Analizi 46

3.3.1. Verinin Tanıtılması 46

3.3.2. Boyutluluk Analizi 48

3.3.3. Model Kestirimleri 49

3.3.4. Gerçek Veri Analizi İçin Değerlendirme Kriterleri 49

BÖLÜM 4 51

BULGULAR 51

4.1 Simülasyon Çalışması Toplam Puan için Bulgular 52

4.1.1. Kestirim Modelleri ve Test Güçlüğü Değişkeni İçin İncelemeler.....53

4.1.2. Çoklu Puanlanan Madde Yüzdesi Değişkeni için İncelemeler 74

4.1.3. Boyutlar Arası Korelasyon Değişkeni için İncelemeler.....92

4.2. Simülasyon Çalışması Alt Puanlar için Bulgular.....97

4.2.1. Kestirim Modelleri ve Test Güçlüğü Değişkeni için İncelemeler 99

4.2.2. Çoklu Puanlanan Madde Yüzdesi Değişkeni için İncelemeler 125

4.2.3. Boyutlar Arası Korelasyon Değişkeni için İncelemeler.....138

4.2.4. Alt Test Sınıflama Doğruluğu Kriteri için İncelemeler.....143

4.3. Gerçek Veri Analizi için Bulgular.....149

4.3.1. Ön Analizler 149

4.3.2. Madde Parametreleri.....151

4.3.3. Yetenek Parametreleri, Kestirim Hatası, Güvenirlik ve Sınıflama Doğruluğu 153

BÖLÜM 5 159

TARTIŞMA, SONUÇ VE ÖNERİLER 159

6.1. Sonuç ve Tartışma.....159

6.2. Sonuçlar 169

6.3. Öneriler.....171

KAYNAKLAR.....173

EKLER.....189

Ek 1. SimuMIRT Örnek Betik Dosyası.....190

Ek 2. BMIRT Örnek Betik Dosyası **192**

Ek 3. Simulasyon Analizi İçin Betik Dosyası **193**

Ek 4. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (*r*=0,2) **197**

Ek 5. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (*r*=0,5) **200**

Ek 6. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (*r*=0,8) **203**

Ek 7. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,2; ÇPMY basit ana etkisi) **206**

Ek 8. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,5; ÇPMY basit ana etkisi) **209**

Ek 9. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,8; ÇPMY basit ana etkisi) **212**

Ek 10. Toplam Test - Model x Test Güçlüğü x Korelasyon Etkileşimi (Korelasyon basit ana etkisi) **215**

Ek 11. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,2) **228**

Ek 12. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,2; alt test farkları) **231**

Ek 13. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,5) **233**

Ek 14. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,5; alt test farkları) **236**

Ek 15. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,8) **238**

Ek 16. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,8; alt test farkları) **241**

Ek 17. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,2; ÇPMY basit ana etkisi).....243

Ek 18. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,5; ÇPMY basit ana etkisi).....246

Ek 19. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,8; ÇPMY basit ana etkisi).....249

Ek 20. Alt Test - Test Güçlüğü x Model x Korelasyon Etkileşimi (Korelasyon basit ana etkisi).....252

Ek 21. Alt Test - Çoklu Puanlanan Madde Yüzdesi %50 için Model x Korelasyon Etkileşimi -Güvenirlik 259

Ek 21. Alt Test – Model x Korelasyon Etkileşimi (sınıflama doğruluğu).....261

Ek 22. Alt Test – Model x Test Güçlüğü Etkileşimi (sınıflama doğruluğu).....262

Ek 23. Alt Test – Model x Çoklu Puanlanan Madde Yüzdesi Etkileşimi (sınıflama doğruluğu) 263

Ek 24. Alt Test – Model x Alt Test Etkileşimi (sınıflama doğruluğu).....264

TABLOLAR LİSTESİ

Tablo 1. *Simülasyon Çalışması Koşullar, Veri Üretimi ve Analizi İle İlgili Özet Tablo* 46

Tablo 2. *Gerçek Veri İçeriği ve Puanlama* 47

Tablo 3. *Gerçek Veriye Ait Betimsel İstatistikler*.....47

Tablo 4. *Poly-DETECT Analizi Boyutluluk İndeksleri ve Değerlendirme Kriterleri* 49

Tablo 5. *Toplam Puan için RMSE, Güvenirlik ve Sınıflama Doğruluğu Değerlerine ait Faktöriyel Karma ANOVA Sonuçları* 52

Tablo 6. *Toplam Puan için Korelasyon 0,2 Düzeyinde Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Karma ANOVA* 53

Tablo 7. *Toplam Puan için Korelasyon 0,2 Düzeyinde Her Bir Çoklu Puanlanan Madde Yüzdesi için Model ve Test Güçlüğü Değişkenleri ile İki Yönlü Karma ANOVA Sonuçları* 56

Tablo 8. *Toplam Puan için Korelasyon 0,2 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması*.....57

Tablo 9. *Toplam Puan için Korelasyon 0,2 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması* 59

Tablo 10. *Toplam Puan için Korelasyon 0,5 Olduğunda Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Faktöriyel Karma ANOVA*.....60

Tablo 11. *Toplam Puan için Korelasyon 0,5 Düzeyinde Her Bir Çoklu Puanlanan Madde Yüzdesi için Model ve Test Güçlüğü Değişkenleri ile İki Yönlü Karma ANOVA Sonuçları* 62

Tablo 12. *Toplam Puan için Korelasyon 0,5 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması*.....63

Tablo 13. *Toplam Puan için Korelasyon 0,5 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması* 66

Tablo 14. *Toplam Puan için Korelasyon 0,8 Düzeyinde Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Faktöriyel Karma ANOVA*.....67

Tablo 15. *Toplam Puan için Korelasyon 0,8 Düzeyinde Çoklu Puanlanan Madde Yüzdesinin Her Bir Düzeyi için Model ve Test Güçlüğü Değişkenleri ile İki Yönlü Karma ANOVA Sonuçları*.....70

Tablo 16. *Toplam Puan için Korelasyon 0,8 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması*.....71

Tablo 17. *Toplam Puan için Korelasyon 0,8 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması* 73

Tablo 18. *Toplam Puan için Korelasyon 0,2 iken Test Güçlüğünün Her Bir Düzeyi için Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile İki Yönlü Karma ANOVA* 78

Tablo 19. *Toplam Puan için Korelasyon 0,5 iken Test Güçlüğünün Her Bir Düzeyi İçin Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile İki Yönlü Karma ANOVA* 84

Tablo 20. *Toplam Puan için Korelasyon 0,8 iken Test Güçlüğünün Her Bir Düzeyi için Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Karma ANOVA*.....90

Tablo 21. *Toplam Puan için Çoklu Puanlanan Madde Yüzdesinin Her Bir Düzeyinde Model, Test Güçlüğü ve Korelasyon Değişkenleri ile Faktöriyel Karma ANOVA*.....92

Tablo 22. *Alt Puanlar için RMSE, Güvenirlik ve Sınıflama Doğruluğu Değerlerine Ait Faktöriyel Karma ANOVA Sonuçları* 98

Tablo 23. *Alt Puanlar için Korelasyon 0,2 Düzeyinde Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Faktöriyel Karma ANOVA (RMSE, Güvenirlik)* 99

Tablo 24. *Alt Puanlar için Korelasyon 0,2 Düzeyinde Her Bir Çoklu Puanlanan Madde Yüzdesi için Model, Test Güçlüğü ve Alt Test Değişkenleri ile ANOVA Sonuçları (RMSE, Güvenirlik)* 101

Tablo 25. *Alt Puanlar için Korelasyon 0,2 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması (RMSE, Güvenirlik)*.....102

Tablo 26. *Alt Puanlar için Korelasyon 0,2 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması (RMSE, Güvenirlik)* 104

Tablo 27. *Alt Puanlar için Korelasyon 0,2 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Alt Testlerin Karşılaştırılması (RMSE, Güvenirlik)* 106

Tablo 28. *Alt Puanlar için Korelasyon 0,5 Düzeyinde Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Faktöriyel Karma ANOVA (RMSE, Güvenirlik)* 108

Tablo 29. *Alt Puanlar için Korelasyon 0,5 Düzeyinde Her Bir Çoklu Puanlanan Madde Yüzdesi için Model, Test Güçlüğü ve Alt Test Değişkenleri ile ANOVA Sonuçları (RMSE, Güvenirlik)* 110

Tablo 30. *Alt Puanlar için Korelasyon 0,5 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması (RMSE, güvenirlik)* 111

Tablo 31. *Alt Puanlar için Korelasyon 0,5 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması (RMSE, Güvenirlik)* 113

Tablo 32. *Alt Puanlar için Korelasyon 0,5 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Alt Testlerin Karşılaştırılması (RMSE, Güvenirlik)* 115

Tablo 33. *Alt Puanlar için Korelasyon 0,8 Düzeyinde Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Faktöriyel Karma ANOVA (RMSE, Güvenirlik)* 117

Tablo 34. *Alt Puanlar için Korelasyon 0,8 Düzeyinde Her Bir Çoklu Puanlanan Madde Yüzdesi için Model, Test Güçlüğü ve Alt Test Değişkenleri İle ANOVA Sonuçları (RMSE)* 119

Tablo 35. *Alt Puanlar için Korelasyon 0,8 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması (RMSE, Güvenirlik)* 120

Tablo 36. *Alt Puanlar için Korelasyon 0,8 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması (RMSE, Güvenirlik)* 122

Tablo 37. *Alt Puanlar için Korelasyon 0,8 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Alt Testlerin Karşılaştırılması (RMSE, Güvenirlik)* 124

Tablo 38. *Alt Puanlar için Korelasyon 0,2 Olduğunda Test Güçlüğünün Her Bir Düzeyi için Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile İki Yönlü Karma ANOVA (RMSE, Güvenirlik)* 129

Tablo 39. *Alt Puanlar için Korelasyon 0,5 Olduğunda Test Güçlüğünün Her Bir Düzeyi için Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Karma ANOVA (RMSE)* 133

Tablo 40. *Alt Puanlar için Korelasyon 0,8 Olduğunda Test Güçlüğünün Her Bir Düzeyi için Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Karma ANOVA (RMSE, Güvenirlik)* 137

Tablo 41. *Alt Puanlar için Çoklu Puanlanan Madde Yüzdesinin Her Bir Düzeyinde Model, Test Güçlüğü ve Korelasyon Değişkenleri ile Faktöriyel Karma ANOVA (RMSE, Güvenirlik)* 139

Tablo 42. *Alt Puanlar Model x Korelasyon Etkileşim Etkisi için Basit Ana Etkiler (Sınıflama Doğruluğu)* 144

Tablo 43. *Alt Puanlar Model x Test Güçlüğü Etkileşim Etkisi için Basit Ana Etkiler (Sınıflama Doğruluğu)* 146

Tablo 44. *Alt Puanlar Model x Çoklu Puanlanan Madde Yüzdesi Etkileşim Etkisi için Basit Ana Etkiler (Sınıflama Doğruluğu)* 147

Tablo 45. *Alt Puanlar Model x Alt Test Etkileşim Etkisi için Basit Ana Etkiler (Sınıflama Doğruluğu)* 149

Tablo 46. *Gerçek Veriye Ait Alt Puanların Birbiri ve Toplam Puan ile Olan Korelasyon Değerleri* 150

Tablo 47. *Poly-DETECT Analizi Sonuçları* 150

Tablo 48. *DFA Model-Veri Uyum İndeksleri ve Kriter Değerler* 151

Tablo 49. *Gerçek Veri Madde Parametrelerine Ait Betimsel İstatistikler* 152

Tablo 50. *Gerçek Veri Yetenek Parametrelerine Ait Betimsel İstatistikler* 154

Tablo 51. *Gerçek Veri Analizi için Toplam Test ve Alt Testlerde Elde Edilen Yetenek Parametrelerine Ait Standart Hata Ortalama ve Standart Sapma Değerleri* 155

Tablo 52. *Gerçek Veri Analizi için Standart Hataların Tekrarlı Ölçümler ANOVA ile Karşılaştırılması* 156

Tablo 53. *Gerçek Veri Analizi için Marjinal Güvenirlik Katsayıları* 157

Tablo 54. *Gerçek Veri Analizi için Sınıflama Doğruluğu Değerleri* 157

ŞEKİLLER LİSTESİ

Şekil 1. Model A: tek boyutlu model; Model B: ilişkili özellik modeli; Model C: ikinci düzey model; Model D: Bifaktör model.....13

Şekil 2. Yeterliliğin sınıflanmasında yanlış pozitif ve yanlış negatifin şekilsel gösterimi .28

Şekil 3. Toplam puan korelasyon 0,2 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE).....54

Şekil 4. Toplam puan korelasyon 0,2 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik).....55

Şekil 5. Toplam puan korelasyon 0,2 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (sınıflama doğruluğu).....55

Şekil 6. Toplam puan korelasyon 0,5 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE).....61

Şekil 7. Toplam puan korelasyon 0,5 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik).....61

Şekil 8. Toplam puan korelasyon 0,5 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (sınıflama doğruluğu).....62

Şekil 9. Toplam puan korelasyon 0,8 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE).....68

Şekil 10. Toplam puan korelasyon 0,8 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik).....69

Şekil 11. Toplam puan korelasyon 0,8 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (sınıflama doğruluğu).....69

Şekil 12. Toplam puan korelasyon 0,2 – test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE) 75

Şekil 13. Toplam puan korelasyon 0,2 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik) 76

Şekil 14. Toplam puan korelasyon 0,2 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (sınıflama doğruluğu) 77

Şekil 15. Toplam puan korelasyon 0,5 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE) 81

Şekil 16. Toplam puan korelasyon 0,5 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik) 82

Şekil 17. Toplam puan korelasyon 0,5 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (sınıflama doğruluğu) 83

Şekil 18. Toplam puan korelasyon 0,8 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE 87

Şekil 19. Toplam puan korelasyon 0,8 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik) 88

Şekil 20. Toplam puan korelasyon 0,8 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (sınıflama doğruluğu) 89

Şekil 21. Toplam puan çoklu puanlanan madde yüzdesi %50 –test güçlüğü x model x korelasyon (RMSE) 94

Şekil 22. Toplam puan çoklu puanlanan madde yüzdesi %50 – test güçlüğü x model x korelasyon (güvenirlik) 95

Şekil 23. Toplam puan çoklu puanlanan madde yüzdesi %50 – test güçlüğü x model x korelasyon (sınıflama doğruluğu) 96

Şekil 24. Alt puanlar korelasyon 0,2 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE) 100

Şekil 25. Alt puanlar korelasyon 0,2 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik) 101

Şekil 26. Alt puanlar korelasyon 0,2 - çoklu puanlanan madde yüzdesi x model x alt test (RMSE) 105

Şekil 27. Alt puanlar korelasyon 0,2 - çoklu puanlanan madde yüzdesi x model x alt test (güvenirlik) 106

Şekil 28. Alt puanlar korelasyon 0,5 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE) 109

Şekil 29. Alt puanlar korelasyon 0,5 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik) 109

Şekil 30. Alt puanlar korelasyon 0,5 - çoklu puanlanan madde yüzdesi x model x alt test (RMSE) 114

Şekil 31. Alt puanlar korelasyon 0,5 - çoklu puanlanan madde yüzdesi x model x alt test (güvenirlik) 115

Şekil 32. Alt puanlar korelasyon 0,8 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE) 118

Şekil 33. Alt puanlar korelasyon 0,8 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik) 118

Şekil 34. Alt puanlar korelasyon 0,8 - çoklu puanlanan madde yüzdesi x model x alt test (RMSE) 123

Şekil 35. Alt test korelasyon 0,8 - çoklu puanlanan madde yüzdesi x model x alt test (güvenirlik) 124

Şekil 36. Alt puanlar korelasyon 0,2 –test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE) 127

Şekil 37. Alt puanlar korelasyon 0,2 –test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik) 128

Şekil 38. Alt puanlar korelasyon 0,5 – test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE) 131

Şekil 39. Alt puanlar korelasyon 0,5 –test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik) 132

Şekil 40. Alt puanlar korelasyon 0,8 –test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE) 135

Şekil 41. Alt puanlar korelasyon 0,8 –test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik) 136

Şekil 42. Alt puanlar çoklu puanlanan madde yüzdesi %5 –test güçlüğü x model x korelasyon (RMSE) 141

Şekil 43. Alt puanlar çoklu puanlanan madde yüzdesi %5 –test güçlüğü x model x korelasyon (güvenirlik) 142

Şekil 44. Alt puanlar –model x korelasyon (sınıflama doğruluğu) 144

Şekil 45. Alt puanlar - model x test güçlüğü (sınıflama doğruluğu) 145

Şekil 46. Alt puanlar –model x çoklu puanlanan madde yüzdesi (sınıflama doğruluğu) .. 147

Şekil 47. Alt puanlar –model x alt test (sınıflama doğruluğu) 148

BÖLÜM 1

GİRİŞ

Bu bölümde problem durumu belirtilmiş, araştırmanın amacı ve önemi hakkında bilgi verilmiştir.

1.1. Problem Durumu

Gün geçtikçe eğitimde kullanılan testler çok boyutlu ve alt testlerden oluşan daha karmaşık bir yapı haline gelmiştir. Eğitimde ve psikolojide kullanılan birçok test birden fazla yeteneği ölçmekte ve bu durum da bu testleri özü itibariyle çok boyutlu yapmaktadır (Reckase, 1985; 1997). Testler planlanan yapısal ya da içerik yapısından ötürü doğası gereği çok boyutlu olabilir (Ackerman, Gierl & Walker, 2003). Farklı içeriklere ait alt alanlardan oluşan testler, genel bir yeteneği ve ek bazı alt yetenekleri ölçer, yani her madde hem genel yeteneği hem de buna ek olarak ikinci bir yeteneği ölçer. Bir başka ifadeyle, genel yetenek kestirilen toplam puan iken içerik kategorileri için elde edilen alt puanlar, ikincil yetenekler olarak ele alınır (DeMars, 2005). Bu ikincil yeteneklerden elde edilen alt puanlar giderek önem kazanmaktadır (DeMars, 2005; Reckase & Xu, 2015; Sinharay, Haberman & Wainer, 2011; Wedman & Lyren, 2015). Eğitimde uygulanan testlerde alt puanlar ya testin içerik yapısına göre ya da öğrencilerin genel yeterliliği ya da performansının farklı yönlerine ilişkin bilgi sunan testin yapısal durumuna göre raporlanmaktadır (Li, 2018). Bir içerik boyutu, örneğin, bir matematik testindeki cebir ya da geometri gibi teste ait bir içerik alanını temsil etmektedir. Bu açıdan, alt puanlar testin ölçtüğü farklı içerik alanlarında yanıtlayıcıların güçlü ve zayıf yönlerinin ortaya konması için önemlidir (Haberman & Sinharay, 2010).

Ölçme uzmanları için alt puanları kullanarak tanılayıcı bilgi sunma durumu bir ikilem oluşturmaktadır. Bir yönüyle bu bilginin sağlanması, yukarıda bahsedilen gerekçelerle, öğrencilerin ya da herhangi bir sınava giren yanıtlayıcıların güçlü ve zayıf yönlerini anlamaları ve böylece gelişmeleri yönünde rehber olması olarak algılanabilmektedir. Bir başka açıdan ise alt puanlar, bazı alt testlerin ya da bütün alt testlerin az sayıda maddeye sahip olması sebebiyle istenilenden düşük ve hatta kabul gören değerlerin altında güvenirliklere sahip olabilirler. Güvenilir olmayan alt puanların ya da performans kategorilerinin raporlanması, yukarıda açıklanan kullanışlılığına rağmen faydadan çok zarar sağlayabilir. Oluşan bu ikilem psikometristlerin dikkatini çekmekte ve daha yüksek güvenirlikle tanılayıcı bilgi sağlanması amacıyla alt puanların doğru kestirilmesi için birçok yaklaşım önerilmektedir (Chin, 2011). Alt puanlara ait psikometrik özelliklerin iyi olması, bu puanların tanılama ve kabul amacıyla politika yapıcıları, yanıtlayıcılar ve eğitimciler tarafından kullanılması açısından önemlidir (Haberman, 2008; Monaghan, 2006). Eğitimde ve Psikolojide Ölçme Standartları'na (2014) göre de "bir test birden fazla puan sunduğunda, bu puanların farklılığı ve güvenirliği gösterilmelidir" (Standart 1.14, s. 27).

Alan yazında alt puanların kestirildiği ve toplam puana olan katkısının ya da öneminin değerlendirildiği yöntemler bulunmaktadır. Haberman (2008) ve Sinharay (2010) Klasik Test Kuramı (KTK) yöntemlerini kullanarak alt puanların toplam puan üzerindeki katma değeri üzerine çalışmışlardır. Brennan (2012) Haberman'ın yöntemine benzer olan fayda indeksini (utility index) geliştirmiştir. Wainer, Sheehan ve Wang (2000) tarafından geliştirilen alt puan takviye (subscore augmentation) yöntemi, testin diğer kısımlarından bilgi almanın alt puanların daha doğru kestirilmesini incelemek için kullanılmaktadır. Benzer şekilde, Kahraman ve Kamata (2004) tarafından önerilen tamamlayıcı bilgi (collateral information) yöntemi, yanıtlayıcılara ait test dışındaki bilgileri kullanarak daha doğru kestirim yapılmasını hedeflemektedir. Alt puan kestirim yöntemlerinin karşılaştırıldığı çalışmalarda birçok farklı yöntem kestirim doğruluğu ve güvenirlik açısından karşılaştırılmıştır (de la Torre & Patz, 2005; DeMars, 2005; Fan, 2016; Haberman, 2008; Haberman & Sinharay, 2010; Md Desa, 2012; Shin, 2007; Sinharay, 2010; Stone, Ye, Zhu & Lane, 2010; Wang, Chen & Cheng, 2004; Yao, 2014; Yao & Boughton, 2007).

Yetenek kestiriminde kullanılan modellerden biri de örtük özellik kuramına dayanan ve katılımcıların yanıtları ve yetenek seviyeleri arasındaki ilişkiyi matematiksel bir fonksiyonla tanımlamak üzere geliştirilmiş olan Madde Tepki Kuramı'dır (MTK; Lord, 1980). Farklı türdeki maddeler ve bunlara verilen yanıtların incelenmesi amacıyla çeşitli MTK modelleri

geliştirilmiştir. Hambleton ve Jones (1993) uygun MTK modelinin seçilmesinin, maddeye verilen yanıtların ikili ya da çoklu puanlanıp puanlanmadığına, yanıt kategorilerinin sıralı olup olmadığına, maddeye verilen yanıta katkı sağlayan kaç tane yetenek olduğuna ve maddeye verilen yanıtların altında yatan yeteneğe/yeteneklere bağlı olduğunu belirtmektedir. Tek boyutlu MTK modeli, testin örtük yapısının tek boyutlu olduğunu varsaymaktadır, yani maddeye verilen yanıtların altında yatan tek bir örtük özellik bulunmaktadır (Hambleton, Swaminathan & Rogers, 1991). Tek boyutluluk varsayımının karşılandığı durumlarda MTK modelinin işlevselliğini gösteren çok sayıda çalışma bulunmaktadır (örn; Harwell, Stone, Hsu & Kirisci, 1996). Fakat araştırmalar, gerçek test ortamlarında tek boyutluluk varsayımının sıklıkla sağlanamadığını göstermektedir (Ackerman, 1994; Nandakumar, 1994; Reckase, 1985; 1997). Bazı durumlarda boyut sayısı olması gerekenden daha az olarak hesaplanmaktadır (Reckase & Hirsh, 1991).

Testin çok boyutluluğu planlanmış ya da planlanmamış şekilde ortaya çıkabilir. Çok boyutlu bir test ortaya koymadaki temel amaç, farklı içeriklerin bir araya getirilmesidir. Örneğin fen bilimlerindeki bir test öğrencilerin fizik, kimya ve biyoloji gibi alt alanlardaki bilgilerini ölçebilir. Bazı durumlarda da ölçülen özelliğin çok boyutlu olması testin de çok boyutlu olmasına neden olabilir. Dinleme, konuşma, okuma ve yazma olmak üzere dil yeterliliğine ait dört beceriyi ölçen bir dil testi buna bir örnektir. Öte yandan, istenmeden ortaya çıkan çok boyutluluk durumları arasında test hızı (Lu & Sireci, 2007), aynı metne bağlılık (Bradlow, Wainer & Wang, 1999; Hartig & Höhler, 2009) ya da madde formatı sayılabilir (Yao & Schwarz, 2006; Lee, Kim, Choi & Kang, 2019).

Planlanmış çok boyutlu bir test basit ya da karmaşık yapılı olabilir. Basit yapıda, her bir madde tek bir örtük özellikle ilişkilidir ve test birden fazla örtük özelliği ölçmektedir (Reckase, 2009). Bir başka ifadeyle, basit yapılı bir test, iki ya da daha fazla tek boyutlu alt testlerden oluşan çok boyutlu bir testtir (Wang vd., 2004). Maddeler boyutlara kümelendikleri için basit yapı aynı zamanda maddeler-arası çok boyutluluk olarak da isimlendirilir (örn; Adams, Wilson & Wang, 1997; Hartig & Höhler, 2008; te Marvelde, Glas, Van Landeghem, & Van Damme, 2006). Maddeler içi çok boyutluluk olarak da bilinen karmaşık yapı ise çok boyutlu bir testte maddelerin birden fazla örtük özellik ile ilişkili olmasına olanak tanımaktadır (Reckase, 2009).

Reckase (2009) eğitim, psikoloji ve istatistik gibi alanlardaki çalışmaların, insana ait bilgi yapısının karmaşık olduğunu ve bunun maddelere verilen yanıtların genel olarak karmaşık

ve çeşitli olduğuna işaret ettiğini belirtmektedir. Aynı anda birden fazla örtük yapının ölçülmesini gerektiren birçok test bulunmaktadır. Örneğin, ACT ve SAT testleri farklı yapıları ölçtüğü düşünülen ve birçok alt alandan oluşan testlerdir. Türkiye'de ÖSYM tarafından uygulanan birçok sınav da farklı alt alanlardan oluşmaktadır. Böyle durumlarda, birden çok alt testten oluşan bir test için tek boyutlu MTK modellerinin uygulanması, tek boyutluluk varsayımının ihlaline ve alt testler arasındaki farklılıkların ihmal edilmesine sebep olabileceği için uygun olmamaktadır. Çok boyutlu veriye tek boyutlu MTK'nin uygulandığı birçok çalışmada, elde edilen parametrelerin yanlı olabileceği ve yanıtlayıcının yeterliliğine ilişkin çıkarımların hatalı olabileceği vurgulanmıştır (örn; Ackerman, 1989; Ansley & Forsyth, 1985; Way, Ansley, & Forsyth, 1988; Walker & Beretvas, 2010). Tek boyutlu MTK her bir alt teste ayrı ayrı uygulanabilir. Fakat bu durumun da bazı dezavantajları ortaya çıkabilmektedir: a) alt testlerde madde sayısı az olabilir veya b) alt testler genellikle birbiri ile ilişkilidir ve ayrı ayrı tek boyutlu MTK analizleri gerçekleştirildiğinde bu durum göz ardı edilmiş olabilir (Wang, 2012). Ayrıca Chalmers (2012) tek boyutlu modeller kullanışlı olabilse de karmaşık yapıya sahip ölçümlerin daha iyi tanımlanabilmesi için ölçümlere ait boyutluluğun dikkate alınması gerektiğini ifade etmektedir. Gerçek test ortamlarında meydana gelebilen çok boyutlu örtük yapılara uyum sağlayan çok boyutlu Madde Tepki Kuramı (ÇBMTK; Ackerman, 1994; Reckase, 1997; 2009) ise çok boyutlu verilerde kullanılabilmektedir. Ayrıca birden çok alt teste sahip bir teste ait verinin topyekûn analiz edilebilmesi amacıyla maddeler arası/basit yapılı (between item) ÇBMTK'nin kullanımı önerilmiştir (örn; de la Torre, 2008; de la Torre & Patz, 2005; Longabach, 2015; Sheng & Wikle, 2007; Wang vd, 2004). Çalışmalarda çoğunlukla, maddelerin örtük özellikler içerisine yuvalandığı basit yapılı testler ele alınmıştır.

Çok boyutlu bir model tanımlayan ÇBMTK, boyutlar arasındaki korelasyonu da dikkate alır. de la Torre ve Patz (2005) yetenekler arasında ilişki olmadığında ÇBMTK modelinin tek boyutlu model ile benzer sonuçlar ürettiğini fakat farklı alt testlerden elde edilen yetenekler arasında korelasyon olduğunda bu korelasyonu dikkate almanın yetenek kestiriminde önemli bir iyileşmeye neden olduğunu belirtmektedir. Birçok çalışma ile çok boyutlu yetenek kestirimlerinin ayrı ayrı tek boyutlu model ile elde edilen yetenek kestirimlerinden daha az hataya sahip olduğunu belirlenmiştir (örn; Chang & Davison, 1992; Sheng &Wikle, 2007; Wang vd., 2004). Ayrıca bu çalışmalar, çok boyutlu yaklaşım ile elde edilen yetenek kestirimlerinin daha az yanlı olduğunu, gerçek yetenek kestirimleri ile daha yüksek

korelasyona sahip olduğunu ve korelasyonel yapısının gerçek korelasyonel yapıya daha çok benzediğini göstermektedir.

Çok boyutlu testlerde bir toplam puan raporlandığında testin ölçtüğü konunun genel yapısı hakkında yanıtlayıcıların başarı düzeyleri ortaya konulmaktadır. Toplam puan yanıtlayıcıların genel profilini sunarken alt puanlar ise yanıtlayıcıların alt alanlardaki zayıf ve güçlü yönleri hakkında ek bilgi sunmaktadır. Birçok test çok boyutlu bir yapıya sahip olduğu için toplam puan ve alt puanların birlikte raporlanmasına ilişkin artan ilgi ve ihtiyaç olduğu belirtilmektedir (Liu, Li & Liu, 2019). Örneğin İngilizce yeterlilik testi olan TOEFL, okuma, dinleme, yazma ve konuşma olmak üzere dört içerik alanından oluşmaktadır. Bu sınav sonunda yanıtlayıcılar, genel İngilizce yeterliliklerinin göstergesi olan bir toplam puana ve her bir alt alanla ilgili alt puanlara sahip olurlar. Bu puanların eş zamanlı kestirilmesi, testle ilgili paydaşlara öğrencilerin testin geneli ve alt alanlarına ait düzeyleri hakkında daha detaylı bilgi sağlamaktadır (Yao, 2010). Ayrıca toplam puanın ve alt puanların ayrı ayrı elde edilmesi yerine birlikte kestirilmesi, bu bilginin tek bir analizle daha kısa sürede elde edilmesi anlamına gelmektedir.

Alan yazında toplam puan ve alt puanların birlikte kestirildiği yöntemleri ele alan çalışmalar bulunmaktadır (de la Torre & Song, 2009; de la Torre & Hong, 2010; Liu vd., 2019; Soysal & Kelecioğlu, 2018; Yao, 2010). Bu çalışmaların tümünde, toplam puan ve alt puanların raporlanması gerektiği durumlarda bu puanlara ait güvenirliğin oldukça önemli olduğu vurgulanmaktadır. Yao (2010) alt puanların ortalamasının alınarak toplam puan elde edilmesinin en sık kullanılan yöntem olduğunu belirtmektedir. Fakat bu yöntemin, farklı alt testlerdeki farklı maksimum ham puan olasılığını, alt testler arasındaki korelasyonu ve farklı puan noktalarında toplam puan ve alt puanlar arasındaki ilişkinin farklı olabilme olasılığını dikkate almadığını belirtmektedir. Bu problemlerin üstesinden gelebilmek için Yao (2010) basitçe ortalama alma yerine, ÇBMTK maksimum bilgi yöntemini önermiştir. Bu yöntemde toplam puan ve alt puanlar arasında lineer bir ilişki varsayılmamıştır. Alt puanlar ÇBMTK ile kestirilmiş ve toplam puan da maksimum bilgi yöntemi ile elde edilmiştir.

de la Torre ve Song (2009) ise genel ve alt alanlara ait yeteneklerin beraber kestirilmesi için Üst Düzey Madde Tepki Modeli'nin (ÜDMTK) kullanılmasını önermişlerdir. ÜDMTK ile alt boyutlara ait yeteneklere ek olarak genel yetenek de hesaplanır. ÇBMTK modelinin aksine ÜDMTK modelinde toplam puan ve alt puanlar arasında lineer bir ilişki olduğunu varsayılmaktadır. Bifaktör modele göre de toplam puan ve alt puanların birlikte kestirildiği

çalışmalar yer almaktadır (DeMars, 2013; Liu vd., 2019). Bu modelde maddeler hem bir alt alan hem de genel boyutla ilişkilidir. ÇBMTK ve ÜDMTK modellerinden farklı olarak Bifaktör model için testte yer alan alt yetenekler birbirlerine ve genel yeteneğe ortogonaldir/diktir.

Toplam puan ve alt puanların birlikte kestirilmesi ile ilgili yapılan araştırmalar az sayıdadır (de la Torre & Song, 2009; 2010; Liu vd., 2019; Soysal & Kelecioğlu, 2018; Yao, 2010). Sass, Schmitt ve Walker (2008), örtük özellik modellerinde madde güçlüğü, ayırt ediciliği ve şans parametreleri ve örtük özellik kişi parametreleri gibi parametreler arasında en önemlisinin örtük özellik parametresi olduğunu belirtmiştir. Çünkü örtük özellik kestirimleri, yanıtlayıcıların yeterlilik sınıflarının belirlenmesinde kullanılabilir. Zhang (2010) modellerin sınıflama performansının önemli olduğunu ve bununla birlikte yanıtlayıcılar hakkında daha fazla tanılayıcı bilgi elde etmek amacıyla alt alanların her biri için ayrı ayrı sınıflama yapılabileceğini belirtmektedir. Test puanlarına dayalı olarak alınan kararların genellikle iki yönlü (örn; geçti-kaldı, yeterli-yeterli değil) olduğu ele alındığında, psikometrik model karşılaştırmalarında sınıflama doğruluğunun dâhil edilmesi önemlidir (Chin, 2011). Fakat birkaç istisna dışında (örn; Yao & Boughton, 2007) toplam puan ve alt puanların birlikte kestirildiği modellerin sınıflama performansına ilişkin yeterli bilgi bulunmamaktadır.

Daha önce yapılan çalışmalarda kullanılan veri setleri çoğunlukla yalnızca ikili puanlanan (örn; de la Torre & Hong, 2010; de la Torre & Song, 2009) ya da yalnızca çoklu puanlanan (örn; Wang, 2012) maddelerden oluşmaktadır. Ancak gerçek durumda her zaman bunun mümkün olmadığı, bazı durumlarda hem çoktan seçmeli hem de yapılandırılmış maddelerin birleşiminden oluşan karma format testlerin birlikte kullanıldığı bilinmektedir (örn; İngilizce yeterlilik sınavları, TIMSS). Hem çoktan seçmeli hem de yapılandırılmış yanıtlı maddelerin birleşiminden oluşan testler birçok testte kullanıldığı için karma-format verinin kullanılması önemli görülmektedir (Lane, 2005; Shin, 2007; Yao & Schwarz, 2006). Ayrıca madde güçlüğünün güvenirliğe olan etkisi alan yazında pek çok çalışmada tartışılmıştır (Allen & Yen, 1979; De Gruijter & van der Kamp, 2008; Paek, 2015). Bu çalışmada da test güçlüğünün ve testte yer alan yapılandırılmış madde sayılarının farklı düzeylerinde ele alınan modellerin performansının incelenmesine de yer verilmiştir.

Bu bölümde detaylı olarak açıklanan alan yazındaki bu boşluğa ilişkin olarak bu araştırmanın problemini, toplam puan ve alt puanların birlikte kestirilmesi amacıyla

kullanılan modeller (ÇBMTK, ÜDMTK ve Bifaktör model) ile kestirim hatalarının, güvenirliklerin ve sınıflama doğruluklarının karma-format veri yapısı üzerinde ve farklı koşullarda yetenek kestirimleri açısından incelenmesi oluşturmaktadır. Ayrıca konu ile ilgili az sayıda çalışmanın bulunduğu alan yazına katkı sunmak hedeflenmektedir.

1.2. Araştırmanın Amacı

Bu araştırmanın ana amacı, toplam test ve alt testlere ilişkin yetenek kestirim modellerinin karşılıklı olarak incelenmesidir. Çalışma kapsamında ele alınan ÇBMTK, ÜDMTK ve Bifaktör modelleri, kestirim doğruluğu, güvenirlik ve sınıflama doğruluğuna dayalı olarak karşılaştırılmıştır. Çalışmada simülasyon veri ve Türkiye'de bir devlet üniversitesi Yabancı Diller Eğitimi Yüksek Okulu tarafından geliştirilip uygulanan İngilizce yeterlilik sınavına ait gerçek veri kullanılmıştır. Bu genel amaç doğrultusunda hem gerçek hem de simülasyon veri gruplarında aşağıdaki sorulara yanıt aranmıştır.

1. ÇBMTK, ÜDMTK, Bifaktör kestirim modelleri ve farklı simülasyon koşulları (test güçlüğü, boyutlar arası korelasyon ve çoklu puanlanan madde yüzdesi) altında simülasyon veri ile elde edilen yetenek kestirimlerine ait RMSE, güvenirlik ve sınıflama doğruluğu değerleri alt puan ve toplam puan bağlamında farklılaşmakta mıdır?
2. Gerçek veri üzerinde ÇBMTK, ÜDMTK, Bifaktör kestirim modelleri ile elde edilen yetenek kestirimlerine ait hata, güvenirlik ve sınıflama doğruluğu değerleri alt puan ve toplam puan bağlamında farklılaşmakta mıdır?

1.3. Araştırmanın Önemi

Bu çalışma, ÇBMTK, ÜDMTK ve Bifaktör model ile elde edilen toplam puan ve alt puanların kestirimlerinin doğruluğu, güvenirlikleri ve sınıflama performansları açısından karşılaştırmayı hedeflemektedir. Testler daha karmaşık ve çok boyutlu bir hal aldıkça test paydaşları genellikle tek bir genel puanın raporlanmasından daha fazlasına ihtiyaç duymaktadır. Örneğin bir yeterlilik sınavına giren adaylar, özellikle testi geçemeyenler, tekrar teste girmeden önce hangi alt alanlarda kendilerini geliştirmeleri gerektiğini bilmek istemektedir. Özellikle ulusal bir değerlendirmede yeterli olarak sınıflanamayan K-12 öğrencilerinin öğretmenleri ve aileleri, öğrencilerin hangi akademik standartları

karşılayamadığını belirlemek isteyebilmektedir. Bu gibi durumlarda, geçerli ve güvenilir tanılayıcı geri bildirim verilmesi, yanıtlayıcıların eksik oldukları alanlarda kendilerini geliştirmek üzere bir plan yapmasına olanak sağlamaktadır. Bu noktada genel yeterliliğe ek olarak testin içindeki alt alanlara ait yeteneğin belirlenmesi önemli görülmektedir. Bununla birlikte testin bütünü için hesaplanan toplam puan, testin ölçmeyi hedeflediği genel yapıya ilişkin bir değer ortaya koyar. Her iki puan türünün de kullanım amaçları doğrultusunda raporlamada yer aldığı durumlar olabilmektedir. Bu çalışma, toplam puan ve alt testlere ait puanları birlikte raporlamaya ilişkin artan bir ilgi ve ihtiyaca karşılık vermeyi ve alan yazında bu konuda yapılan çalışmaların eksikliğini gidermeyi hedeflemesi açısından önemlidir. Bu çalışmanın sonuçları, toplam test ve alt testlere ait yetenek kestiriminde kullanılan modellerin performansı ile ilgili alan yazına katkı sunması beklenmektedir. Ayrıca sonuçların, alanda yaptıkları uygulamalar için bir model seçmek durumunda olan uygulayıcılara da katkı sağlayacağı düşünülmektedir.

Hem toplam puan hem de alt puanlar için bu puanlara ilişkin yeterli güvenirlik ve geçerlik kanıtları sunulması gerekmektedir. Bu noktada, bu çalışma farklı yöntemler kullanılarak hesaplanan alt puanların ve toplam puanın kestirilen güvenirliklerinin ve hataların karşılaştırılması bakımından önemli olduğu düşünülmektedir. Ayrıca bu puanlara ilişkin geçerlik kanıtları, kestirilen puanlar ile alınan sınıflama kararlarının doğruluğu incelenerek elde edilmiştir. Elde edilen kestirimlerin doğruluğu ve güvenirliğine ek olarak sınıflama performanslarının da değerlendirilmesi modeller hakkında daha detaylı bilgi edinmek açısından önemlidir. Sınıflama performansı testten elde edilen yetenek kestirimlerine dayalı olarak alınan kararların doğruluğu açısından incelenmektedir.

Alt testlere ait yeteneklerin kestirildiği yöntemlerin incelendiği çalışmalar içerisinde az sayıda çalışma özellikle İngilizce yeterliliğinin ölçülmesi ve değerlendirilmesi üzerinedir (örn; Longabach, 2015; Reckase & Xu, 2015). Ölçülecek örtük özelliklerin yapısal olarak farklı olabilmesi (örn; matematik, fen bilgisi, dil becerisi vb.) ve boyutlar arasındaki korelasyonun sınıf seviyesine ya da yetenek düzeylerine göre değişebilmesi sebebiyle, farklı test türleri ile kestirim modellerinin incelenmesi önerilmektedir (de la Torre, Song & Hong, 2011). İngilizce yeterliliği en genel hali ile dinleme, okuma, yazma ve konuşma alt becerilerinden oluşmaktadır. Genel İngilizce yeterliliği için bu dört temel becerinin ayrı ayrı kazanılması ve bu yeterliliği ölçen bir sınavın bu alt alanları içermesi gerekmektedir. Bu durum özü itibariyle bu testleri çok boyutlu yapmaktadır. Bu sebeplerle, İngilizce yeterliliğinin ölçülmesinin temel alınması ve özellikle bu tür testlerde sıklıkla kullanılan

karma-format test yapısının ele alınması önemli görülmektedir. Ele alınan üç modelin karşılaştırılmasına ek olarak bu çalışma karma-format test yapısını dâhil ederek daha önceki çalışmaları geliştirmektedir. Geniş ölçekli testlerde alt puanların da raporlanması gerekliliği ve karma-format testlerin yaygın kullanımı açısından karma-format testlerde farklı kestirim yöntemlerinin karşılaştırıldığı bir çalışma yapma ihtiyacı ortaya çıkmıştır. Bu çalışma sonuçları ile ilgili alan yazına katkı sunmanın yanı sıra Türkiye'de Milli Eğitim Bakanlığı ya da Öğrenci Seçme ve Yerleştirme Merkezi tarafından geliştirilip uygulanan geniş ölçekli, farklı içerik alanlarına sahip alt testlerden oluşan testlerin puanlanmasına katkı sunabileceği düşünülmektedir.

1.4. Sınırlılıklar

1. Simülasyon çalışmasında verilerin üretilmesinde kullanılan model, ilişkili çok boyutlu modeldir. Ayrıca dört alt boyut için üretilen gerçek yetenek kestirimleri kullanılarak gerçek genel yetenek kestirimi üretilmesinde çok boyutlu MTK maksimum bilgi yöntemi kullanılmıştır.
2. Gerçek verideki çoklu puanlanan madde sayısı üç olup oldukça azdır. Ayrıca yazma alt testi sadece iki çoklu puanlanan maddeden oluşmaktadır.

BÖLÜM 2

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde çok boyutlu örtük özellik, çok boyutlu madde tepki kuramı modellerinin tanıtılması, özellikleri ve sınıflama doğruluğu ile ilgili genel bilgiler sunulmuş ve ilgili araştırmalar özetlenmiştir.

2.1. Çok Boyutlu Örtük Özellik

"Örtük özellik" teriminin ilk kez Lazarsfeld (1950) tarafından sunulduğu ve "örtük özellik kuramı"nın ise Lord'un (1952, 1953a, 1953b) çalışmaları ile meydana geldiği ifade edilmektedir (Akt. Hambleton, Swaminathan, Cook, Eignor & Gifford, 1977). Örtük özellik kuramı, bir testteki yanıtlayıcı performansının, özellikler olarak nitelenen yanıtlayıcı özelliklerini tanımlayarak, yanıtlayıcıların bu özelliklere ait puanlarını kestirerek ve bu puanları test performansının belirlenmesinde kullanarak tahmin edilebileceğini ya da açıklanabileceğini varsaymaktadır. Özelliklerin direkt olarak ölçülememesi sebebiyle örtük özellikler ya da yetenekler olarak ifade edilir (Lord & Novick, 1968, s. 358). Örtük özellik teorisine dayanan Madde Tepki Kuramı, yanıtlayıcıların maddelere verdikleri yanıtlar ve onların yetenekleri arasındaki ilişkiyi matematiksel bir fonksiyon ile tanımlamak üzere geliştirilmiştir (Lord, 1980, s. 12).

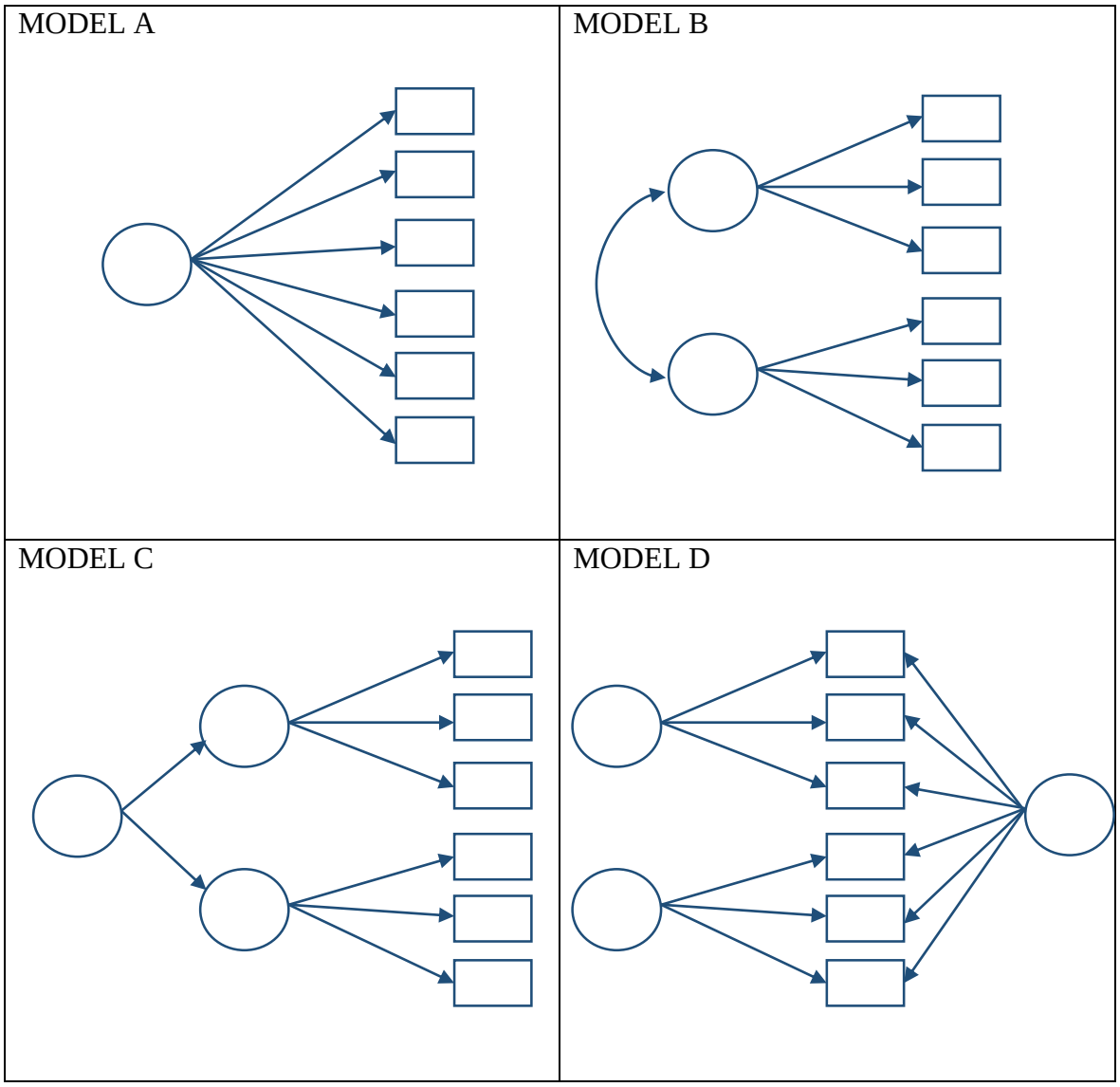
Örtük alanın boyutluluğu, yerel bağımsızlık ve madde karakteristik eğrisi, örtük özellik kuramı ile bağlantılı olarak ortaya çıkan üç önemli husustur (Lord & Novick, 1968, s. 359-365; Lord, 1980, s. 19; Hambleton vd., 1977). Yerel bağımsızlık, aynı yetenek düzeyindeki bireylerin oluşturduğu bir yanıtlayıcı grubu için maddelere doğru yanıt verme olasılığının birbirinden bağımsız olması anlamına gelmektedir (Lord & Novick, 1968, s. 360-362). Bir

başka ifadeyle aynı θ koşulunda seçilen herhangi bir yanıtlayıcının maddelere verdiği yanıtların istatistiksel olarak birbirinden bağımsız olması ve bir maddeye verdiği yanıtın diğer maddelere verdiği yanıtlardan etkilenmiyor olması anlamına gelmektedir (Hambleton & Swaminathan, 1985, s. 23-25). Madde karakteristik eğrisi, maddenin doğru yanıtlanma olasılığı ve yetenek düzeyi arasındaki ilişkiyi gösterir. Bir j maddesinin doğru yanıtlanma olasılığı en genel haliyle $P_i(\theta)$ şeklinde gösterilir. Madde niteliklerine ve ölçülen örtük özelliğe bağlı belirlenen bu eğri, monotonik olarak artan bir fonksiyonu tanımlamaktadır. (Baker, 2001, s. 21; Embretson & Reise, 2000, s. 47; Hambleton vd., 1991, s. 9).

Örtük özellik kuramı boyutluluk açısından ele alındığında, yanıtlayıcının performansının açıklanmasında tek bir örtük yeteneğin yeterli olduğu örtük özellik modellerinin tek boyutlu modeller olarak ele alındığı görülmektedir. Yanıtlayıcının test performansının daha doğru bir şekilde ortaya konmasında birden fazla yeteneğin gerektiği modeller ise çok boyutlu modeller olarak tanımlanmaktadır (Hambleton vd., 1977). Tek boyutlu MTK modellerinde maddeye verilen yanıtların altında tek bir örtük özellik bulunduğu varsayımı vardır. Tek boyutluluk varsayımının ihlal edilmesi, testin tek bir baskın özellikten daha fazlasını ölçtüğü anlamına gelip yapılan kestirimlerin geçerliliği ve güvenirliği açısından sorun oluşturabilmektedir (Hambleton vd., 1991, s. 9; Embreson & Reise, 2000, s. 48). Çok boyutlu MTK modelleri ise maddeye verilen yanıtlarla ilişkili birden çok yetenek olduğu durumlar için modelleme yapabilmek üzere geliştirilmiştir (Ackerman, 1994; Reckase, 1997, 2009). Reckase (2009, s. 57) eğitim, psikoloji, test geliştirme, psikometri ve istatistik alanındaki gelişmelerle birlikte insana ait bilginin daha derinleşmesi ve test maddelerine verilen yanıt süreçlerini daha karmaşık ve çeşitli hale gelmesinin bir sonucu olarak çok boyutlu MTK'nın ortaya çıktığını belirtmektedir.

Eğitimde yapılan ölçmelerde ve psikolojik yapıların ölçülmesinde çok boyutluluğa sebep olan birçok durum görülmektedir. Wainer ve Thissen (1996) bir test birden çok farklı içeriğe sahip alt alanları ölçmek üzere geliştirildiğinde, çok boyutluluğun ortaya çıkmasının doğal ve muhtemel olduğunu belirtmiştir. Testin farklı madde formatlarından olması da potansiyel çok boyutluluk kaynağı olarak görülmektedir (Zhang, 2016). Ayrıca Yen (1993) maddelerin ortak bir köke bağlı olmasını, yanıtlayıcı hızı, puanlayıcı etkilerini ve uygun olmayan bazı dışsal müdahaleleri yerel bağımlılığın olası sebepleri arasında göstermiş ve bu durumun da boyutluluğu etkileyeceğini ifade ederek çok boyutluluğa neden olabileceğini vurgulamıştır. Bu çalışmada testin içerik yapısından kaynaklanan çok boyutluluk ele alınmıştır.

Reise, Moore ve Haviland (2010) farklı sebeplerle ortaya çıkan çok boyutlu örtük modellerle ilgili grafiksel gösterim sunmuştur. Tek boyutlu bir model ve çok boyutlu üç modele ait örnek grafikler Şekil 1'de yer almaktadır. Daireler boyutları ya da örtük özellikleri ve kareler ise örtük özellikleri ölçen maddeleri simgelemektedir.



Şekil 1. Model A: tek boyutlu model; Model B: ilişkili özellik modeli; Model C: ikinci düzey model; Model D: Bifaktör model. Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of personality assessment, 92*(6), 544-559. https://doi.org/10.1080/00223891.2010.496477

Şekil 1'de yer alan Model A, tek boyutlu modeli göstermektedir. Testte yer alan her bir madde, tek bir ortak özellik ile ilişkilidir. Diğer üç modelde çok boyutlu yapılar yer almaktadır. Model B çok boyutlu ilişkili özellik modelidir ve modelde yer alan her bir boyut

testteki bazı maddelerle ilişkilidir. Bu modelin, test benzer içeriklere ve birbiriyle yüksek ilişkili maddelere sahip birden çok kümeden oluştuğunda kullanılması yaygındır. Fakat Model B'de ölçülen ya da madde varyansını direkt olarak etkileyen ortak bir boyut yoktur. Model C ise örtük özelliklere ait üst düzey yapıyı içeren çok boyutlu bir modeldir ve ikinci düzey faktör model olarak da anılır. Model C ve Model B arasındaki fark, Model C'de birinci düzey özellikler arasındaki farkı açıklayan bir ikinci düzey özelliğin yer almasıdır. Ayrıca Model C, ölçme yapısını faktörler arasındaki korelasyona dayandırmıştır ve faktörler ortak bir nedene sahip oldukları için birbirleri ile ilişkilidir. Bir başka ifadeyle, bu üst düzey model, alt boyutlar arasındaki korelasyonun neden kaynaklandığını gösteren "üst düzey" ya da "ikinci düzey" boyutun hedef yapı olduğunu belirtmektedir. Burada üst düzey boyut ile maddeler arasında direkt bir ilişki yoktur; alt faktörler, maddeler ve üst faktör arasındaki ilişkiye aracılık etmektedir. Son olarak, Model D bifaktör model yapısını temsil etmektedir. Bifactör modelde iki tür faktör yapısı vardır: 1) bütün maddelerle ilişkili "genel" faktör ve 2) maddelerden bazıları ile ilişkili ve genel faktör tarafından açıklanmayan ek varyansı gösteren alt faktörler. Bu genel faktör, maddelerin ortak özelliğini ve bu hedef boyuttaki bireysel farklılıkları yansıtmaktadır. Alt faktörler, madde yanıtlarındaki genel faktör tarafından açıklanmayan varyansı açıklayan olası boyutlardır. Bifaktör modelde, bütün maddelerin ortak bir örtük özelliği ölçtüğü fakat madde varyansının buna ek olarak ikinci bir alt faktörden de etkilendiği varsayılmaktadır (Reise, Morizot & Hays, 2007; Reise vd., 2010).

2.2. Çok Boyutlu Madde Tepki Kuramı

Çok Boyutlu Madde Tepki Kuramı, kişilerin çok boyutlu alandaki yerini, test maddesine doğru yanıt verme olasılıkları ile ilişkilendirir. Madde yanıtlarından kestirilen parametreler ile gösterilen madde özellikleri, bu ilişkiye dolaylı yoldan katkı sağlar. Kuram genel olarak matematiksel işlemlere dayanır ve bu işlemler bir takım varsayımlar içermektedir (Reckase, 2009, s. 8-9):

1. Değerlendirilen kişilerin yeri (yeteneği), test süresince değişmez.
2. Bir test maddesinin özellikleri, kullanıldığı ölçme durumlarının hepsinde sabittir.
3. Yanıtlayıcının bir test maddesine verdiği yanıtlar, diğer test maddelerine verdiği yanıtlardan bağımsızdır. Maddelerin, diğer maddelere verilen yanıtları etkilememesi

beklenmektedir. Bu varsayım, Madde Tepki Kuramı bağlamında "yerel bağımsızlık" olarak adlandırılmaktadır.

4. Çok boyutlu alandaki her bir noktanın ilişkili olduğu sadece bir tane doğru yanıt olasılığı vardır. Bu varsayım, kişiler ile maddeler arasındaki etkileşimi temsil eden modelin matematiksel ifadeleri için önemlidir. Bir diğer ifadeyle, yanıtlayıcıların yeteneği arttıkça (en azından azalmaması istenir) test maddesini doğru yanıtlama olasılıklarının artmasıdır. Bu varsayım, "monotonik artış" varsayımı olarak da adlandırılmaktadır.

Çok boyutlu MTK modelleri, kişiler ve test maddeleri arasındaki ilişkiyi tanımlamada tek boyutlu MTK modellerinden farklılaşmaktadırlar. Çünkü ÇBMTK'de tek bir hipotetik yapı değil birden fazla hipotetik yapının yanıtlayıcı performansını etkilediği ele alınır. Bu yüzden "çok boyutlu" olarak isimlendirilirler (Reckase, Ackerman & Carlson, 1988). Diğer MTK modellerinde olduğu gibi analizin temel odağı maddelere verilen yanıtlardır. En çok kullanılan modeller iki kategorili olanlardır; çok kategorili çok boyutlu MTK modelleri de mevcuttur. Çok Boyutlu MTK modelleri, gerçek yanıtları değil, yanıtların olasılıklarını modeller. Modellerin en temel ifadesi şu şekildedir (Reckase, 2009, s. 59):

$$P\,(U = u|\theta) = f\,(u, \theta, \gamma) \qquad \text{(Eşitlik 1)}$$

U yanıtlayıcının madde puanı

u maddeye verilebilecek olası puanlar arasından atanan puanın değeri

θ çok boyutlu alanda yanıtlayıcının yerini tanımlayan parametreler vektörü

γ maddenin özelliklerini tanımlayan parametreler vektörü

Farklı ÇBMTK modelleri, yetenek boyutları ve performans arasında farklı istatistiksel ilişkiler olduğunu varsayar. Ayrıca boyutlar ve maddeler arasındaki ilişkiler örüntüsü, basit yapılı bir yük matrisi (a loading matrix with a simple structure – between item multidimensionality) ya da karmaşık bir yük yapısı (within item multidimesionality) ile tanımlanabilir. Bunun dışında, bir madde için yanıtlama performansını etkileyen birden çok boyutun etkileşimini ele alan telafisel (compensatory) ve telafisel olmayan (non-compensatory) yöntemler söz konusudur.

2.2.1. Veri Setinin Yapısına Göre: Basit Yapı – Karmaşık Yapı (Between- and within-item multidimensionality)

ÇBMTK modelleri arasındaki önemli farklardan biri, her bir madde için başarı olasılığının modeldeki boyutların yalnızca birinden etkilenip etkilenmediği ya da bir maddeye verilen yanıtların aynı anda birden çok yetenek boyutuna bağlı olarak modellenip modellenmediğidir. İlk durum, maddeler arası çok boyutluluk (between-item multidimensionality) ve diğeri ise madde içi çok boyutluluk (within-item multidimensionality) olarak tanımlanmaktadır (Adams, Wilson & Wang, 1997). Maddeler arası çok boyutluluk modellerinde, modeldeki her bir boyutu ölçmek için birbirinden ayrı madde kümeleri kullanılır. Faktör analizi terimleri ile ifade edecek olursak, bu modeller yüklerin (loadings) basit bir yapısı ile tanımlanır ve birçok tek boyutlu ölçme modelinin tek ve ortak bir modelde birleştirilmesi olarak düşünülebilirler. Bu birleşim, gizil yetenek boyutları arasındaki ilişkilerin modellenmesine olanak sağlar. Madde içi çok boyutluluk modellerinde, birden çok yeteneğin karışımı, her bir maddeye verilen yanıtın altında yatan etken olarak modellenir (Hartig & Höhler, 2009).

Maddeler arası çok boyutlu olan bir test birden fazla alt testten oluşur. Her bir alt test tek bir yeteneği veya özelliği ölçer ve kendi içinde homojendir. Her bir madde, yer aldığı alt test ile ölçülmeye çalışılan tek bir özellik ile ilişkilidir. Bu türdeki çok boyutlu testlere çoklu-tek boyutlu (multi-unidimensional) adı verilmektedir (Sheng 2005; Sheng & Wikle, 2007; Bulut, 2013). Madde içi çok boyutlu modelde ise maddeler tek bir yetenek ile ilişkili değildir. Hem testin bütünü hem de testteki maddeler birden çok yetenek ile ilişkilidir. Bunun bir sonucu olarak, faktör analizi terimleri ile ifade edecek olursak, maddeler ölçülmek istenen birden fazla yeteneğe faktör yükü verir. Yük yapısının seçimi, gizil boyutların yorumlanması ile doğrudan ilişkilidir. Hangi modelin diğerine tercih edilmesi gerektiği ile ilgili öneri, yalnızca araştırma önerisi doğrultusunda yapılabilir. Çünkü farklı modeller veriye uyum açısından eşdeğer olabilir (Hartig & Höhler, 2008).

Madde içi çok boyutluluğu kapsayan modeller, farklı yetenekler ve görevler arasındaki ilişkiyi modelleme için uygundur. Bir maddeyi yanıtlama olasılığı, farklı yetenek boyutlarının birleşiminin bir fonksiyonu olarak modellenebilir. Madde içi çok boyutlu modeller, maddelerin yanıtlanması için ihtiyaç duyulan yetenekler hakkında varsayımlar gerektirmektedir. Bu varsayımlar, yeteneklerin kombinasyonlarının telafisel olup olmadığına ilişkin bilgiyi gerektirir. Sonuç olarak, madde içi çok boyutluluk modelleri, tek

bir yetenek boyutu ile açıklanamayan karmaşık görevlerdeki performansı modellemede kullanılır. Yani, yanıtlayıcının yetenekleri ve test maddeleri arasındaki ilişkinin modellenmesini sağlar (Hartig & Höhler, 2009).

Madde içi çok boyutlu modellere göre daha az karmaşık olması ve gizil değişkenlerin kolaylıkla yorumlanabilmesi, maddeler arası çok boyutlu modelin avantajlarındandır. Gizil değişken için kestirilen puanlar, belirlenen madde seti için performansın net ölçümlerini gösterir. Eğer araştırmada ana amaç, belli konu alanları için performans ölçümleri elde etmek ise maddeler arası çok boyutlu model, daha karmaşık olan maddeler içi çok boyutlu modelden daha uygun olacaktır (Hartig & Höhler, 2009).

2.2.2. Boyutlar Arasındaki Telafisel (Compensatory) ve Telafisel Olmayan (Non-compensatory) İlişki

Maddeler içi çok boyutlu modellerde, maddelerin yanıtlanması için gerekli olan boyutlar farklı şekillerde bütünleştirilir. En önemli özellik, bu bütünleşmenin telafisel fonksiyon içerip içermediğidir. Çok boyutlu MTK modellerinin çoğu telafiseldir. Bu modeller, bir boyuttaki yüksek yeteneğin diğer boyutlardaki düşük yeteneği telafi etmesini sağlar. Bir diğer ifadeyle, bu telafi, diğer boyutlardaki düşük yeteneğin maddeyi doğru yanıtlama olasılığını çok düşürmesini engellemektir. Telafisel olmayan modelde ise soruyu yanıtlama olasılığının yüksek olması bütün boyutlarda yüksek yetenek sahibi olunduğu anlamına gelir. Her boyut ayrı ayrı ele alınır ve en son kestirilen olasılık her bir olasılığın çarpımıdır (Reckase, 2009, s. 79).

Boyutların farklı birleşmeleri arasındaki fark, iki basit model ile gösterilebilir. Maddenin her iki boyuta da yüklendiği ve madde güçlüğünün sıfır olduğu iki boyutlu Rasch modeli için madde yanıt fonksiyonu (item response function) şu şekildedir:

$$\Pr(x|\theta_1, \theta_2) = \text{logit}(\theta_1 + \theta_2) \qquad \text{(Eşitlik 2)}$$

Telafisel modelde, başarı olasılığı yetenek boyutları θ_1 ve θ_2'nin toplamına bağlıdır. Bir yetenek diğerini telafi edebileceği için θ_1 ve θ_2 toplamı, maddeyi doğru yanıtlama olasılığı ile aynı sonucu verecektir. Bu model telafisel olmayan model ile karşılaştırılacak olursa, güçlük eklendiğinde ve şans parametresi sıfır olduğunda iki boyut için madde tepki fonksiyonu aşağıdaki gibidir:

$$\Pr(x|\theta_1, \theta_2) = \text{logit}(\theta_1) \cdot \text{logit}(\theta_2) \qquad \text{(Eşitlik 3)}$$

Telafisel olmayan modelde başarı olasılığı yetenek boyutları θ_1 ve θ_2'nin çarpımına bağlıdır. Yani bir madde için başarı olasılığı yalnızca o maddenin çözümü için gerekli yeteneklerin hepsi için yüksek değerler olduğunda yüksektir (Hartig & Höhler, 2009).

2.3. Çok Boyutlu MTK Modelleri ve Kestirimi

Birçok çok boyutlu MTK modeli geliştirilmiştir; bunlardan bazıları Rasch modelin çok boyutlu uzantısı (Adams vd., 1997), çok boyutlu 2PL model (Reckase, 1985), çok boyutlu 3PL model (Reckase, 2009), normal ogiv modelin çok boyutlu uzantısı (Bock & Schilling, 2003; McDonald, 1999; Samejima, 1974), çok boyutlu kısmi puan modeli (Kelderman & Rijkes, 1994), çok boyutlu genelleştirilmiş kısmi puan modeli (Yao & Schwarz, 2006) ve çok boyutlu aşamalı tepki modeli (Muraki & Carlson, 1995) olarak sayılabilir. Bu çalışmada ikili puanlanan maddeler için çok boyutlu 3PL model (M-3PL) ve çoklu puanlanan maddeler için çok boyutlu genelleştirilmiş kısmi puan modeli (M-2PPC) kullanılmıştır. Aşağıda bu modeller sırasıyla açıklanmıştır.

N yanıtlayıcı ve J maddeden oluşan bir veri seti olduğu varsayılsın. Madde yanıtları $X = \{X_{ij}\}$ şeklindeki bir matriste yer alır; burada $i = 1, 2, \dots, N$ ve $j = 1, 2, \dots, J$ şeklinde gösterilebilir. *i.* yanıtlayıcının yanıtlarını içeren vektör $\vec{X}_i = (X_{i1}, X_{iJ})$ şeklindedir. Yanıtlayıcılara ait yetenek parametreleri aşağıdaki eşitlikle ifade edilmektedir.

$$\theta = \left(\vec{\theta}_i, \dots, \vec{\theta}_N\right)^T \qquad \text{(Eşitlik 4)}$$

Burada, $\vec{\theta}_i = (\theta_{i1}, \dots, \theta_{iD})$ olup *i.* kişi için D boyut sayısı kadar yetenek parametresi olduğunu göstermektedir. İkili puanlanan bir *j* maddesi için $\vec{\theta}_i$ yeteneğine sahip bir yanıtlayıcının bu maddeye doğru yanıt verme olasılığı M-3PL model için aşağıdaki eşitlik ile gösterilmektedir (Reckase, 1997):

$$P_{ij1} = P(x_{ij} = 1 \mid \vec{\theta}_i, \vec{\beta}_j) = \beta_{3j} + \frac{1 - \beta_{3j}}{1 + e^{(-\vec{\beta}_{2j} \odot \vec{\theta}_i^T + \beta_{1j})}} \qquad \text{(Eşitlik 5)}$$

x_{ij} 0 ya da 1'dir ve *i.* yanıtlayıcının *j* maddesine verdiği yanıt

$\vec{\beta}_j = (\vec{\beta}_{2j}, \beta_{1j}, \beta_{3j})$ *j* maddesine ait parametreler

$\vec{\beta}_{2j} = (\beta_{2j1}, \dots, \beta_{2jD})$ madde ayırt edicilik parametrelerinin D boyut sayısı kadar parametreyi içeren vektörü

β_{1j} madde güçlük parametresi

β_{3j} alt asimptot / şans parametresi

$\vec{\beta}_{2j} \odot \vec{\theta}_i^T$ iki vektörün skaler çarpımı

Çoklu puanlanan j maddesi için $\vec{\theta}_i$ yeteneğine sahip bir yanıtlayıcının $k-1$ kategorisinde yanıt verme olasılığı M-2PPC model için aşağıdaki eşitlik ile gösterilmektedir (Yao & Schwarz, 2006):

$$P_{ijk} = P(x_{ij} = k-1 \mid \vec{\theta}_i, \vec{\beta}_j) = \frac{e^{(k-1)\vec{\beta}_{2j}\odot\vec{\theta}i - \sum_{t=1}^{k}\beta_{\delta_t j}}}{\sum_{m=1}^{K_j} e^{\left((m-1)\vec{\beta}_{2j}\odot\vec{\theta}_i^T - \sum_{t=1}^{m}\beta_{\delta_t j}\right)}} \quad \text{(Eşitlik 6)}$$

$x_{ij} = 0, \dots, K_j - 1$ i. yanıtlayıcının j maddesine verdiği yanıt

$\vec{\beta}_j$=($\vec{\beta}_{2j}$, $\beta_{\delta_2 j}$, …, $\beta_{\delta_{K_j} j}$) j maddesine ait parametreler

$\vec{\beta}_{2j} = (\beta_{2j1}, \dots, \beta_{2jD})$ madde ayırt edicilik parametrelerinin D boyut sayısı kadar parametreyi içeren vektörü

$\beta_{\delta_t j}$ k kategori için eşik parametreleri, $k = 1, 2, \dots, K_j$. $\beta_{1j} = 0$ ve K_j = j maddesi için yanıt kategori sayısı

M-3PL ve M-2PPC modeller için yanıt olasılıkları sırasıyla Eşitlik 7 ve 8 ile de ifade edilebilir:

$$P_{ij} = P\left(X_{ij} \mid \vec{\theta}_i, \vec{\beta}_j\right) = P_{ij1}^{1(X_{ij}=1)}(1 - P_{ij1})^{1(X_{ij}=0)} \quad \text{(Eşitlik 7)}$$

$$P_{ij} = P\left(X_{ij} \mid \vec{\theta}_i, \vec{\beta}_j\right) = \prod_{k=1}^{K_j} P_{ij1}^{1(X_{ij}=k-1)} \quad \text{(Eşitlik 8)}$$

Buradaki yanıt kategorilerine ait gösterim ise Eşitlik 9'da yer almaktadır.

$$1_{(X_{ij}=k)} = \begin{cases} 1 & eğer\ X_{ij} = k \\ 0 & değilse \end{cases} \quad \text{(Eşitlik 9)}$$

Bütün maddeler için madde parametreleri şöyledir:

$$\beta = (\vec{\beta}_1, \dots, \vec{\beta}_j, \dots, \vec{\beta}_J)^T \quad \text{(Eşitlik 10)}$$

Olabilirlik fonksiyonuna ait matematiksel gösterim Eşitlik 11'de yer almaktadır.

$$P(X\,|\theta,\beta) = \prod_{i=1}^{N} P\left(\vec{X}_i\,|\vec{\theta}_i,\ \beta\right) = \prod_{i=1}^{N}\prod_{j=1}^{J} P\left(X_{ij}|\vec{\theta}_i,\vec{\beta}_j\right) \qquad \text{(Eşitlik 11)}$$

$P(\theta|\lambda)$, λ koşulunda θ yeteneğindeki bir yanıtlayıcı evreni için olasılık dağılımı olsun. Örneğin, eğer θ'nın normal dağıldığı varsayılırsa $\lambda = (\mu, \sigma)$ olur ve burada da μ ortalama ve σ varyans-kovaryans matrisidir. Birleşik sonsal dağılıma ait matematiksel gösterim Eşitlik 12'de yer almaktadır ve Eşitlik 13 ve 14 ile açıklanmaktadır.

$$P(\theta,\beta,\lambda|X \propto P(X|\theta,\beta,\lambda)P(\theta|\beta,\lambda)P(\beta)P(\lambda)\ = P(X|\theta,\beta)P(\theta|\lambda)P(\beta)P(\lambda) \quad \text{(Eşitlik 12)}$$

$$P(X\,|\theta,\beta) = \prod_{i=1}^{N}\prod_{j=1}^{J} P\left(X_{ij}|\vec{\theta}_i,\vec{\beta}_j\right) \qquad \text{(Eşitlik 13)}$$

$$P(\theta|\lambda) = \prod_{i=1}^{N} P\left(\vec{\theta}_i|\lambda\right) \qquad \text{(Eşitlik 14)}$$

Burada, bir yanıtlayıcının bir maddeye verdiği yanıtın yanıtlayıcının yeteneğine ve madde parametrelerine dayandığı varsayılmaktadır (Yao & Schwarz, 2006).

Çok boyutlu MTK'deki bütün parametrelerin gösterimi tek boyutlu modellerdeki gösterime benzerdir. Tek fark yetenek ve madde ayırt edicilik parametrelerinin vektör olarak yer almasıdır. Bu teknik farklılık ve boyut sayısının fazla olması çok boyutlu MTK modellerini daha karmaşık hale getirmektedir (Reckase, 2009, s. 63). MCMC yöntemi, boyut sayısı çok olduğunda sorun oluşturabilecek türev almadan kurtularak sonsal dağılımı kullanmaktadır. Model parametreleri (θ,β,λ), birleşik sonsal dağılımdan $P(\theta,\beta,\lambda|X)$ örneklem alan Metropolis-Hastings algoritmasını kullanarak kestirilir (Patz & Junker, 1999a, 1999b).

2.3.1. Çok Boyutlu MTK İçin Toplam Puanın Hesaplanması – Maksimum Bilgi Yöntemi

Çok boyutlu MTK modeline dayalı olarak toplam puanın hesaplanması amacıyla Yao (2010), farklı alt testlerdeki farklı maksimum ham puan olasılığını, alt testler arasındaki korelasyonu ve farklı puan noktalarında toplam puan ve alt puanlar arasındaki ilişkinin farklı olabilme olasılığını dikkate aldığını belirttiği maksimum bilgi yöntemini (maximum information method) basitçe ortalama alma yerine önermiştir. Yöntem, test bilgi fonksiyonlarına dayanmaktadır. Yao ve Schwarz (2006) M-3PL ve M-2PPC modelleri için

çok boyutlu madde ve test bilgi fonksiyonlarını geliştirmiştir. M-3PL modelde bir *j* maddesi için $\vec{\theta}$ yetenek vektöründeki bilgi fonksiyonuna ait matematiksel ifade Eşitlik 15'te yer almaktadır.

$$I_j(\vec{\theta}) = \frac{\left(P_{j1} - \beta_{3j}\right)^2 (1 - P_{j1)}}{P_{j1}(1 - \beta_{3j})^2} \vec{\beta}_{2j} \otimes \vec{\beta}_{2j} \qquad \text{(Eşitlik 15)}$$

Burada yer alan P_{j1} Eşitlik 15'te gösterilmektedir. $\vec{\theta}$ yetenek vektöründeki olasılığı göstermek amacıyla Eşitlik 7'den farklı olarak burada *i* yoktur. $\otimes$ simgesi dış çarpımı ifade etmektedir ve $\vec{\beta}_{2j} \otimes \vec{\beta}_{2j}$ D × D bir matristir. Bu matristeki m. satır ve n. sütundaki eleman $\vec{\beta}_{2j}$'deki m. satır ve n. sütundaki elemanların çarpımıdır.

M-2PPC modelde bir *j* maddesi için $\vec{\theta}$ yetenek vektöründeki bilgi fonksiyonuna ait matematiksel ifade ise Eşitlik 16'da sunulmuştur.

$$I_j(\vec{\theta}) = \sigma^2 \vec{\beta}_{2j} \otimes \vec{\beta}_{2j} \qquad \text{(Eşitlik 16)}$$

Burada σ^2 Eşitlik 17 ile açıklanır.

$$\sigma^2 = \sum_{k=1}^{K_j} (k-1)^2 P_{jk} - \left[\sum_{k=1}^{K_j} (k-1) P_{jk}\right]^2 \qquad \text{(Eşitlik 17)}$$

Eşitlik 17'de yer alan P_{jk} Eşitlik 6'da verilen denklemin *i* olmadığı şeklidir.

$\vec{\theta}$ yetenek vektöründeki test bilgi fonsiyonu Eşitlik 18 ile gösterilmektedir ve madde bilgi fonksiyonlarının toplamıdır.

$$I(\vec{\theta}) = \sum_{j=1}^{J} I_j(\vec{\theta}) \qquad \text{(Eşitlik 18)}$$

Yetenek vektörüne ait varyans $V(\vec{\theta})$ test bilgi fonksiyonunun tersi $I(\vec{\theta})^{-1}$ ile belirlenir. $\alpha = (\alpha_1, \dots, \alpha_D)$ yönünde, bileşik puan, varyansı ve standart hatası sırasıyla Eşitlik 19, 20 ve 21'de yer almaktadır.

$$\theta_\alpha = \sum_{l=1}^{D} \theta_l w_l \qquad \text{(Eşitlik 19)}$$

$$V(\theta_\alpha) = \vec{w} V(\vec{\theta}) \vec{w}^T \qquad \text{(Eşitlik 20)}$$

$$SEM(\theta_\alpha) = V(\theta_\alpha)^{1/2} \qquad \text{(Eşitlik 21)}$$

Burada yer alan ağırlık vektörü $\vec{w} = (cos^2\alpha_1, \dots, cos^2\alpha_D)$ ile ifade edilir ve ağırlık toplamları 1'e eşittir, $\sum_{l=1}^{D} w_l = 1$. $\vec{w}^T$ ise $\vec{w}$ vektörünün transpozudur. Yön $\alpha_{\vec{\theta}} = (\alpha_1, \dots, \alpha_D)$ şeklinde olduğunda $\vec{w}V(\vec{\theta})\vec{w}^T$ olası bütün açılarda minimum değeri alır. Daha sonra $\alpha_{\vec{\theta}}$ yönünde yetenek vektöründeki $(\vec{\theta})$ bileşik puan, θ_α, maksimum bilgiye ulaşır ve böylece en güvenilir bileşik puan meydana gelmiş olur (Yao, 2010, 2011, 2012a, 2012b).

2.4. Üst Düzey (Higher-Order) Madde Tepki Kuramı

Yetenekler arasındaki korelasyonu dikkate alan bir diğer kestirim süreci Üst Düzey Madde Tepki Kuramı modeli (de la Torre & Song, 2009) ile yapılabilmektedir. Üst düzeyde daha geniş ve genel boyutları ve alt düzeyde daha özel alt alanları içeren ÜDMTK modeli, yeteneklerin hiyerarşik olarak modellenmesinin tipik örneklerinden biridir. Bu açıdan ÜDMTK modeli psikolojik araştırmalarda ve uygulamada sıklıkla kullanılan hiyerarşik yetenek yapısı ile benzerdir (örn; Carrol, 1993; Cronbach & Snow, 1977). Eğitim araştırmalarında da alt puanların ölçülmesindeki gelişmelerle birlikte bu model de alt alanlara ait yeteneklerin ve toplam yeteneğin birlikte kestirilmesine olanak sağlamıştır (de la Torre vd., 2011). Birinci düzey model, her maddenin bir boyuta yüklendiği maddeler-arası ÇBMTK modeli ile benzerlik göstermektedir. İkinci düzey ise genel ve alt boyutlardaki yetenekler arasındaki ilişkiyi tanımlayan lineer bir modeldir. Bu lineer ilişki alt boyutlar arasındaki korelasyonu belirlemektedir.

ÜDMTK, alt alanların korelasyonel yapısı ile ortaya konan test içi tamamlayıcı bilgiyi kullanan çoklu-tek boyutlu (multi-unidimensional) bir modeldir. Böylece farklı alt alanlardan alt testlerin ya da testin genelinin bu model ile analiz edilmesi ile daha doğru parametre kestirimlerinin elde edilmesi beklenir (de la Torre & Hong, 2010). ÜDMTK modele ait formüller ve kestirim algoritması de la Torre ve Song (2009) tarafından ortaya konmuştur ve aşağıda kısaca özetlenmektedir.

ÜDMTK modelde, her alt alanın tek boyutlu olduğu varsayılır ve her bir alt alana ilişkin yetenek, $\theta_i^{(d)}$, i yanıtlayıcısının d (d = 1, 2, …, D) alt alanına ilişkin performansına işaret etmektedir. Bu farklı alanlar aynı yeteneği ölçtüğünde testin bütünü tek boyutlu düşünülebilir. Üst düzey bir genel yeteneğin olması [$\theta_i \sim N(0,1)$] alt yetenekler arasındaki korelasyon ile açıklanmaktadır. Bu modelde birinci düzey, alt yetenek için madde

performansını tanımlayan bir MTK modelidir. İkinci düzey alt yetenekler ve genel yetenek arasında lineer ilişkileri tanımlar. Alt yetenekler, genel yeteneğin lineer fonksiyonları şeklinde ifade edilir:

$$\theta_i^{(d)} = \lambda^{(d)}\theta_i + \varepsilon_{id} \qquad \text{(Eşitlik 22)}$$

$\lambda^{(d)}$ d yeteneğinin genel yeteneği yordamasındaki örtük katsayı, $|\lambda^{(d)}| \leq 1$; $\lambda^{(d)} \sim$ U(-1,1)

ε_{id} hata terimi, $\varepsilon_{id} \sim$ N(0, 1 - $\lambda^{(d)2}$), diğer hata terimlerinden bağımsız

Alt yeteneklerin genel yetenek koşulunda birbirinden bağımsız oldukları varsayılır. Genel yetenek ve alt yetenekler arasındaki korelasyon $\lambda^{(d)}$ ile gösterilirken d ve d' alt yetenekleri arasındaki korelasyon $\lambda^{(d)} \times \lambda^{(d')}$ şeklinde ifade edilir. Matematiksel olarak $\lambda^{(d)}$ değeri negatif olabilir fakat çoğu uygulamada negatif olmaması beklenir. Çünkü alt yetenekler genel yetenekle ilişkili ise doğal olarak birbirleri ile de pozitif korelasyona sahiptirler.

Yanıtlayıcılara ait yeteneklerinin belirlenmesi için alt test ya da toplam teste ilişkin yetenek parametrelerine ek olarak modele ait örtük regresyon parametrelerinin ($\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(D)}$) de kestirilmesi gerekmektedir. Model, hiyerarşik Bayes yaklaşımı kullanılarak Eşitlik 23, 24 ve 25 ile ifade edilmektedir.

$$\theta_i \sim N(0,1) \qquad \text{(Eşitlik 23)}$$

$$\lambda^{(d)} \sim U(-1.0, 1.0) \qquad \text{(Eşitlik 24)}$$

$$\theta_i^{(d)} \mid \theta_i, \lambda^{(d)} \sim N\left(\lambda^{(d)}\theta_i, 1 - \lambda^{(d)2}\right) \qquad \text{(Eşitlik 25)}$$

Genel yetenek θ_i normal dağılımdan ve regresyon katsayısı tek biçimli dağılımdan elde edilir. Daha sonra alt alanlara ait yetenekler, MCMC yöntemi ile kestirilen genel yetenek ve regresyon katsayılarına dayalı olarak üretilir. Regresyon katsayılarının bilindiği varsayıldığında, genel ve alt alanlara ait yetenek kestirimleri, maksimum olabilirlik gibi geleneksel yöntemlerle kolaylıkla elde edilebilir. Fakat yetenek kestirimlerine ek olarak regresyon katsayılarının da kestirilmesi, boyutluluk ve karmaşıklıktan ortaya çıkan problemleri arttırmaktadır. Bu sebeple model parametreleri MCMC algoritması kullanılarak hesaplanmaktadır. Genel yetenek, tam koşullu dağılım kullanılarak elde edilmiştir: $N(c \cdot \frac{\sum_{d=1}^{D}\lambda^{(d)}\theta_i^{(d)}}{1-\lambda^{(d)2}}, c)$ ve burada $c^{-1} = 1 + \sum_{d=1}^{D}\lambda^{(d)}/(1 - \lambda^{(d)2})$. Burada verilen koşullu dağılımdan da anlaşılacağı üzere genel yetenek (θ_i) en temel haliyle alt boyutlara ait

yeteneklerin ($\theta_i^{(d)}$) ağırlıklandırılmış bir ortalamasıdır (de la Torre & Hong, 2010; de la Torre & Song, 2009).

2.5. Bifaktör Model

Genel bir örtük özelliğe ait alt boyutların ölçülmesi amacıyla Gibbons ve Hedeker (1992), ikili puanlanan maddeler için Tam-Bilgi Madde Bifaktör Analizi'ni (Full-information Item Bifactor Analysis) geliştirmişlerdir. Doğrulayıcı çok boyutlu MKT modelinin özel bir hali olan madde Bifaktör analizi, ölçme aracının boyutluluğu ve bireysel farklılıkların ölçeklenmesi ile ilgili stratejiler hakkında bilgi sağlar. Gibbons vd. (2007) ise bu analizi, çoklu puanlanan maddelerde kullanılacak şekilde genişletmişlerdir. Her iki çalışmada da hesaplamalar normal ogiv modele dayalı olarak üretilmiştir. Cai, Yang ve Hansen (2011) çalışmasında Bifaktör modelin, lojistik fonksiyonu kullanan çok boyutlu MTK modelleri ile kullanımını ortaya koymuştur. Ek olarak bu çalışmada, hem ikili hem de çoklu puanlanan maddelere yönelik modeller ele alınmıştır. Bifaktör model, aynı anda genel ve alt yeteneklere ait kestirimler sunmaktadır. Modelde her madde hem genel faktöre hem de alt faktörlerden birine yüklenmektedir. Alt faktörler, hem genel faktöre hem de birbirlerine diktir (orthogonal), yani ilişkili olmadığı varsayılır. Modelin genel amacı genel yeteneği ölçmektir; alt alanlara özgü boyutlar, nüans boyutlarıdır ve artıkları açıklamaktadırlar (DeMars, 2013; Yao, 2010).

Cai, Yang ve Hansen (2011) standart madde bifaktör ölçme yapısının aşağıdaki faktör örüntüsüne sahip olduğunu belirtmektedir:

$$\begin{pmatrix} a_{10} & a_{11} & 0 \\ a_{20} & a_{21} & 0 \\ a_{30} & a_{31} & 0 \\ a_{40} & 0 & a_{42} \\ a_{50} & 0 & a_{52} \\ a_{60} & 0 & a_{62} \end{pmatrix}$$

Burada altı madde, bir genel boyut ve iki alt boyut yer almaktadır. Görüldüğü üzere ilk faktör genel faktör, diğer ikisi ise alt faktörlerdir. Burada a'lar faktör yükleri ile benzeşen madde ayırt edicilik parametrelerini simgelemektedir. Madde ayırt ediciliklerin altındaki ilk indis maddeyi, ikincisi ise faktörü belirtmektedir. 0 alt indisi genel faktörü, 1 ve 2 de sırasıyla alt faktörleri göstermektedir. Bu örnekte ilk üç madde hem genel faktöre hem de birinci alt

faktöre yüklenmektedir. Son üç madde ise hem genel faktöre hem de ikinci alt faktöre yüklenmektedir.

2PL model için *i.* yanıtlayıcının *j* maddesine doğru yanıt verme olasılığı Eşitlik 26'da yer almaktadır (Liu vd., 2019).

$$P(\vec{\theta}_i) = \frac{e^{1.7(\vec{a}_j'\vec{\theta}_i + d_j)}}{1 + e^{1.7(\vec{a}_j'\vec{\theta}_i + d_j)}} \qquad \text{(Eşitlik 26)}$$

$\vec{a}_j' = (\beta_{aj}, 0, \dots, \beta_{sj}, \dots, 0)$ madde ayırt edicilik parametreleri vektörü

d_j madde güçlüğü

$\vec{\theta}_i = (\theta_{ia}, \theta_{i1}, \dots, \theta_{is}, \dots, \theta_{iS})$ her bir yanıtlayıcıya ait yetenek vektörü

Bir *j* maddesi için madde ayırt edicilik parametreleri vektöründe sıfırdan farklı iki değer vardır; biri genel faktör (β_{aj}) için diğeri de ilişkili olduğu *s*. alt faktör (β_{sj}) içindir. Her bir yanıtlayıcıya ait yetenek vektörü ise genel faktör için bir genel yetenek (θ_{ia}) ve her alt faktör için bir tane olmak üzere alt faktör sayısı kadar yetenek parametresi ($\theta_{i1}, \dots, \theta_{is}, \dots, \theta_{iS}$) içermektedir.

Bifaktör modele dayalı olarak toplam puan (θ_{toplam}), genel faktörün (θ_{ia}) ve diğer bütün alt faktörlerin ($\theta_{i1}, \dots, \theta_{is}, \dots, \theta_{iS}$) ağırlıklandırılmış birleşimidir (weighted composite). *s*. alt boyut için alt puan (θ_{alt_s}) ise genel faktörün (θ_{ia}) ve ilgili alt faktörün (θ_{is}) ağırlıklandırılmış birleşimidir:

$$\theta_{toplam} = w_{1a}\theta_{ia} + \sum_{s=1}^{S} w_{1s}\theta_{is} \qquad \text{(Eşitlik 27)}$$

$$\theta_{alt_s} = w_{2a}\theta_{ia} + w_{2s}\theta_{is} \qquad \text{(Eşitlik 28)}$$

w_{1a} toplam puan için genel faktörün ağırlığı

w_{1s} toplam puan için alt faktörlerin ağırlığı

w_{2a} alt puan için genel faktörün ağırlığı

w_{2s} alt puan için alt faktörün ağırlığı

Puan kestiriminde ağırlıklar büyük rol oynamaktadır. Bu çalışmada, Yao (2010) çalışması ile benzer şekilde ağırlıklandırma 1 ve 0 olacak şekilde belirlenmiştir:

$$w_{1a} = 1, w_{1s} = 0 \qquad \text{(Eşitlik 29)}$$

$$w_{2a} = 0, w_{2s} = 1 \qquad \text{(Eşitlik 30)}$$

Bu ağırlıklandırmada toplam puan genel faktörle ve alt puan da ilgili alt faktörle temsil edilmektedir. Liu vd. (2019) çalışmasında farklı ağırlıklandırma yöntemleri geliştirmiş ve detaylı olarak sunmuştur.

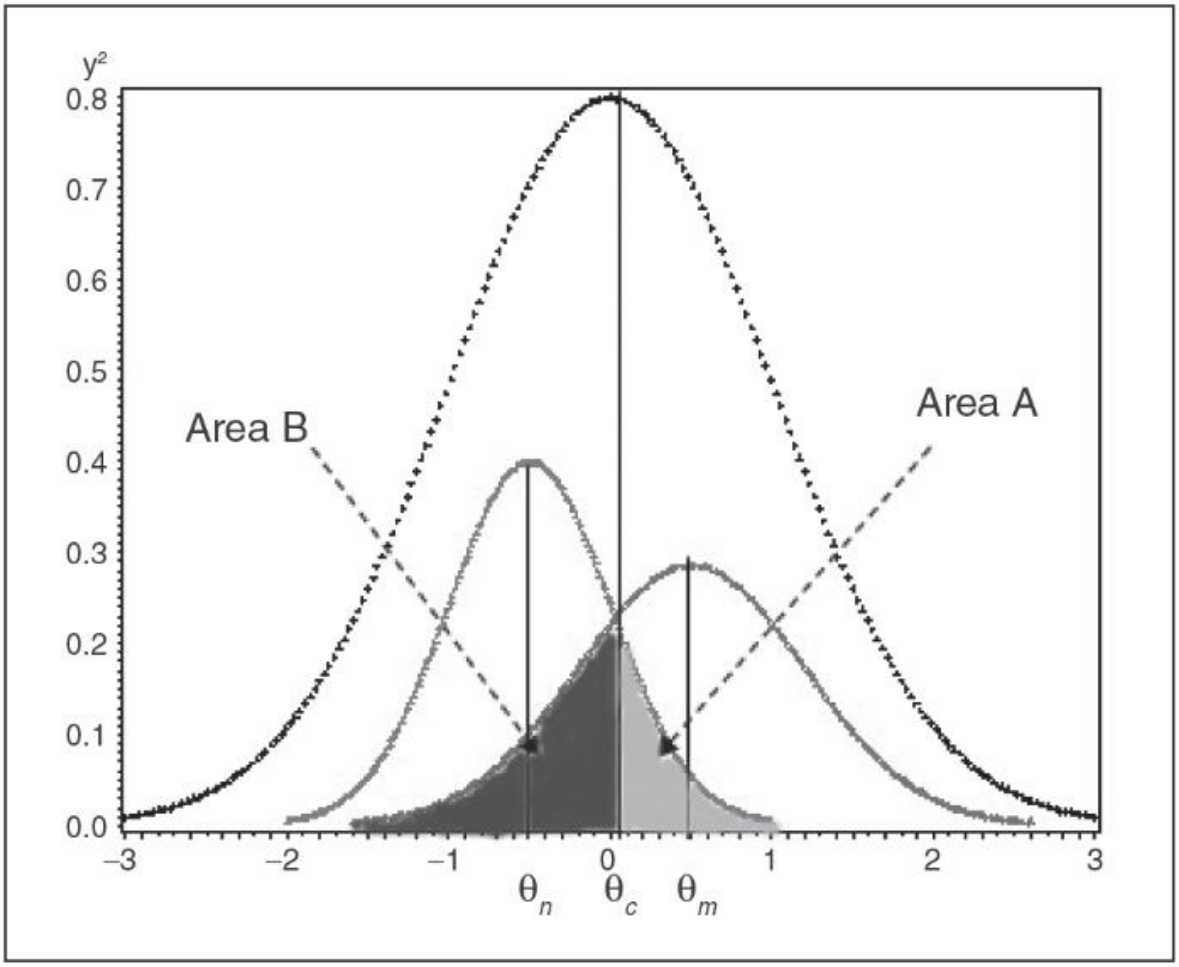
2.6. Elde Edilen Puanlara Ait Sınıflama Performansı

Kriter referanslı testler, öğrencilerin beklenen öğrenme düzeyine ulaşıp ulaşmadıklarının ya da seviyelerinin belirlenmesi amacı kapsamında eğitimde en sık kullanılan testlerdir. Öğrenci başarıları hakkında daha doğru kararlar alınması öncelikli bir konudur (Wang, Song, Ding & Meng, 2016). Fakat testler bütünüyle hatadan arınık değildir ve hataların en aza indirilerek önemli kararların alınması gerekmektedir. Sınıflama tutarlılığı, yanıtlayıcıların tekrarlı ya da paralel ölçümlerde de aynı kategoriye tekrar sınıflanma düzeyidir. Sınıflama doğruluğu ise gözlenen sınıflamaların gerçek sınıflamalar ile uyum derecesine işaret etmektedir (Lee, 2010). Bu bağlamda sınıflama tutarlılığı terimi, tutarlılık tanımının tekrarlı ölçümler gerektirmesinden dolayı sınıflama güvenirliği olarak anılmaktadır. Sınıflama doğruluğu ise gözlenen puanlara göre yapılan sınıflamanın, gerçek puanlara göre yapılan sınıflamaya göre doğruluğunun derecesini değerlendirdiği için daha çok sınıflamanın geçerliliği ile ilgilidir. Sınıflama doğruluğu ve tutarlılığı indekslerinin elde edilmesi, sınıflama sürecine ait psikometrik özelliklerin değerlendirilmesi açısından önemlidir (Lee, Hanson & Brennan, 2002). Sınıflama tutarlılığı ve sınıflama doğruluğu kavramları, sınıflandırmaların güvenilirliğini ve geçerliliğini tanımlayan indeksler olarak önerilmiştir (Hambleton & Novick, 1973).

Sınıflama doğruluğu ve tutarlılığının hesaplanması için birçok yöntem geliştirilmiştir. Yapılan ilk çalışmalarda iki uygulamaya dayalı indeksler geliştirilmiştir; bunlardan en yaygın olanları *P* uyum indeksi (Hambleton & Novick, 1973) ve Cohen *kappa*'dır (Swaminathan, Hambleton & Algina, 1974). İki uygulamaya dayalı yöntemler, aynı yanıtlayıcıları iki kez test etmeyi gerektirdiği için uygulamada zorluklara neden olmaktadır. Bu sebeple tek uygulamaya dayalı sınıflama doğruluğu ve tutarlılığı indekslerinin kestirilmesine yönelik çalışmalar yapılmıştır. Bu yöntemler temelde geliştirildikleri kuramlara göre KTK'ye dayalı yöntemler (Huynh, 1976; Hanson & Brennan, 1990; Livingstone & Lewis, 1995; Subkoviak, 1976) ve MTK'ye dayalı yöntemler (Hambleton & Han, 2004; Huynh, 1990; Wang, Kolen & Harris, 2000; Lee, 2010; Rudner, 2001, 2005) olarak iki kategoriye ayrılmaktadır.

MTK'ye dayalı sınıflama tutarlılığı ve doğruluğu indekslerine ilişkin ilk çalışma ikili puanlanan maddeler için Huynh (1990) tarafından yapılmıştır. Daha sonra Wang vd. (2000) çoklu puanlanan MTK modellerini kullanarak MTK yöntemlerini genişletmişlerdir. Lee (2010) marjinal toplam puan dağılımlarını integral ile örtük özellik (θ) dağılımı ile birleştirmeyi gerektiren indeksleri geliştirmiştir. Rudner (2001, 2005) ise yetenek (θ) ölçeğine dayanan indeksler üretmiştir. Bu yöntemde, sınıflamaların beklenen olabilirliği hesaplanarak sınıflama doğruluğu elde edilmektedir. Aşağıda Rudner yöntemine ait ikili sınıflama (geçti-kaldı) için açıklamalar yer almaktadır.

Kesme puanı θ_c, A yanıtlayıcısına ait gerçek yetenek θ_n ve B yanıtlayıcısına ait gerçek yetenek θ_c olsun. Bu değerlerin yetenek ölçeğindeki gösterimi Şekil 2'de yer almaktadır (Zhang, 2010). Yetenek kestirimi içerisinde hatayı da içermesi sebebiyle her bir gerçek yetenek kestirimi aslında koşullu bir dağılıma sahiptir. Şekilde θ_n kesme puanından (θ_c) daha düşük olduğu için A yanıtlayıcısının, bütün kestirimlerde "kaldı" sınıfında yer alması gerekmektedir. Benzer şekilde θ_m kesme puanından yüksek olduğu için B yanıtlayıcısı "geçti" şeklinde sınıflanmalıdır. Fakat A yanıtlayıcısının geçenler sınıfında yer alma ihtimali bulunmaktadır ve Şekil 2'de yetenek kestirimlerinin kesme puanından daha büyük olduğu A bölgesi ile gösterilmektedir. Sınıflama terminolojisi ile bu ihtimal, yani gerçekte "kaldı" sınıfında olup "geçti" olarak kestirilmesi yanlış pozitif olarak tanımlanmaktadır.



Şekil 2. Yeterliliğin sınıflanmasında yanlış pozitif ve yanlış negatifin şekilsel gösterimi. Zhang, B. (2010). Assessing the accuracy and consistency of language proficiency classification under competing measurement models. *Language Testing, 27*(1), 119–140. https://doi.org/10.1177/0265532209347363

Kesme puanının üstünde gözlenen puan elde etme olasılığı olan A bölgesi, normal dağılım eğrisi altında Eşitlik 31'deki z puanının sağında kalan alandır.

$$z = \frac{\theta_c - \theta_n}{se(\theta_n)} \qquad (Eşitlik\ 31)$$

Eşitlik 31'de yer alan $se(\theta_n)$, A yanıtlayıcısına ait θ_n kestirimlerinin standart hatasıdır. A bölgesi (yanlış pozitif), yani gerçek sınıfı "kaldı" olan bir A yanıtlayıcısının "geçti" sınıfında yer alma olasılığı $P(\hat{\theta}_n > \theta_c | \theta_n)$ ile ifade edilir. Bütün yanıtlayıcılar için gerçek durumu "kaldı" olup "geçti" olarak sınıflananların beklenen oranı Eşitlik 32 ile elde edilir.

$$P(m,n) = \sum_{n=1}^{N} \left[P(\hat{\theta}_n > \theta_c | \theta_n) d(\theta_n) \right] \qquad (\text{Eşitlik } 32)$$

Eşitlik 32'de N gerçekte kalanların sayısıdır ve $\hat{\theta}_n$ ise gerçek yeteneğe (θ_n) ait kestirimdir. $d(\theta_n)$ gerçek yeteneğe ait evren yoğunluğunu yani gerçek puanı (θ_n) olan yanıtlayıcıların beklenen oranını göstermektedir. Benzer şekilde gerçek sınıfı "geçti" olup yanlış bir şekilde "kaldı" sınıfında yer alma ihtimali olanların (yanlış negatiflerin) beklenen oranı Eşitlik 33 ile elde edilmektedir.

$$P(n,m) = \sum_{m=1}^{M} \left[P(\hat{\theta}_m > \theta_c | \theta_m) d(\theta_m)\right] \qquad \text{(Eşitlik 33)}$$

Eşitlik 33'te yer alan M gerçek sınıfı "geçti" olanların sayısıdır. Her iki sınıf için de doğru sınıflama oranları da benzer mantıkla elde edilmektedir. Dolayısıyla sınıflama doğruluğu kişilerin gerçek sınıflarında bulunma olasılığının yüksek olması ile ilişkilidir.

2.7. İlgili Araştırmalar

Wang vd. (2004) tek boyutlu ve çok boyutlu yaklaşım ile alt puan kestirimini ele almışlardır. Çalışmada biri ikili puanlanan maddelerden oluşan beş alt teste sahip fen bilgisi testi ve diğeri dörtlü Likert türünde çoklu puanlanan maddelerden oluşan 10 alt boyuta sahip Öğretmek Kişilik Ölçeği olan iki ayrı gerçek veri kullanılmıştır. Analizler Rasch modelleri ve marjinal maksimum olabilirlik kestirim yöntemi ile gerçekleştirilmiştir. Her iki veri seti için de sonuçlar, örtük özellikler arasındaki korelasyonu dikkate alan çok boyutlu yaklaşımın, tek boyutlu yaklaşıma göre daha doğru ölçümler yaptığını göstermektedir. Ek olarak ölçme doğruluğundaki iyileşme, örtük özelliklerin sayısı ve test uzunluğuna göre değişmektedir. Örtük özellik sayısı arttıkça ve testte yer alan madde sayısı azaldıkça, iyileşme daha anlamlı bir hale gelmiştir. Aynı ölçme doğruluğunun elde edilmesi için çok boyutlu yaklaşımda gereken madde sayısı tek boyutlu için gereken madde sayısının yarısından daha az olduğu tespit edilmiştir.

DeMars (2005) MTK temelli alt puan kestirim modellerini ikili puanlanan çoktan seçmeli maddelerden oluşan testler kullanarak incelemiştir. Alt testlere ait yetenek kestirimleri Bifaktör model, faktörlerin ilişkili olduğu çok boyutlu model, takviyeli puan yaklaşımı (augmented score approach; Wainer vd., 2001) ve her alt test için ayrı ayrı olmak üzere tek boyutlu model ile elde edilmiştir. Sonuçlar çok boyutlu MTK ve diğer faktörlerdeki bilgiyi dikkate alan Wainer takviyeli puan modelleri arasındaki farkın önemsenmeyecek kadar az

olduğunu göstermektedir. Fakat alt puanların birbirinden bağımsız kestirildiği tek boyutlu MTK modeli diğerlerinden daha fazla yanlılığa ve RMSE değerine sahiptir.

Shin (2007) karma puanlı testler için alt puanların doğru ve güvenilir kestirimini elde etmeyi hedefleyen beş yöntemi, güvenirlik, RMSE, yanlılık, kestirimin standart sapması ve güven aralığı gibi çeşitli kriterlere göre karşılaştırmıştır. Çalışmada ele alınan yöntemler, Bock MTK yöntemi, Yen nesnel performans indeksi (objective performance index; OPI) yöntemi, Wainer'ın ham puanlara dayalı takviye (augmentation) yöntemi, Shin MCMC yordanan puan yöntemi ve doğru oranı ile ham puan yöntemidir. İkili ve çoklu puanlanan maddelerden oluşan veri sırasıyla 3PLM ve genelleştirilmiş kısmi puan modeli kullanılarak simülasyon ile üretilmiştir. Değerlendirme ölçütlerine olan etkisi incelen simülasyon koşulları ise yanıtlayıcı sayısı (250, 500 ve 1,000), alt test uzunluğu (6, 12 ve 18), alt testler arasındaki korelasyon (0,5, 0,8 ve 1,0) ve testte yer alan yapılandırılmış yanıtlı/çoktan seçmeli (CR/MC) maddelerin oranıdır (0, %20 ve %50). Sonuçlar, madde sayısı, alt testler arasındaki korelasyon ve CR/MC oranının güvenirliği etkilediğini göstermektedir. Madde sayısı, boyutlar arasındaki korelasyon ve CR/MC oranı arttıkça güvenirlik artmıştır. RMSE ve yanlılık değerlerini etkileyen koşullar, her bir alt testteki madde sayısı ve boyutlar arasındaki korelasyondur. Madde sayısı arttıkça ve korelasyon yükseldikçe RMSE ve yanlılık değerleri azalmıştır. Karşılaştırılan yöntemler arasında Wainer ve Shin yöntemi daha yüksek güvenirlik ve daha düşük RMSE değeri üretirken Yen ve ham puan yöntemleri görece daha düşük güvenirliğe sahiptir.

de la Torre vd. (2011) MTK'ye dayalı alt puan kestirim yöntemlerinden dördünü karşılaştırmıştır: çok boyutlu MTK, korelasyona dair bilgiyi kullanan takviyeli puanlama (Wainer vd., 2001), üst düzey (higher-order) MTK modeli (de la Torre & Song, 2009) ve alt puanların iyileştirilmesinde testin genelinden elde edilen bilgiyi kullanan OPI (Yen, 1987) yöntemi. Çalışmada, simülasyon veri ile test uzunluğu, alt test sayısı ve alt testler arasındaki korelasyonun alt testler için yetenek kestirimini nasıl etkilediği incelenmiştir. Ayrıca modeller gerçek bir veri seti üzerinde de uygulanmıştır. Simülasyon koşullarının çoğunda korelasyona dayalı olan çok boyutlu MTK, takviyeli puanlama ve üst düzey MTK modelleri birbirine benzer sonuçlar üretmiştir. Modellerin performansı, testteki madde sayısı az, korelasyon ve boyut sayısı yüksek olduğunda daha iyidir. OPI, koşulların birçoğunda diğer modellerden daha kötü performans göstermiştir. Gerçek veriye dayalı analiz sonuçlarının da similasyon çalışmasına ait sonuçlar ile benzerlik gösterdiği tespit edilmiştir.

Longabach (2015) birden çok içeriğe (dinleme, okuma, yazma ve konuşma) dayalı alt alanlardan oluşan İngilizce dil yeterliliğinin ölçülmesini ele almıştır. Çalışmada alt puan kestirim yöntemlerinden doğru sayısına dayalı KTK, tek boyutlu MTK, takviyeli MTK puanlama (Wainer vd., 2001) ve çok boyutlu MTK yöntemlerini incelemiştir. İlk iki model alt alanlara ait yeterliliği ayrı ayrı ele alırken son iki model bu alanlar arasındaki ilişkiyi de dikkate almaktadır. Bu modeller, ikili ve çoklu puanlanan maddelerden oluşan karma-formatta gerçek bir veri seti üzerinde yapılan analizlerle güvenirlik ve ölçme doğruluğu açısından kıyaslanmıştır. Alt alanlardaki yeterliliği ayrı ayrı ölçen KTK ve tek boyutlu MTK modelleri birbirine yakın sonuçlar üretmiştir. Ayrıca bu modeller, alt boyutlar arasındaki ilişkileri dikkate alan diğer iki modelden daha düşük güvenirlik ve ölçme doğruluğu değerleri üretmişlerdir. En güvenilir kestirimler çok boyutlu MTK modeli ile elde edilmiştir.

Yao ve Boughton (2007) karma format bir testte alt puanları kestirmek için ikili puanlanan maddelerde M-3PL ve çoklu puanlanan maddelerde M-2PPC çok boyutlu MTK modellerini kullanmış ve kestirim yöntemi olarak Bayes MCMC algoritması ele alınmıştır. Analizler BMIRT programı ile gerçekleştirilmiştir. Bayes kestirim ile çok boyutlu MTK modeli, doğru sayısı yüzdesi, OPI yöntemi, maksimum olabilirlik yöntemi ile çok boyutlu MTK modeli ve Bayes tek boyutlu MTK modeli ile karşılaştırılmıştır. Bu modeller ile elde edilen alt puanların incelenmesi için değerlendirme kriterleri, kestirim doğruluğu ve sınıflama doğruluğudur. Simülasyon çalışmasında manipüle edilen değişkenler örneklem büyüklüğü (1000, 3000 ve 6000) ve testte yer alan dört alt boyut arasındaki korelasyondur (0, 0,1, 0,3, 0,5, 0,7 ve 0,9). Çalışma sonuçları, Bayes MCMC algoritması ile çok boyutlu MTK modelinin, kestirim doğruluğu ve sınıflama performansı açısından diğerlerinden daha iyi olduğunu göstermektedir. Korelasyon arttıkça kestirimlerin ve sınıflamanın hataları azalmıştır. Korelasyon düşük olduğunda Bayes kestirim ile elde edilen sonuçlar, maksimum olabilirlik çok boyutlu MTK ve tek boyutlu MTK ile elde edilen sonuçlarla benzerdir. Korelasyon yüksek olduğunda ise OPI yöntemi hem parametre kestirim doğruluğu hem de sınıflama doğruluğu açısından Bayes çok boyutlu MTK ile yakın sonuçlar ortaya koymaktadır.

Fan (2016) alt puanların güvenirliğini ve sınıflama performanslarını incelediği çalışmasında beş yetenek kestirim modelini karşılaştırmıştır: Ham puan, tek boyutlu MTK, takviyeli ham puan, takviyeli tek boyutlu MTK, çok boyutlu MTK. Bu karşılaştırma için bir simülasyon çalışması tasarlanmış ve sınıflama tutarlılığı ve doğruluğunu etkileyen faktörler (alt boyutlar arası korelasyon, alt test uzunluğu, alt test sayısı ve kesme puanının seçimi) incelenmiştir.

Sonuçlar, alt puanların güvenirliği ve sınıflamasına ilişkin kestirimlerin incelenen faktörlerin bir fonksiyonu olduğunu göstermiştir. Alt testlerde yer alan madde sayısı 20 ya da daha fazla olduğunda güvenirlik açısından modeller arasındaki fark düşüktür. Madde sayısı az olduğunda ise boyutlar arasındaki korelasyonu dikkate alan modeller (çok boyutlu MTK, takviyeli MTK ve takviyeli ham puan) daha iyi performans göstermiştir. Alt puanlara ait sınıflama performanslarına ilişkin sonuçlar madde sayısı ve güvenirlik arasındaki ilişkiye benzer sonuçlar üretmiştir. Puan dağılımının ortasında tek bir kesme puanı olduğunda beş madde için bütün modellere ilişkin sınıflama tutarlılığı ve doğruluğu kestirimleri 0,70 civarındadır. Kesme puanı dağılımın merkezine uzak olduğunda ise sınıflama tutarlılığı ve doğruluğu daha yüksektir. Bu durum, beş maddeden oluşan bir alt test puanına ait güvenirlik düşük olsa bile kesme puanının yerine göre bu puanlar ile doğru sınıflama yapılabildiğini göstermektedir.

de la Torre ve Song (2009) çok boyutlu bir test yapısında toplam puan ve alt puanların birlikte kestirilmesini amacıyla üst düzey MTK modelini önermiştir. Önerilen model çeşitli koşullar altında simülasyon çalışması ile detaylı olarak incelenmiş ve gerçek bir veri seti üzerinde de uygulanmıştır. Yetenek kestirimine etkisi olup olmadığı incelenen simülasyon koşulları test uzunluğu (10, 20 ve 30), alt test sayısı (2 ve 5), alt testler arasındaki korelasyon (0, 0,4, 0,7 ve 0,9) ve örneklem büyüklüğüdür (1000, 2000 ve 4000). Üst düzey MTK model parametreleri, MCMC kestirim yöntemi kullanılarak elde edilmiştir. Ayrıca önerilen model ile elde edilen toplam test ve alt testlere ilişkin yetenek parametreleri, tek boyutlu MTK modeli ile elde edilen yetenek parametreleri ile karşılaştırılmıştır. Sonuçlar, alt boyutlar arasındaki korelasyon yüksek olduğunda tek boyutlu MTK ve üst düzey MTK ile elde edilen genel yeteneklerin benzer olduğunu göstermiştir. Yine de birden çok tek boyutlu MTK uygulamasının yapılmasının yetersiz olduğu ve özellikle alt testlerde madde sayısı az olduğunda güvenilir olmayan kestirimler yaptığı belirtilmektedir. Alt boyutlar arasındaki korelasyon düşük olduğunda ise üst düzey MTK ile elde edilen genel yetenek kestirimlerine ait yanlılık daha düşüktür. Alt testler açısından ise üst düzey MTK kullanıldığında daha doğru yetenek kestirimleri elde edilmiştir. Kestirim doğruluğu alt test sayısı ve boyutlar arasındaki korelasyon arttıkça artmıştır.

Yao (2010) toplam ve alt puanlar raporlamak amacıyla dört modelin performansını incelemiştir. Çalışmada ele alınan modeller, tek boyutlu MTK, çok boyutlu MTK, üst düzey MTK ve genel Bifaktör modeldir. Çok boyutlu MTK modeli için toplam puan hesaplanması amacıyla maksimum bilgi yöntemi önerilmiştir. Yetenek kestirimlerine etkisi olduğu

düşünülen örneklem büyüklüğü (500, 1000 ve 2000), alt boyutlar arasındaki korelasyon (0, 0,3, 0,5, 0,7 ve 0,9) ve test uzunluğu (20, 32, 48 ve 60) koşulları ile bir simülasyon çalışması gerçekleştirilmiş ve ek olarak gerçek veri ile uygulama yapılmıştır. İkili ve çoklu puanlanan maddelerden oluşan karma-format test yapısı ele alınmıştır. Genel ve alt yeteneklerin kestiriminin değerlendirilmesindeki kriterler RMSE, yanlılık ve güvenirliktir. Genel yetenek kestiriminde, korelasyon 0,8'den yüksek olduğunda elde edilen sonuçlar dört model için de birbirine yakındır. Fakat korelasyon 0,0 ya da çok düşük olduğunda çok boyutlu MTK maksimum bilgi yöntemi ve üst düzey MTK modeli, tek boyutlu MTK ve Bifaktör modelden daha iyi performans göstermiştir. Alt puanlar açısından da çok boyutlu MTK ve üst düzey MTK modelleri ile birbirine yakın ve diğer modellerden daha iyi sonuçlar elde edilmiştir. Boyutlar arasındaki korelasyon arttıkça Bifaktör model ile elde edilen alt puan kestirimlerine ait hata artmış ve güvenirlik düşmüştür.

Liu vd. (2019) toplam ve alt test puanlarının raporlanmasında Bifaktör modelin kullanılmasını incelemiştir. Çalışmada, genel ve alt faktörlerin Bifaktör model altında farklı ağırlıklandırmalar ile altı farklı yöntem önerilmiş ve Yao'nun maksimum bilgi yöntemi ile toplam puanın hesaplandığı çok boyutlu MTK modeli ile karşılaştırılmıştır. Genel Bifaktör modelde alt faktörler birbirleri ve genel faktör ile ilişkisizdir. Önerilen diğer beş yöntemde Bifaktör model altında bu faktörlere ait puanların hesaplanmasında faktörler arası ilişkiyi dikkate alan beş farklı ağırlıklandırma yer almaktadır. Gerçek ve simülasyon verinin birlikte ele alındığı çalışmada yer alan simülasyon koşulları, alt test uzunluğu (20, 40 ve 60), alt test sayısı (2 ve 5), alt testler arası korelasyon (0, 0,3, 0,5 ve 0,7) ve örneklem büyüklüğüdür (1000 ve 5000). Analizler MCMC kestirim yöntemi ile BMIRT programı kullanılarak gerçekleştirilmiştir. Toplam test ve alt testlere ait yetenek parametreleri kestirilişmiş ve bunlara ilişkin yanlılık, RMSE ve güvenirlik değerleri hesaplanmıştır. Sonuçlar, koşulların birçoğunda ağırlıklandırmalarda ilgili faktörlerdeki maddelere ait ayırt edicilikleri dikkate alan Bifaktör-M4 ve Bifaktör-M6 yöntemleri ile daha doğru ve güvenilir toplam ve alt puanlar elde edildiğini göstermektedir. Bifaktör-M4 ve Bifaktör-M6 yöntemleri ile çok boyutlu MTK sonuçları birbirlerine benzerlik göstermektedir. Faktörler arası ilişki tanımlamayan Bifaktör genel model (Bifaktör-M1) yetenek kestiriminde hem toplam hem de alt test bazında en kötü performansı göstermiştir. Bifaktör-M1 modeli ile genel yetenek kestiriminin hatası, boyutlar arası korelasyon ve boyut sayısı azaldıkça artış göstermiştir. Alt testlere ait yetenek kestiriminde ise durum tam tersi olup korelasyon arttıkça hata artmıştır.

BÖLÜM 3

YÖNTEM

Bu bölümde araştırmanın türü, simülasyon çalışması ve gerçek veri analizi hakkında bilgi sunulmaktadır.

3.1. Araştırmanın Türü

Bu çalışmada, ÇBMTK, ÜDMTK ve Bifaktör modellerine dayalı olarak kestirimin doğruluğu, güvenirlik ve sınıflama performansı karşılaştırmalı olarak incelenmiştir. Simülasyon çalışması ile yöntemlerin, boyutlar arası korelasyon, çoklu puanlanan maddelerin toplam madde sayısına oranı ve test güçlüğü faktörlerine göre değerlendirme kriterleri üzerine etkileri incelenmiştir. Ayrıca modellerin göreceli performansı gerçek bir veri seti üzerinde de test edilmiştir. Bu doğrultuda bu çalışma, betimsel düzeyde ve simülatif bir çalışmadır.

3.2. Simülasyon Çalışması

Bu kısımda simülasyon koşuları, verilerin üretilmesi, madde ve yetenek parametrelerinin kestirilmesi – model kestirimleri, simülasyon çalışması için değerlendirme kriterleri ve analizler açıklanmıştır.

3.2.1. Simülasyon Koşulları

3.2.1.1. Sabit Koşullar

Bu çalışmada gerçekleştirilen simülasyon sürecinde bazı koşullar sabit olarak ele alınmıştır. Sabit olarak ele alınan koşullar, örneklem büyüklüğü, boyut sayısı, madde sayısı ve yetenek parametresinin dağılımıdır. Örneklem büyüklüğünün değişken olarak ele alındığı birçok çalışma bulunmaktadır (örn. de la Torre & Hong, 2010; Shin, 2007; Yao & Boughton, 2007). Örneklem büyüklüğü 2000 ve üzeri olması ÇBMTK kestirimleri için önerilmektedir (Akerman, 1994; Reckase, 1995). Bu durum ÇBMTK yöntemlerinin geniş ölçekli testlerin incelenmesi için daha uygun olduğunu göstermektedir (Lin, 2008). Bazı çalışmalarda ise tek ve çok boyutlu yetenek kestiriminde örneklem büyüklüğünün etkisinin az olduğu bulunmuştur (de la Torre & Patz, 2005; de la Torre & Song, 2009). Bu sebeplerle, kararlı kestirimlerin elde edileceği varsayılan sabit bir örneklem büyüklüğü, N=5000, simülasyon çalışmasının bütün koşulları için kullanılmıştır.

Boyut sayısının değişken olarak ele alındığı çalışmalarda 2 ilâ 7 arasında değişen sayılarda çok boyutlu test verisinin üretildiği görülmektedir (Bulut 2003; de la Torre & Patz, 2005; de la Torre vd., 2011; Lee, 2012; Yavuz, 2014). Diğer bazı çalışmalarda ise boyut sayısı 2 (Kuo & Sheng, 2016; Shin, 2007) ya da 4 olacak şekilde sabit tutulmuştur (Chin, 2011; Yao, 2010; Yao & Boughton, 2007). Benzer çekilde bu çalışmada üretilen veri dört boyuttan oluşmaktadır. Bunun sebebi, ülkemizde yapılan geniş ölçekli sınavların bir çoğunun dört alt alandan oluşması, İngilizce becerisinin ölçüldüğü ulusal ya da uluslar arası sınavların dile ait dört temel beceriyi ölçmeye çalışması ve ek olarak bu tez kapsamında kullanılan gerçek verinin dört boyutlu bir test olmasıdır. Örneğin Türkiye'de uygulanan liselere geçiş ve yükseköğretime geçiş sınavlarında Türkçe, Matematik, Sosyal Bilgiler ve Fen Bilgisi olmak üzere dört temel alan ele alınmaktadır. Uluslararası uygulanan İngilizce yeterlilik sınavı TOEFL ise okuma, yazma, dinleme ve yazma alt alanlarından oluşmaktadır.

Yapılan çalışmalarda farklı alt test uzunlukları incelenmiştir. Örneğin, Yao (2010) 20-32-48-60 olmak üzere farklı test uzunluklarını, Yavuz (2014) ise 15, 30 ve 60 maddeden oluşan test uzunluklarını incelemiştir. Bulut (2013) ise kısa, orta ve uzun olarak tanımladığı 10, 20 ve 40 maddeden oluşan alt testleri ele almıştır. Chin (2011) madde sayısının güvenirliğe direkt etki ettiğini ve genellikle, alt test ne kadar uzunsa güvenirliğin de yüksek olacağını belirtmiştir. Böylece alt testler ile sunulan bilgi daha güvenilir olmaktadır. Öte yandan madde sayısının değişken olarak ele alınmadığı ve sabit olarak ele alındığı çalışmalar da

bulunmaktadır. Örneğin, Jang (2014) boyut sayısının iki olduğu çalışmasında herbir boyuttaki madde sayısını 30 olarak ele almıştır. Benzer şekilde bu çalışmada da madde sayısı sabit tutulmuş ve dört alt testin her biri için 30 olarak belirlenmiştir.

Alt testlere ait yetenek parametreleri üretilirken farklı dağılım özelliklerinin dikkate alındığı çalışmalar bulunmakla birlikte, birçok çalışmada (de la Torre & Patz, 2005; Shin, 2007; Yao & Boughton, 2007; Yao & Schwarz, 2006;) yetenek parametreleri çok değişkenli normal dağılıma göre üretilmiştir. Bu çalışma kapsamında gerçek yetenek parametrelerinin üretiminde çok değişkenli normal dağılım, $\theta_i \sim MVN(0, \Sigma)$, kullanılmıştır. Ayrıca alt testler arasındaki ikili korelasyon değeri aynı olacak şekilde belirlenmiştir. Bir başka ifadeyle, 1. ve 2. alt test arasındaki korelasyon (r), 1-3, 1-4, 2-3, 2-4 ve 3-4 alt testleri arasında da aynıdır. 4 boyutlu varyans-kovaryans matrisi Eşitlik 34'te gösterilmektedir.

$$\Sigma = \begin{bmatrix} 1 & r & r & r \\ r & 1 & r & r \\ r & r & 1 & r \\ r & r & r & 1 \end{bmatrix} \qquad \text{(Eşitlik 34)}$$

Eşitlik 34'te görüldüğü üzere korelasyon değerleri aynıdır. Korelasyon büyüklüğü ise değişken olarak ele alınmış ve bir sonraki kısımda açıklanmştır.

3.2.1.2. Manipüle Edilen (Değişimlenen) Koşullar

Bu çalışmada değişken olarak ele alınan koşullar boyutlar arası korelasyon, testteki çoklu puanlanan madde yüzdesi ve test güçlüğüdür. Bu üç koşul birlikte (korelasyon: 3, testteki çoklu puanlanan madde yüzdesi: 4 ve test güçlüğü: 5) toplamda 60 çapraz koşul meydana getirmişlerdir. Değişkenler, aşağıda ayrıntılı olarak açıklanmaktadır.

Boyutlar arası korelasyon. Gerçek yetenek parametreleri arasındaki korelasyon değeri, bağımsız değişken olarak çalışmaya dahil edilmiştir. Alt testlerden elde edilen yetenek kestirimleri arasındaki korelasyon değeri yöntemlerin performansını etkilemektedir (Chin, 2011). ÇBMTK ve ÜDMTK yöntemlerinden farklı olarak, Bifaktör modelde boyutların birbiri ile ve genel faktör ile korelasyonu olmadığı varsayımı vardır. Bu varsayımın ihlal edildiği durumlarda kestirimin kalitesinin değerlendirilebilmesi için bu model için de farklı korelasyon değerleri ele alınmıştır. Jang (2014) de Bifaktör model ile analiz yaptığı çalışmasında 0,2 ve 0,8 korelasyon değerlerini incelemiştir. Literatür incelendiğinde farklı kestirim yöntemleri ile yapılan çalışmalarda birçok korelasyon değerinin incelendiği

görülmektedir. Örneğin Finch (2010) boyutlar arası korelasyonda dört farklı değeri kullanmıştır: ilişkinin olmadığı 0, düşük düzey korelasyon için 0,3, orta düzey için 0,5 ve nispeten yüksek korelasyon için 0,8 korelasyon. Bulut (2013) düşük, orta ve yüksek olarak ifade ettiği 0,3, 0,5 ve 0,8 korelasyon değerlerini incelemiştir. Chin (2011) eğitimde kullanılan ölçmelerde sıklıkla görülen değerler olduğunu ifade ederek 0,5, 0,7 ve 0,9 korelasyon değerlerini ele almıştır. Korelasyon değeri manipüle edilen diğer bazı çalışmalarda ise 0,2, 0,5 ve 0,8 korelasyon değerleri, düşük orta ve yüksek ilişkinin göstergesi olarak kullanılmıştır (Kuo, 2015; Kuo & Sheng, 2016; Li, 2018). Benzer şekilde bu çalışmada veri üretim aşamasında dört alt test için gerçek yetenek parametreleri arasındaki korelasyon değerleri düşük, orta ve yüksek korelasyon olacak şekilde sırasıyla 0,2, 0,5 ve 0,8 olarak belirlenmiştir.

Testteki çoklu puanlanan madde yüzdesi. Bu çalışmada toplam madde sayısı sabit tutularak testin içindeki çoklu puanlanan maddelerin yüzdesi değişken olarak ele alınmıştır. Shin (2007) çalışmasında, karma format testlerde alt puan hesaplama yöntemlerini karşılaştırdığı çalışmasında çoklu puanlanan maddelerin ikili puanlanan maddelere oranını değişken olarak ele almış ve bu durumun güvenirliğe etkisini incelemiştir. Bu oran için koşullar %0, %20 ve %50 olarak belirlenmiştir. Bu çalışmada da karma format bir test formu tasarlanmış ve 120 maddeden oluşan testin içerisinde farklı sayılarda çoklu puanlanan madde yer almıştır. Çoklu puanlanan maddelerin testteki toplam madde sayısı içindeki yüzdesi koşulu için ele alınan durumlar sırasıyla %5, %10, %25 ve %50'dir. Değişkene ait kategori %5 olduğunda 120 maddeden 6 tanesi, %10 olduğunda ise 12 tanesi çoklu puanlanan maddedir. Bu maddelerin hepsi son 30 madde (4. alt test) içerisinde yer almaktadır. Çoklu puanlanan madde yüzdesi %25 olduğunda, son 30 maddenin hepsi yani 4. alt testin tamamı ve %50 olduğunda ise son 60 madde yani 3. ve 4. alt test maddelerinin tamamı çoklu puanlanmaktadır.

Test güçlüğü. Yapılan çalışmalarda güçlük parametreleri genellikle standart normal dağılımda olacak şekilde üretilmektedir. Örneğin, Finch (2010), Chin (2011) ve Woods ve Thissen (2006) güçlük parametrelerini standart normal dağılımda üretmişlerdir. Li ve Lissitz (2012) güçlük parametrelerini üretirken ortalamaları -0,5 ile 0,5 arasında ve standart sapması 1 olacak şekilde normal dağılım kullanmışlardır. Az sayıdaki diğer çalışmalarda ise veri üretiminde güçlük parametreleri farklılaştırılarak test güçlüğünün farklılığı koşulu olarak ele alınmıştır. Örneğin, Embretson (2007, s. 63) kolay, orta güçlükte ve zor olarak ifade ettiği test durumları için sırasıyla ortalamaları 0, 1, 2 ve standart sapmaları 1 olacak şekilde normal

dağılım sergileyen güçlük parametrelerine sahip maddeler üretmiştir. Benzer şekilde, Smith (2016) de ortalama güçlüğü 0 olan kolay bir test, ortalama güçlüğü 1 olan orta düzey bir test ve ortalama güçlüğü 2 olan zor bir test oluşturmuştur. Morse, Johanson ve Griffeth (2012) kolay, orta ve zor test durumları için madde güçlüklerinin standart sapması 1 ve ortalamaları sırasıyla -1,5, 0 ve 1,5 olacak şekilde bir normal dağılım şeklinde tanımlamıştır. Paek (2015) de ortalama güçlük değerlerini -1,5, 0 ve 1,5 olarak belirlemiş ve ortalamadaki bu artışın, güçlük parametreleri dağılımını, sırasıyla pozitif kayışlı, normal ve negatif kayışlı yaptığını ifade etmektedir. Bu çalışmada da test güçlüğü değişken olarak ele alınmıştır. Madde parametrelerine ilişkin güçlük değerleri standart sapması 1 ve ortalaması sırasıyla -2, -1, 0, 1 ve 2 olan normal dağılıma sahip beş farklı durumda olacak şekilde üretilmiştir. Bu testlerin sırasıyla çok kolay, kolay, orta güçlükte, zor, çok zor olduğu varsayılmıştır.

3.2.1.3. Replikasyon Sayısı

Harwell, Stone, Hsu ve Kirisci (1996) MTK analizi içeren simülasyon çalışmalarında replikasyon sayısını inceledikleri çalışma sonucunda, daha kararlı sonuçlar elde edebilmek için bu sayının 25'ten az olmaması gerektiğini ifade etmişlerdir. Simülasyon çalışması içeren alan yazın incelendiğinde farklı sayılarda replikasyon yapıldığı görülmektedir (örn; 20 replikasyon –Yao, 2010; Yao & Boughton, 2007; 25 replikasyon –de la Torre & Hong, 2010; Şahin & Yıldırım, 2018; 50 replikasyon – Fu & Qu, 2010; Jang, 2014; Levy, Mislevy & Sinharay, 2009; Soysal, 2017; 100 replikasyon – Ames, 2015, Chin, 2011; Li & Lissitz, 2012; Sass vd., 2008; Sinharay, 2006; Toribio & Albert, 2011). Önceki çalışmalar ışığında ve daha kararlı sonuçlar elde etmek amacıyla bu çalışmada tasarlanan çapraz koşulların her biri için 100 replikasyon gerçekleştirilmiştir.

3.2.2. Verilerin Üretilmesi

Bölüm 3.2.1.2'de açıklanan simülasyon koşulları veri üretim aşamasında manipüle edilmiştir. Bu çalışmada veri üretimi için 60 çapraz koşul bulunmaktadır: 3 korelasyon x 4 testteki çoklu puanlanan madde yüzdesi x 5 test güçlüğü. Her bir çapraz koşul için 100 replikasyon kullanılmış ve toplamda 6000 veri üretilmiştir. Kişi sayısı bütün koşullarda aynı olup 5000'dir. Her biri 30 maddeden oluşan 4 boyutlu testin basit yapılı olduğu varsayılmıştır. Her alt test kendi içinde tek boyutlu iken testin geneli çok boyutludur. Bir

başka ifadeyle, her madde yalnızca bir boyut ile ilişkilidir. 6000 madde yanıt örüntüsünün oluşturulmasında izlenen aşamalar aşağıda maddeler halinde sunulmuştur.

1. Önce madde parametreleri üretilmiştir. Madde ayırt edicilik parametreleri 0,5 - 2,5 aralığında olacak şekilde (örn; Ackerman, 1988; Bulut, 2013; Jang, 2014; Lin, 2008) tek biçimli olarak üretilmiştir (a_i ~ U [0.5, 2.5]). Güçlük parametreleri Bölüm 3.2.1.2'de açıklandığı üzere ortalama değerleri farklı olacak şekilde ($b_i \sim N$ (-2, 1), $b_i \sim N$ (-1, 1), $b_i \sim N$ (0, 1), $b_i \sim N$ (1, 1), $b_i \sim N$ (2, 1)) normal dağılım ile üretilmiştir. Şans parametresi de beta dağılım (c_i ~ beta (6, 16)) ile elde edilmiştir (Yao, 2010). Madde parametrelerinin üretiminde WinGen3 programı (Han & Hambleton, 2007) kullanılmış olup madde ve yetenek ilişkisinin tanımlanmasında ikili puanlanan maddeler için 3 parametreli lojistik model (3PL) ve çoklu puanlanan maddeler için iki parametreli kısmi puan modeli (2PPC) tanımlanmıştır. Çoklu puanlanan maddeler 4 kategoriden oluşmaktadır (0-1-2-3).
2. Dört alt teste ilişkin gerçek yetenek parametreleri, önceden belirlenen bir varyans-kovaryans matris ile çok değişkenli normal dağılımdan, $\theta_i \sim MVN(0, \Sigma)$, üretilmiştir. Her bir boyut için yetenek parametrelerinin ortalaması 0 ve varyansı 1'dir. Simülasyon koşulları bölümünde açıklanan boyutlar arası korelasyona (r=0,2, r=0,5 ve r=0,8) dayalı olan korelasyon matrisleri aşağıda gösterilmektedir.

$$\Sigma = \begin{bmatrix} 1 & 0,2 & 0,2 & 0,2 \\ 0,2 & 1 & 0,2 & 0,2 \\ 0,2 & 0,2 & 1 & 0,2 \\ 0,2 & 0,2 & 0,2 & 1 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0,5 & 0,5 & 0,5 \\ 0,5 & 1 & 0,5 & 0,5 \\ 0,5 & 0,5 & 1 & 0,5 \\ 0,5 & 0,5 & 0,5 & 1 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0,8 & 0,8 & 0,8 \\ 0,8 & 1 & 0,8 & 0,8 \\ 0,8 & 0,8 & 1 & 0,8 \\ 0,8 & 0,8 & 0,8 & 1 \end{bmatrix}$$

SimuMIRT (Yao, 2003) ile üretilen gerçek yetenek parametreleri, korelasyon koşulları için ayrı ayrı olacak şekilde elde edilmiştir. Toplam puan için gerçek yetenek kestirimleri, burada dört alt boyut için üretilen gerçek yetenek parametreleri kullanılarak BMIRT (Bayesian Multivariate Item Response Theory – Bayes Çok Değişkenli Madde Tepki Kuramı; Yao, 2003; Yao, 2016) programında maksimum bilgi yöntemiyle (Yao, 2010) elde edilmiştir. Yao (2010) ve Liu vd. (2019) tarafından gerçekleştirilen çalışmalarda da toplam puan için gerçek yetenek kestirimleri aynı yolla elde edilmiştir.

3. Madde-yanıt örüntüleri SimuMIRT (Yao, 2003) ile üretilmiştir. Programın betik dosyasına kişi sayısı, madde sayısı, boyut sayısı, WinGen3 ile elde edilmiş madde

parametreleri, yetenek parametrelerine ait ortalama, varyans ve varyans-kovaryans değerleri tanımlanmıştır. Her bir çapraz koşuldaki 100 replikasyon için sadece rastgele sayılar (random seed) değiştirilmiş 100 kontrol dosyası oluşturularak çalıştırılmıştır. Böylece, 5000 kişi, 120 madde ve 4 alt testten oluşan 60 çapraz koşula (3x4x5) ait 6000 veri seti üretilmiştir. SimuMIRT için bu çalışma kapsamında oluşturulan örnek bir betik dosyası Ek 1'de sunulmuştur.

3.2.3. Madde ve Yetenek Parametrelerinin Kestirilmesi – Model Kestirimleri

Veri üretim aşamasında üretilen 6000 veri seti ÇBMTK, ÜDMTK ve Bifaktör yöntemleri ile ayrı ayrı analiz edilmiştir. Madde ve yetenek parametrelerinin kestiriminde BMIRT programı (Yao, 2003) kullanılmıştır. BMIRT, MCMC (Markov Chain Monte Carlo) yöntemine dayalı olarak parametre kestirimi yapan Bayes bir yaklaşım ile analiz yapmaktadır. MCMC algoritmasının (Patz & Junker, 1999a, 1999b) geliştirilmesi ile çok boyutlu ve çok kategorili veri gibi daha karmaşık durumların üstesinden gelinebilmektedir (örn; Edwards, 2010; Yao, 2010). MCMC algoritması, ÇBMTK için popüler bir yöntem olarak öne çıkmaktadır; karmaşık ve çok boyutlu modellerde kullanılabilirliğinden dolayı MCMC kestirimi ile Bayes yaklaşım, ÇBMTK modellerinde sıklıkla kullanılmaktadır (örn; Ames, 2015; Chin, 2011; de la Torre & Patz, 2005; Kim, 2008; Kuo, 2015; Lee, 2012; Yao & Schwarz, 2006). BMIRT programı, madde, yetenek ve evren dağılımına ait parametreleri MCMC yöntemi altındaki Metropolis-Hastings algoritmasını kullanarak kestirir.

BMIRT programının bir diğer öne çıkan özelliği, hem ikili hem çoklu puanlanan veri için kestirim yapabilmesidir. Program çok boyutlu 3 parametreli lojistik model (M-3PL) ve çok boyutlu iki parametreli kısmi puan (M-2PPC) modelini aynı anda uygulayabilmektedir. Bu sayede karma format bir yapıya sahip olan veriler için kullanılabilmektedir. Ayrıca program kullanımı ücretsizdir.

MCMC kestirimi için iterasyon sayısı ve öncül (priors) değerlerinin belirlenmesi gerekmektedir. Alan yazın incelendiğinde farklı sayılarda iterasyon ve hazırlık (burn-in) değerlerinin kullanıldığı görülmektedir. Kuo (2015) 5000'i hazırlık olmak üzere 10000 iterasyon tanımlarken, Liu vd. (2019) 2000 hazırlık ve 5000 iterasyon tanımlamışlardır. De la Torre ve Patz (2005) ise 2000 hazırlık ve toplamda 10000 iterasyon ile analiz yapmışlardır. Yao ve Schwarz (2006) karma format veri yapısı kullandıkları çalışmalarında tanımladıkları iterasyon sayısı 15000 olmak üzere ilk 3000'i hazırlık aşaması için atanmıştır. Benzer

şekilde bu çalışmada da iterasyon sayısı 15000 ve hazırlık sayısı ise 3000 olarak belirlenmiştir. Yakınsamayı sağlamak ve daha doğru parametre kestirimleri elde etmek için iterasyon sayısı yüksek tutulmuştur. Madde parametrelerine ait öncüller için Yao ve Schwarz (2006) tarafından önerilen değerler kullanılmıştır: $a_i \sim logN$ (1,5, 1,5); $b_i \sim N$ (0, 1); $c_i \sim$ beta (100, 400). Yetenek parametresi için çok değişkenli standart normal dağılım tanımlanmıştır. Boyutlar arası ilişki için ise veri üretim aşamasında ele alınan korelasyon değerleri girilmiştir. Sadece Bifaktör model için korelasyon öncül değeri boyutlar arasındaki diklik varsayımından ötürü 0 olarak girilmiştir. BMIRT programı ile gerçekleştirilen analizler için örnek bir kontrol dosyası Ek 2'de sunulmuştur.

MCMC algoritmasını kullanan programlar uzun çalışma sürelerine ihtiyaç duyarlar (Chin, 2011). BMIRT ve flexMIRT programlarının karşılaştırıldığı bir çalışmada, MCMC yöntemini kullanmasından ve bu yöntemin yüksek sayıda iterasyon ile daha kararlı sonuçlar üretmesinden dolayı kestirim süresinin flexMIRT programına göre daha uzun olduğu belirtilmektedir (Yavuz & Hambleton, 2017). Bu çalışma kapsamında yer alan tek bir analiz için araştırmacının kişisel bilgisayarında gereken süre 90 dakikadır. Gerçek veri analizlerinin yapıldığı bilgisayar özellikleri Lenovo Yoga 710 Intel(R) Core™ i7-7500U CPU @ 2,70GHz 8 GB RAM ve 64 bit işletim sistemi şeklindedir. Simülasyon veri analizleri ise toplam sürenin (18000 analiz x 90 dk.) oldukça fazla olması sebebiyle, Türk Ulusal Bilim e-Altyapısı (TRUBA) kaynakları kullanılarak gerçekleştirilmiştir. TÜBİTAK ULAKBİM bünyesinde yer alan Yüksek Başarımlı ve Grid Hesaplama Merkezi araştırmacılara yüksek performanslı hesaplama ve veri depolama alanı sağlayan ulusal bir yapılanmadır (TRUBA, 2018; ayrıntılı bilgi için: www.truba.gov.tr). TRUBA sistemi Linux işletim sistemi ile çalışmaktadır. Bu sebeple BMIRT bat dosyaları Linux'ta çalışacak şekilde sh betik dosyalarına çevrilmiştir. Bu çalışma için hazırlanmış, analizlerin Linux sisteminde gerçekleştirilmesini sağlayan Python dili ile yazılmış betik dosyası Ek 3'te sunulmuştur.

Bu çalışma kapsamında toplam test ve alt testlere ilişkin yetenek kestiriminde kullanılan modeller Bölüm 2'de ayrıntılı olarak açıklanmıştır. Aşağıda, bu modeller ile çalışma verisi üzerinde yapılanlar kısaca özetlenmiştir:

- ÇBMTK: Dört alt test için bireylere ait yetenek puanlarının hesaplanmasında basit yapılı ÇBMTK analizi kullanılmıştır. Testin bütününe ilişkin yetenek puanı ise dört alt teste ait puanların kullanıldığı maksimum bilgi yöntemiyle elde edilmiştir (Yao, 2010; Yao, 2012a).

- ÜDMTK: Bu modelde toplam puan ve alt puanlar arasında lineer bir ilişki olduğu varsayılır. Böylece toplam ve alt boyutlara ait yetenek parametreleri eş zamanlı olarak kestirilmiştir.
- Bifaktör Model: Bifaktör modelde alt boyutlar birbiri ile ve genel boyut ile birbirlerine diktir. Bir başka ifadeyle boyutlar arasında korelasyon olmadığı varsayımı vardır. Bu model ile beş yetenek kestirilmiştir. İlki genel boyuta ve diğer dört yetenek parametresi de alt boyutlara aittir. ÜDMTK ile benzer şekilde burada da toplam ve alt puanlara ait kestirimler aynı çıktıda yer almaktadır.

3.2.4. Simülasyon Çalışması İçin Değerlendirme Kriterleri

SimuMIRT ile 60 çapraz koşulda üretilen veriler, BMIRT programı kullanılarak ÇBMTK, ÜDMTK ve Bifaktör modellere göre analiz edilmiştir. Analiz sonucunda elde edilen toplam ve alt testlere ait yetenek parametrelerinin doğruluğu, RMSE (root mean square error), gerçek parametreler ile aralarındaki korelasyon ve sınıflama tutarlılığı ve doğruluğu açısından incelenmiştir. RMSE değerleri Eşitlik 35 kullanılarak R 4.0.3 (R Development Core Team, 2016) ile elde edilmiştir.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\theta_i - \hat{\theta}_i)^2} \qquad \text{(Eşitlik 35)}$$

Burada $\hat{\theta}_i$ i. birey için ilgili kestirim yöntemleri ile elde edilen toplam puan ya da alt test puanı, θ_i ise gerçek puana işaret etmektedir. N kişi sayısı olup bu çalışmada sabit tutulduğu için bütün koşular için N=5000 yanıtlayıcı bulunmaktadır. Gerçek alt test puanları her bir boyut için üretilen yetenek parametreleridir. Gerçek toplam test puanı ise alt testler için üretilen gerçek yetenek parametreleri kullanılarak maksimum bilgi yöntemi ile hesaplanan gerçek toplam puan yetenek kestirimleridir.

Bir diğer değerlendirme kriteri gerçek yetenek parametreleri ile kestirilen yetenek parametreleri arasındaki korelasyondur [$cor(\theta, \hat{\theta})$]. Toplam puan ve alt test puanlarına ait gerçek ve kestirilen yetenek parametreleri arasında Pearson korelasyon katsayısı hesaplanmıştır. Hesaplanan bu değerler, güvenirliğin bir göstergesi olarak ele alınmıştır. Bu kriter, tezin diğer bölümlerinde güvenirlik olarak isimlendirilmiştir. RMSE ile benzer

şekilde her bir replikasyon için korelasyon değerleri R (R Development Core Team, 2016) ile elde edilmiştir.

Sınıflama performansının incelenmesi için ise Rudner (2001, 2005) sınıflama tutarlılığı ve doğruluğu indeksleri hesaplanmıştır. Geçti-kaldı, yeterli- yetersiz, başarılı-başarısız şeklinde ikili sınıflama yapılmıştır. Toplam test ve alt testlerin her biri için gerçek yetenek parametrelerinde grubun %65'inin geçeceği/başarılı sayılacağı nokta kesme puanı olarak belirlenmiştir. Sınıflama tutarlılığı ve doğruluğu indeksleri her bir replikasyon için R *cacIRT* (Lathrop, 2015) paketi kullanılarak elde edilmiştir. Her bir koşul altında elde edilen sınıflama tutarlılığı ve doğruluğu değerleri arasındaki korelasyon 1,00'a oldukça yakın olduğu için değişkenlerin karşılaştırılmasında yapılan analizlerde sadece sınıflama doğruluğu ele alınmıştır.

3.2.5. Veri Analizi

Manipüle edilen simülasyon koşullarının ve kestirimde kullanılan analiz modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) RMSE, güvenirlik ve sınıflama doğruluğu üzerindeki etkisini incelemek için toplam puana ilişkin incelemelerde faktöriyel karma ANOVA (Gamst, Meyers & Guarina, 2008, s.391) yapılmıştır. ANOVA yapılmasının amacı, koşulların ve kestirim yöntemlerinin karşılaştırılmasında istatistiksel çıkarım yapabilmektir. Analize ilişkin normallik ve küresellik varsayımları incelenmiştir. Bütün değerlendirme kriterlerinde her bir çapraz koşul için yapılan incelemelerde neredeyse hepsinin normal dağılım sergilediği görülmüştür. Kirk (2013, s. 99) F istatistiğinin, normallik varsayımı ihlallerine karşı güçlü olduğunu belirtmektedir. Mauchly test ile değerlendirilen küresellik varsayımı ise sağlanamamıştır. Ancak Gamst vd. (2008, s.269) bu varsayımın karşılanamadığını durumlarda Greenhouse-Geisser ya da Huynh-Feldt düzeltmesinin yapılabileceğini belirtmektedir. Bu çalışmada da Huynh-Feldt düzeltmesi uygulanmıştır.

RMSE, güvenirlik ve sınıflama doğruluğu değerlendirme kriterlerinin her biri için ayrı analizler yapılmıştır. Faktöriyel karma ANOVA'da simülasyon koşulları (boyutlar arası korelasyon, çoklu puanlanan madde yüzdesi ve test güçlüğü) faktörler arası değişkenler (between-factor variables) ve kestirim modeli (ÇBMTK, ÜDMTK ve Bifaktör model) faktör içi değişken olarak tanımlanmıştır. Dört bağımsız değişkene sahip olan ANOVA için yorumlamalar anlamlı olan en yüksek etkileşim etkisi için incelenmiştir. Dört yönlü etkileşim etkisi anlamlı olduğunda bir değişkenin her bir düzeyinde diğer iki değişken

arasındaki basit üç yönlü etkileşimlerin anlamlı olup olmadığı incelenmiştir. Bu analizlerde üç yönlü etkileşim etkisi anlamlı olduğunda bir değişkenin her bir düzeyinde diğer iki değişken arasındaki basit iki yönlü etkileşimlerin incelenmesi yaklaşımı kullanılmıştır. İki yönlü etkileşimler anlamlı olduğunda her bir değişken için basit ana etkiler incelenmiştir. Basit ana etki testleri bir değişkene ait ortalamaları diğer değişkenin her bir düzeyinde karşılaştırmaktadır. Etkileşim etkisi olmadığı durumlarda ise ilgili değişken(ler) için ana etkiler incelenmiştir. Farkın anlamlı çıktığı ana etkiler ve basit ana etkiler için farkın hangi gruplar arasında olduğunu belirlemek üzere Bonferroni düzeltmesi ile ikili karşılaştırmalar gerçekleştirilmiştir.

Etki büyüklüğü ölçümü olarak kısmi eta kare (η_p^2) raporlanmıştır. Kısmi eta kare değerlerinin yorumlanmasında 0,01, 0,06 ve 0,14 değerleri sırasıyla düşük, orta ve yüksek etki olarak ifade edilmektedir (Cohen, 1988, s. 286). Etki büyüklüğü 0,01 değerinin altında olduğunda anlamlılık değeri manidar olsa bile yorumlamaya dâhil edilmemiştir. Böylece istatistiksel olarak anlamlı ancak pratik olarak önemsiz farklılıkların ortaya konmasının engellenmesi amaçlanmıştır. Kirk (1996) de pratik anlamlılık kavramından bahsetmiştir. Buna göre, yalnızca sonucun istatistiksel anlamlılığına odaklanmak yerine, bağımsız değişkenlerin oluşturduğu etkinin kullanıldığı bağlamı ele almak gerekmektedir.

Alt puanların incelenmesinde de benzer şekilde faktöriyel karma ANOVA yapılmıştır. Toplam puandan farklı olarak alt puan analizlerinde boyutlar arası korelasyon, çoklu puanlanan madde yüzdesi, test güçlüğü ve yetenek kestirim modeline ek olarak alt puanlar da bir değişken olarak dâhil edilmiştir. Burada simülasyon koşulları (boyutlar arası korelasyon, çoklu puanlanan madde yüzdesi ve test güçlüğü) faktörler arası değişkenler (between-factor variables) ve kestirim yöntemi (ÇBMTK, ÜDMTK ve Bifaktör model) faktör içi değişken olarak tanımlanmış, alt testler ise blok değişken olarak ele alınmıştır.

3.2.6. Simülasyon Aşamaları – Özet

Bu kısımda, Bölüm 4.2'de ayrıntılı olarak açıklanan simülasyon çalışması ile ilgili bilgiler özetlenmiştir. Tablo 1'de simülasyon çalışmasında ele alınan sabit ve manipüle edilen koşullar, verilerin üretilmesi ve analiz aşamaları ile ilgili özet bilgiler yer almaktadır.

Tablo 1.

Simülasyon Çalışması Koşullar, Veri Üretimi ve Analizi İle İlgili Özet Tablo

Durum	**Açıklama**
Örneklem büyüklüğü	5000
Boyut sayısı	4
Madde sayısı	Her bir boyut için 30 ve toplamda 120
Boyutlar arası korelasyon	r=0,2, r=0,5 ve r=0,8
Testteki çoklu puanlanan madde yüzdesi	%5, %10, %25 ve %50
Test güçlüğü	Standart sapması 1 ve ortalaması sırasıyla -2, -1, 0, 1 ve 2 olan normal dağılıma sahip güçlük parametreleri ile çok kolay, kolay, orta güçlükte, zor, çok zor testler
Ayırt edicilik parametresi	$a_i \sim$ U [0.5, 2.5]
Güçlük/yer parametresi	$b_i \sim N$ (-2, 1), $b_i \sim N$ (-1, 1), $b_i \sim N$ (0, 1), $b_i \sim N$ (1, 1), $b_i \sim N$ (2, 1)
Şans parametresi	$c_i \sim$ beta (6, 16)
Yetenek parametresi	$\theta_i \sim MVN(0, \Sigma)$
Model	İkili puanlana maddeler için M-3PL Çoklu puanlanan maddeler için M-2PPL
Replikasyon sayısı	100
Madde-yanıt örüntülerinin üretilmesi	SimuMIRT (Yao, 2003)
Toplam puan ve alt puanların elde edildiği yöntemler	ÇBMTK, ÜDMTK ve Bifaktör
Yetenek kestirimlerinin elde edilmesi	BMIRT (Yao, 2003)
Kestirim yöntemi	MCMC - Metropolis-Hastings algoritması
İterasyon sayısı	15000 (3000 hazırlık)
Değerlendirme kriterleri	RMSE, gerçek yetenek parametreleri ile kestirilen yetenek parametreleri arasındaki korelasyon (güvenirlik) ve sınıflama doğruluğu
Analiz	Faktöriyel karma ANOVA
Etki büyüklüğü	Kısmi eta kare (η_p^2): 0,01, 0,06 ve 0,14 –sırasıyla düşük, orta ve yüksek etki

3.3. Gerçek Veri Analizi

3.3.1. Verinin Tanıtılması

Çalışmanın son aşamasında, toplam puan ve alt test puanlarının birlikte kestirilmesi için kullanılan yöntemlerin (ÇBMTK, ÜDMTK ve Bifaktör) karma formatta olan gerçek bir test verisi üzerinde gösterilmesi amacıyla Türkiye'de bir devlet üniversitesinin Yabancı Diller Yüksek Okulu tarafından geliştirilip uygulanan İngilizce yeterlilik sınavı verileri kullanılmıştır. Sınava girenlerin çoğu hazırlık sınıfı öğrencilerinden oluşmakta ve lisans

eğitimine devam edebilmeleri için en az 60 puan almaları gerekmektedir. Belirlenen puanın altında puana sahip olan öğrenciler hazırlık sınıfını tekrar okumak zorundadır. Sınava ait veri seti 5085 kişiden oluşmaktadır. Sınava ilişkin bilgiler Tablo 2'de sunulmuştur.

Tablo 2.

Gerçek Veri İçeriği ve Puanlama

Bölümler	Madde sayısı - Türü	Puanlama	Toplam puan
Dinleme	30 madde – ikili puanlama	0 – 1	30
Okuma	33 madde – 32 madde ikili puanlama ve 1 madde çoklu puanlama	0 – 1 0 – 6	38
Kelime	20 madde – ikili puanlama	0 – 1	20
Yazma	2 madde – çoklu puanlama	0 – 6	12
Testin bütünü	85 madde – karma test		100

Araştırma kapsamında kullanılan İngilizce yeterlilik sınavı dinleme, okuma, kelime ve yazma olmak üzere dört boyuttan oluşmaktadır. Tablo 2'de görüldüğü üzere, boyutlardaki madde sayısı 2 ila 33 arasında değişmektedir. Hem ikili hem çoklu puanlanan maddelerden oluşan test karma format test yapısı göstermektedir. Testte toplam 85 madde yer alırken bunlardan 3 tanesi çoklu, 82 tanesi ikili puanlanmaktadır. İkili puanlanan maddeler 0-1 şeklinde iki kategoride ve çoklu puanlanan maddeler 0-1-2-3-4-5-6 olacak şekilde 7 kategoride puanlanmaktadır. Tablo 3'te veriye ait betimsel istatistikler bulunmaktadır.

Tablo 3.

Gerçek Veriye Ait Betimsel İstatistikler

Alt test/ toplam	Madde sayısı	$\bar{X}$	ss	En düşük puan	En yüksek puan	Çarpıklık	Basıklık	α
Dinleme	30	20,826	5,259	5,000	30,000	-,290	-,692	,823
Okuma	33	27,517	6,482	5,000	38,000	-,602	-,414	,833
Kelime	20	9,528	4,343	0,000	20,000	,270	-,662	,817
Yazma	2	7,701	2,483	0,000	12,000	-,640	,187	,601
Toplam	85	65,573	15,973	19,000	98,000	-,294	-,654	

Seksen beş maddeden oluşan İngilizce yeterlilik sınavından elde edilen puanların ortalaması 65,57'tür. Sınavdan alınan en düşük puan 19,00 ve en yüksek puan ise 98,00'dır. Toplam teste ve alt testlere ait çarpıklık ve basıklık katsayıları incelendiğinde dağılımın normal olduğu söylenebilir. Dinleme, okuma, kelime, yazma alt testleri için hesaplanan güvenirlik katsayıları sırasıyla 0,82, 0,83, 0,82 ve 0,60'tır.

3.3.2. Boyutluluk Analizi

Veriden elde edilen puanların kullanımının ve yorumlanmasının daha doğru olması için geçerlik kanıtlarının elde edilmesinde verinin boyut yapısının incelenmesi önemlidir (Reckase & Xu, 2015). Boyutluluk, test ile ölçülmesi hedeflenen örtük yapı hakkında bilgi veren, test ile yanıt örüntüleri arasındaki ilişkiyi gösterir. Wainer ve Thissen (1996) sabit (fixed) ve seçkisiz (random) boyutluluk kavramlarından bahsetmektedir. Seçkisiz boyutluluk, bazı "beklenmedik" boyutlarla karşılaşma olasılığını açıklayan bir kavram iken, sabit boyutluluk bir yönüyle "beklenen" bir durumdur. Özellikle, testin genel yeteneğe ek olarak alt alanlarına sahip olduğu durumlarda, puanlarda çok boyutluluğun görülmesi olasıdır. Test geliştiren uzmanlar alt alanlara sahip bir test oluşturduklarında veri için zaten çok boyutlu bir yapı varsaymış olurlar. Bu durumda, doğrulayıcı boyutluluk yaklaşımlarının kullanılmasının daha mantıklı olabileceği belirtilmiştir (Zhang, 2016). Çalışma kapsamında yer alan gerçek veri seti de İngilizce genel yeterliliğini ölçmek amacıyla ilgili kurum tarafından alt testlerden oluşacak şekilde hazırlanmış ve uygulanmıştır. Bu sebeple verinin boyutluluğuna ilişkin analizde doğrulayıcı yöntemler kullanılmıştır. Bunlardan ilki Doğrulayıcı Faktör Analizi (DFA) ve bir diğeri ise içerik temelli doğrulayıcı Poly-DETECT (Zhang & Stout, 1999a, 1999b; Zhang, 2007) analizidir.

Veriye ait dört boyutlu yapı için DFA, M-plus programı ile yapılmıştır. Model uyumunun değerlendirilmesinde kullanılan uyum indeksleri, CFI (Comparative Fit Index – Karşılaştırmalı Uyum İndeksi), TLI (Tucker-Lewis Index –Tucker-Lewis İndeksi) ve RMSEA'dır (Root Mean Square Error of Approximation –yaklaşık hataların ortalama karekökü). CFI ve TLI indekslerinin 0,95'ten büyük ve RMSEA değerinin 0,5'ten küçük olması, modelin iyi uyum sağladığını göstermektedir (Hu & Bentler, 1999; Tabachnick & Fidell, 2013, s. 720-723).

Poly-DETECT analizi R programı *sirt* paketi ile yapılmıştır. Analiz sonucunda DETECT, ASSI ve RATIO indeksleri elde edilmektedir. Bu indekslerin değerlendirilmesi ile ilgili bilgiler Tablo 4'te sunulmuştur (Jang & Roussos, 2007; Zhang, 2007).

Tablo 4.

Poly-DETECT Analizi Boyutluluk İndeksleri ve Değerlendirme Kriterleri

İndeks	Kritik değerler	Değerlendirme
DETECT	DETECT > 1,00	Güçlü çok boyutluluk
	,40 < DETECT < 1,00	Orta düzey çok boyutluluk
	,20 < DETECT < ,40	Zayıf çok boyutluluk
	DETECT < ,20	Tek boyutlu
ASSI	ASSI=1	Basit yapı altında en yüksek değer
	ASSI > ,25	Tek boyutluluktan sapma
	ASSI < ,25	Tek boyutlu
RATIO	RATIO=1	Basit yapı altında en yüksek değer
	RATIO > ,36	Tek boyutluluktan sapma
	RATIO < ,36	Tek boyutlu

DETECT indeksi testteki çok boyutluluğun miktarını gösterir. DETECT değerinin 1'den büyük olması çok boyutluluğun güçlü olduğunu, 0,40 -1,00 arası orta düzey, 0,40 değerinin altı zayıf olduğunu göstermektedir. DETECT değerinin 0,2 altında olması tek boyutluluğa işaret etmektedir. ASSI ve RATIO indeksleri için kritik değerler sırasıyla 0,25 ve 0,36'dır. Bu değerlerden yüksek olan ASSI ve RATIO indeksleri çok boyutluluğun göstergesi olarak ele alınır (Jang & Roussos, 2007; Zhang, 2007).

3.3.3. Model Kestirimleri

İngilizce yeterliliğini ölçen toplam puan ve dinleme, okuma, kelime ve yazma alt test puanlarının hesaplanmasında ÇBMTK, ÜDMTK ve Bifaktör model kullanılmıştır. İngilizce genel yeterlilik ve dört alt beceri için yetenek parametrelerinin kestirimi BMIRT (Yao, 2003; Yao, 2016) programı ile yapılmıştır. MCMC–Metropolis-Hastings algoritmasının kullanıldığı analizdeki iterasyon sayısı ve öncüller simülasyon veri için belirlenenler ile aynıdır (Bkz. Bölüm 3.2.3). Boyutlar arası korelasyon ÇBMTK ve ÜDMTK için r=0,5 ve Bifaktör model için r=0,0 olarak tanımlanmıştır. ÇBMTK model için genel yetenek kestirimi maksimum bilgi yöntemi ile elde edilmiştir (Yao, 2010).

3.3.4. Gerçek Veri Analizi İçin Değerlendirme Kriterleri

İngilizce yeterlilik sınavına ait toplam test ve alt testlere ilişkin yetenek parametrelerinin doğruluğu, ölçmenin koşullu standart hatası (CSEM-conditional standard error of measurement) ile değerlendirilmiştir. BMIRT programı ilgili modellerin hepsinde her bir yanıtlayıcının her bir yetenek parametresi için CSEM değeri hesaplamaktadır. Hatalar

arasındaki farkın istatistiksel olarak anlamlı olup olmadığının değerlendirilmesinde tekrarlı ölçümler ANOVA kullanılmıştır. Toplam ve alt test puanlarına ilişkin hataların her biri için olmak üzere beş adet ANOVA yapılmıştır. CSEM değeri daha düşük olan modelin ilgili veri seti için diğer modellere göre daha doğru kestirim yaptığı sonucuna ulaşılabilir.

ÇBMTK, ÜDMTK ve Bifaktör model ile kestirilen yetenek parametrelerinin güvenirliği bir diğer değerlendirme kriteridir. Bu amaçla, Bayes marjinal güvenirlik (de la Torre & Patz, 2005) ya da görgül güvenirlik (Brown & Croudace, 2015) olarak adlandırılan güvenirlik kestirim yöntemi kullanılmıştır. Eşitlik 36, 37 ve 38 ile gösterilmektedir.

$$\rho_d = \frac{var\,(\hat{\theta}_d)}{var\,(\hat{\theta}_d) + \overline{Pvar\,(\hat{\theta}_d)}} \qquad \text{(Eşitlik 36)}$$

Burada d, toplam test ya da alt testi işaret etmektedir. N sayıda yanıtlayıcıdan oluşan bir örneklemde, toplam ya da alt teste ilişkin yetenek kestirimlerinin gözlenen ve marjinal sonsal varyansı, kestirilen yetenek parametrelerinden ($\hat{\theta}$) ve bunlara ilişkin standart hatalardan (SE) hesaplanmaktadır:

$$var\left(\hat{\theta}_d\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_i - \overline{\hat{\theta}}\right)^2 \qquad \text{(Eşitlik 37)}$$

$$\overline{Pvar\,(\hat{\theta}_d)} = \frac{1}{N}\sum_{i=1}^{N} SE^2\left(\hat{\theta}_i\right) \qquad \text{(Eşitlik 38)}$$

Marjinal güvenirlik katsayısı ilgili modeller kullanılarak hesaplanan toplam ve alt test puanlarının hepsi için ayrı ayrı hesaplanmıştır. Yüksek marjinal güvenirlik değeri, test edilen modellerden elde edilen puanların güvenirliğini yüksek olduğunu göstermektedir (Md Desa, 2012).

Son olarak, modellerin performansı Rudner (2001, 2005) sınıflama tutarlılığı ve doğruluğu indeksleri hesaplanarak incelenmiştir. Hesaplamalar R cacIRT (Lathrop, 2015) paketi ile yapılmıştır. İngilizce yeterlilik sınavı toplam puanı için geçme notu 60'tır. Ham veri incelendiğinde grubun %65'inin geçtiği görülmektedir. ÇBMTK, ÜDMTK ve Bifaktör modelleri ile elde edilen toplam ve alt test puanlarının her biri için kesme puanı, kestirilen yetenek parametrelerine göre grubun %65'inin geçeceği şekilde belirlenmiştir.

BÖLÜM 4

BULGULAR

Bu bölümde çalışmanın araştırma problemleri doğrultusunda elde edilen bulgular sunulmuştur. Çalışmada hem simülasyon veri hem de gerçek veri kullanılmış olup sırasıyla raporlanmıştır. Simülasyon çalışmasında toplam puan ve alt puanlar için gerçekleştirilen analizler ayrı ayrı sunulmuştur. ÇBMTK, ÜDMTK ve Bifaktör modeller ile elde edilen yetenek kestirimleri kestirim doğruluğu (RMSE), güvenirlik ve sınıflama doğruluğu kriterlerine göre karşılaştırılmıştır. Kestirim modelleri ve simülasyon çalışmasında ele alınan değişkenler ile birlikte ilgili kriterlerin karşılaştırılması için faktöriyel karma ANOVA yapılmış ve anlamlı çıkan en büyük etkileşim etkisinden başlayarak incelemeler gerçekleştirilmiştir. Toplam puan ve alt puanlar için ayrı ayrı olmak üzere anlamlı çıkan etkileşim etkilerinin incelenmesinde sırasıyla kestirim modelleri ve test güçlüğü, çoklu puanlanan madde yüzdesi ve boyutlar arası korelasyon basit ana etkileri ele alınmıştır. Böylece çalışmada ele alınan kestirim modellerinin ve diğer değişkenlerin yetenek kestiriminin doğruluğu, güvenirlik ve sınıflama doğruluğu değerleri açısından hem toplam test hem de alt testler için nasıl bir etkiye sebep olduğu açıklanmaya çalışılmıştır. Son bölümde bir İngilizce yeterlilik sınavı verileri için boyutluluk analizi, madde ve yetenek parametrelerine ait betimsel istatistikler ve ele alınan kestirim modellerinin hem toplam test hem de alt testler için yetenek kestirim doğruluğu, güvenirlik ve sınıflama doğruluğu değerlerine göre karşılaştırılmasına yer verilmiştir.

"ÇBMTK, ÜDMTK, Bifaktör kestirim modelleri ve farklı simülasyon koşulları (test güçlüğü, boyutlar arası korelasyon ve çoklu puanlanan madde yüzdesi) altında elde edilen yetenek kestirimlerine ait RMSE, güvenirlik ve sınıflama doğruluğu değerleri alt puan ve toplam puan bağlamında farklılaşmakta mıdır?" araştırma sorusuna yönelik bulgular Bölüm 5.1 ve Bölüm 5.2'de sunulmuştur. "Gerçek veri üzerinde ÇBMTK, ÜDMTK, Bifaktör kestirim modelleri ile

elde edilen yetenek kestirimlerine ait hata, güvenirlik ve sınıflama doğruluğu değerleri alt test ve toplam test bağlamında farklılaşmakta mıdır?" araştırma sorusuna yönelik bulgular ise Bölüm 5.3'te yer almaktadır.

4.1 Simülasyon Çalışması Toplam Puan için Bulgular

Bu bölümde simülasyon veri ile gerçekleştirilen analizlerde toplam puan açısından incelemelere yer verilmiştir. ÇBMTK, ÜDMTK, Bifaktör kestirim modellerinin ve simülasyon koşullarının (test güçlüğü, boyutlar arası korelasyon ve çoklu puanlanan madde yüzdesi) yetenek kestirimlerine ait RMSE, güvenirlik ve sınıflama doğruluğu değerlerine olan etkisi faktöriyel karma ANOVA sonuçlarına göre alt başlıklar halinde incelenmiştir.

ÇBMTK, ÜDMTK ve Bifaktör modelleri ile yapılan analiz sonrasında elde edilen toplam puanlar için RMSE, güvenirlik ve sınıflama doğruluğu değerleri hesaplanmıştır. Bu değerlerin analiz modellerine ve simülasyon koşullarına göre karşılaştırılması amacıyla her bir bağımlı değişken (RMSE, güvenirlik ve sınıflama doğruluğu) için ayrı ayrı olmak üzere faktöriyel karma ANOVA yapılmıştır. Sonuçlar Tablo 5'te sunulmuştur.

Tablo 5.

Toplam Puan için RMSE, Güvenirlik ve Sınıflama Doğruluğu Değerlerine ait Faktöriyel Karma ANOVA Sonuçları

	RMSE		Güvenirlik		Sınıflama doğruluğu	
Kaynak	F	η_p^2	F	η_p^2	F	η_p^2
Gruplar içi						
M	2772860,209*	,998	472973,711*	,988	3566439,853*	,998
M *x* ÇPMY	19176,388*	,906	12155,917*	,860	33175,400*	,944
M *x* TG	3935,895*	,726	4128,322*	,735	4013,482*	,730
M *x* K	422537,013*	,993	228352,277*	,987	526408,817*	,994
M *x* TG *x* ÇPMY	4964,935*	,909	4121,037*	,893	4900,978*	,908
M *x* ÇPMY *x* K	3896,690*	,797	5181,584*	,840	5395,170*	,845
M *x* K *x* TG	3264,826*	,815	2922,098*	,797	1692,996*	,695
M *x* TG *x* ÇPMY *x* K	3083,225*	,926	3157,756*	,927	3195,750*	,928
Gruplar arası						
ÇPMY	17422,074*	,898	1325,628*	,401	44215,816*	,957
TG	209822,420*	,993	1837,820*	,553	58031,714*	,975
K	608927,672*	,995	56909,576*	,950	2321356,798*	,999
ÇPMY *x* TG	6342,922*	,928	347,886*	,413	3710,243*	,882
ÇPMY *x* K	3784,933*	,793	529,056*	,348	1148,164*	,537
TG *x* K	2175,088*	,746	198,730*	,211	2287,259*	,755
ÇPMY *x* TG *x* K	2340,575*	,904	235,695*	,488	2237,204*	,900

*p<,001 M: model; ÇPMY: çoklu puanlanan madde yüzdesi; TG: test güçlüğü; K: korelasyon

Tablo 5'te yer alan RMSE, güvenirlik ve sınıflama doğruluğu değerleri ile gerçekleştirilen analizler sonucunda model, çoklu puanlanan madde yüzdesi, test güçlüğü ve korelasyon değişkenleri arasındaki dört yönlü etkileşimin anlamlı ve yüksek etki büyüklüğüne sahip olduğu görülmektedir. Bu dört yönlü etkileşimin yorumlanması için öncelikle boyutlar arası korelasyon değişkeninin her bir düzeyinde (0,2, 0,5 ve 0,8) çoklu puanlanan madde yüzdesi, model ve test güçlüğü değişkenleri arasındaki üç yönlü etkileşim etkisi incelenmiştir. Ardından çoklu puanlanan madde yüzdesi değişkeninin her bir düzeyinde model, korelasyon ve test güçlüğü değişkenleri arasındaki üç yönlü etkileşim etkisi incelenmiştir. Üç yönlü etkileşim etkisinin incelenmesine yönelik analizler aşağıda ayrı ayrı sunulmaktadır.

4.1.1. Kestirim Modelleri ve Test Güçlüğü Değişkeni İçin İncelemeler

Bu bölümde gerçekleştirilen analizlerde simülasyon veride toplam puan açısından yetenek kestirim modellerinin ve test güçlüğünün, ele alınan değerlendirme kriterleri açısından yetenek kestirimine olan etkisi ortaya konmuştur.

Korelasyon 0,2 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

Korelasyon 0,2 düzeyinde model, çoklu puanlanan madde yüzdesi ve test güçlüğü değişkenleri arasındaki üç yönlü etkileşim etkisinin incelenmesi amacıyla bu değişkenler ile RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için ayrı ayrı faktöriyel karma ANOVA yapılmıştır. Analiz sonuçları Tablo 6'da yer almaktadır.

Tablo 6.

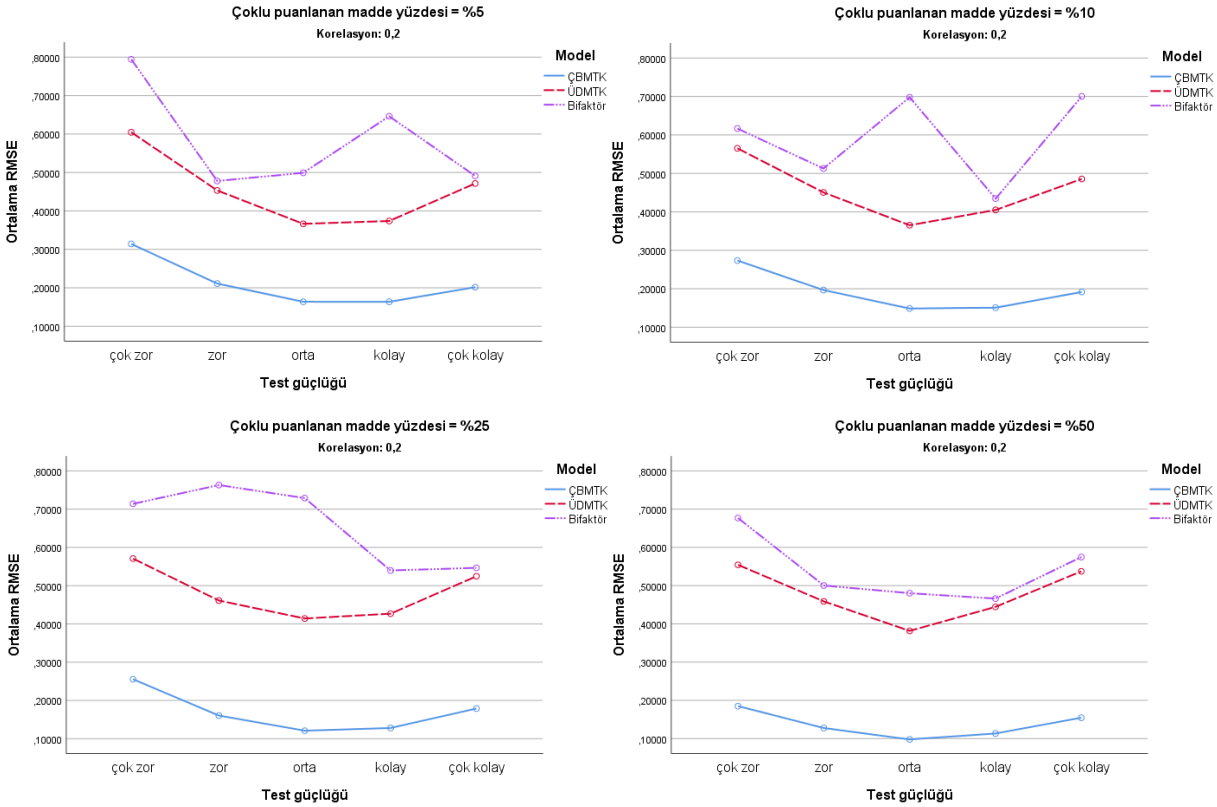
Toplam Puan için Korelasyon 0,2 Düzeyinde Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Karma ANOVA

	RMSE		Güvenirlik		Sınıflama doğruluğu	
Kaynak	F	η_p^2	F	η_p^2	F	η_p^2
Gruplar içi						
M	1000326,630*	,998	303094,731*	,994	1380153,546*	,999
M *x* ÇPMY	6255,411*	,905	6988,992*	,914	13450,181*	,953
M *x* TG	3973,618*	,889	3388,942*	,873	2208,134*	,817
M *x* TG *x* ÇPMY	4274,594*	,963	3577,091*	,956	4359,054*	,964
Gruplar arası						
ÇPMY	6846,526*	,912	7824,080*	,922	10845,090*	,943
TG	43226,018*	,989	4560,459*	,902	7002,664*	,934
ÇPMY *x* TG	4131,195*	,962	2856,510*	,945	4301,441*	,963

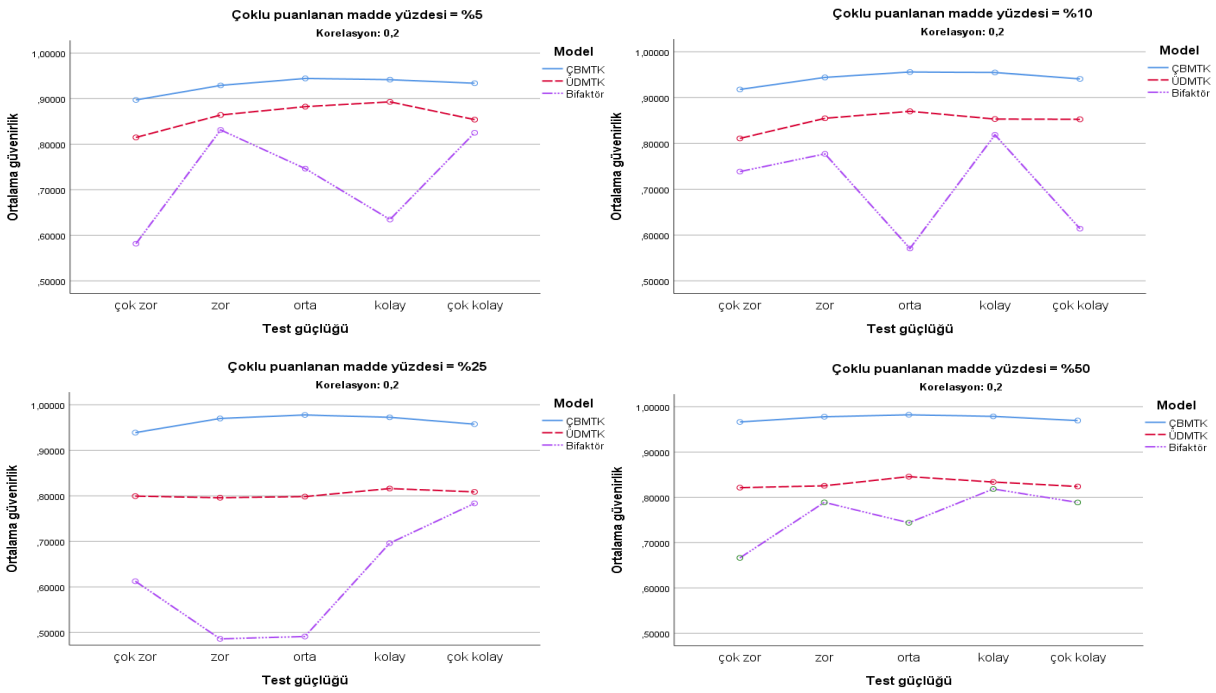
*p<,001 M: model; ÇPMY: çoklu puanlanan madde yüzdesi; TG: test güçlüğü

Boyutlar arası korelasyon 0,2 düzeyinde RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için gerçekleştirilen ANOVA sonucunda model, test güçlüğü ve çoklu puanlanan madde yüzdesi değişkenleri arasındaki üç yönlü etkileşimin her üç değerlendirme kriteri için de anlamlı olduğu tespit edilmiştir. Bu üç yönlü etkileşimleri incelemek amacıyla basit iki yönlü etkileşim yaklaşımı kullanılmıştır. Bunun için çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü arasındaki etkileşim etkisi ve ardından test güçlüğünün her bir düzeyinde (çok zor, zor, orta, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki etkileşim etkisi incelenmiştir.

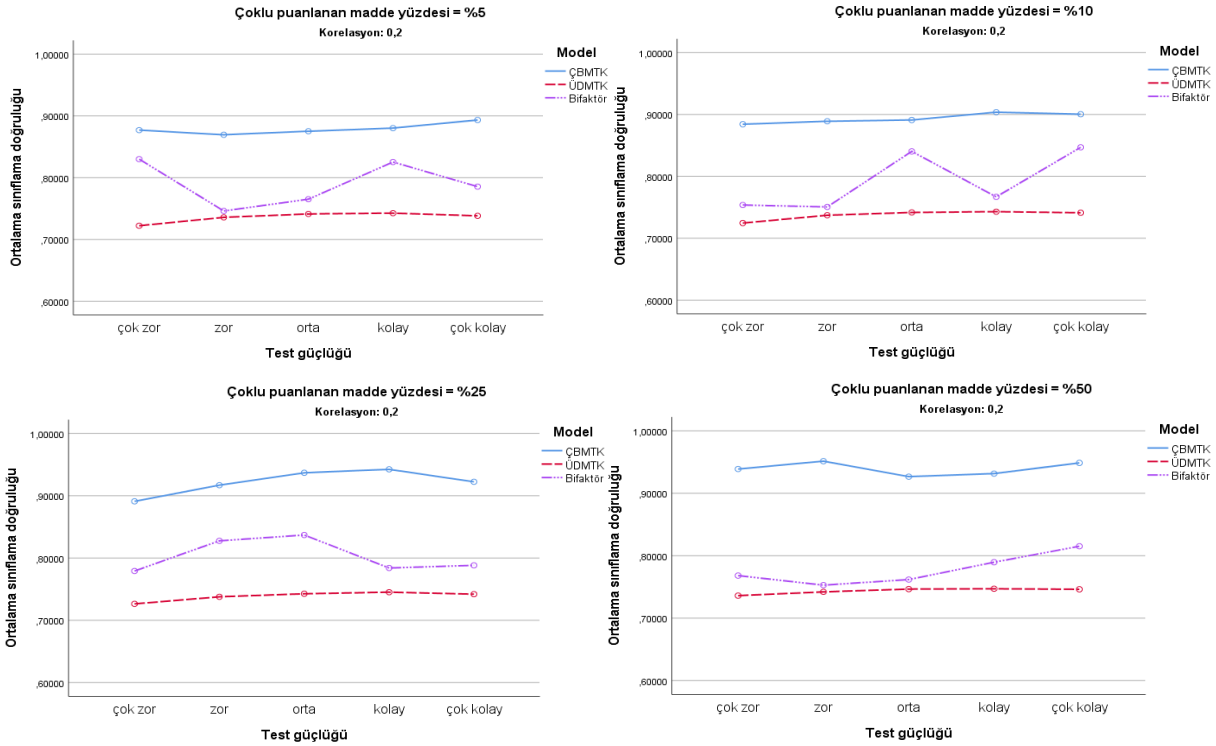
Öncelikle çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü değişkenleri ile RMSE, güvenirlik ve sınıflama doğruluğu değerleri için iki yönlü karma ANOVA yapılmış ve sonuçlar Tablo 7'de sunulmuştur. Çoklu puanlanan madde yüzdesinin her bir düzeyi için model ve test güçlüğü düzeylerinde elde edilen ortalama RMSE, güvenirlik ve sınıflama doğruluğu değerlerini içeren grafikler sırasıyla Şekil 3, Şekil 4 ve Şekil 5'te yer almaktadır.



Şekil 3. Toplam puan korelasyon 0,2 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE)



Şekil 4. Toplam puan korelasyon 0,2 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik)



Şekil 5. Toplam puan korelasyon 0,2 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (sınıflama doğruluğu)

Tablo 7.

Toplam Puan için Korelasyon 0,2 Düzeyinde Her Bir Çoklu Puanlanan Madde Yüzdesi için Model ve Test Güçlüğü Değişkenleri ile İki Yönlü Karma ANOVA Sonuçları

			RMSE		Güvenirlik		Sınıflama doğruluğu	
ÇPMY	Kaynak		F	η_p^2	F	η_p^2	F	η_p^2
%5	Gruplar içi	M	159914,984*	,997	54115,465*	,991	160591,694*	,997
		M *x* TG	3596,819*	,967	3734,696*	,968	3071,483*	,961
	Gruplar arası	TG	20863,065*	,994	5225,328*	,977	2662,826*	,956
%10	Gruplar içi	M	287498,936*	,998	94041,642*	,995	257417,793*	,998
		M *x* TG	7585,850*	,984	5599,760*	,978	5558,038*	,978
	Gruplar arası	TG	14416,002*	,991	3531,873*	,966	7648,739*	,984
%25	Gruplar içi	M	247607,088*	,998	100886,591*	,995	355246,585*	,999
		M *x* TG	3839,357*	,969	3702,744*	,968	2198,967*	,947
	Gruplar arası	TG	6637,657*	,982	3011,853*	,961	5068,944*	,976
%50	Gruplar içi	M	414281,917*	,999	66523,983*	,993	1723815,278*	1,000
		M *x* TG	1881,032*	,938	1224,672*	,908	7179,586*	,983
	Gruplar arası	TG	16325,971*	,992	1363,696*	,917	6056,085*	,980

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi; M: model; TG: test güçlüğü

Tablo 7'de yer alan boyutlar arası korelasyon 0,2 düzeyi ve çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) gerçekleştirilen analiz sonuçları incelendiğinde model ve test güçlüğü arasındaki etkileşimin RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için anlamlı olduğu görülmektedir. Bu etkileşim etkilerine ait etki büyüklükleri yüksektir. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi için sırasıyla model ve test güçlüğü değişkenlerinin her bir düzeyinde diğer değişkene ait basit ana etkiler tek yönlü ANOVA ile incelenmiştir. İlk olarak test güçlüğü basit ana etkisi incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde test güçlük düzeyleri (çok zor, zor, orta güçlük, kolay, çok kolay) için hesaplanan RMSE, güvenirlik ve sınıflama doğruluğu değerleri ayrı ayrı karşılaştırılmıştır. Test güçlüğü basit ana etkisi için elde edilen sonuçlar Tablo 8'de verilmiştir.

Tablo 8.

Toplam Puan için Korelasyon 0,2 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması

		RMSE		Güvenirlik		Sınıflama doğruluğu	
ÇPMY	Model	F	η_p^2	F	η_p^2	F	η_p^2
%5	ÇBMTK	58970,361*	,998	12479,496*	,990	3223,925*	,963
	ÜDMTK	52439,937*	,998	5325,560*	,977	1258,040*	,910
	Bifaktör	6013,836*	,980	4157,271*	,971	3106,179*	,962
%10	ÇBMTK	38141,471*	,997	11811,422*	,990	2909,988*	,959
	ÜDMTK	26843,640*	,995	1832,838*	,937	1358,724*	,917
	Bifaktör	7084,847*	,983	5078,368*	,976	7208,729*	,983
%25	ÇBMTK	52547,122*	,998	24687,011*	,995	25533,030*	,995
	ÜDMTK	10095,559*	,988	114,044*	,480	1511,546*	,924
	Bifaktör	3377,537*	,965	3865,695*	,969	2272,818*	,948
%50	ÇBMTK	22984,656*	,995	8372,082*	,985	8380,895*	,985
	ÜDMTK	21516,460*	,994	401,572*	,764	816,174*	,868
	Bifaktör	4862,321*	,975	1344,605*	,916	8494,737*	,986

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 8'de çoklu puanlanan madde yüzdesi ve kestirim modellerinin tüm düzeylerinde RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her birinin test güçlüklerine göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. Anlamlı bulunan bu farkların kaynağının belirlenmesi için ikili karşılaştırmalar yapılmıştır (Ek 4). Buna göre çoklu puanlanan madde yüzdesi %10, %25 ve %50 olduğunda ÇBMTK ve ÜDMTK için elde edilen RMSE değerleri test güçlük düzeylerine göre en düşükten en yükseğe doğru sırasıyla orta, kolay, zor, çok kolay ve çok zor şeklindedir. Çoklu puanlanan madde yüzdesi %5 olduğunda ise bu sıralama orta, kolay, çok kolay, zor ve çok zor şeklindedir ve ÇBMTK için orta ve kolay test güçlükleri için hesaplanan RMSE değerleri arasındaki fark anlamlı değildir. Bifaktör model için de ikili karşılaştırmalarda anlamlı farklılıkların olduğu ancak ÇBMTK ve ÜDMTK modellerinde olduğu gibi bu anlamlı farklılıkların test güçlük düzeylerine göre belli bir eğilimde olmadığı sonucuna ulaşılmıştır.

Çoklu puanlanan madde yüzdesi %10, %25 ve %50 olduğunda ÇBMTK için elde edilen güvenirlik değerleri test güçlük düzeylerine göre en yüksekten en düşüğe doğru sırasıyla orta, kolay, zor, çok kolay ve çok zor şeklindedir. Çoklu puanlanan madde yüzdesi %5 olduğunda ise bu sıralama orta, kolay, çok kolay, zor ve çok zor şeklindedir. ÜDMTK model için çoklu puanlanan madde yüzdesi %10 ve %50 olduğunda en yüksek güvenirlik test güçlüğü orta olduğunda elde edilirken çoklu puanlanan madde yüzdesi %5 ve %25 olduğunda en yüksek güvenirlik test güçlüğü kolay olduğunda elde edilmiştir. Ancak çoklu puanlanan madde yüzdesi %10 olduğunda test güçlüğü zor, kolay ve çok kolay arasında, %25 olduğunda orta, zor ve çok

zor arasında ve %50 olduğunda zor, çok zor ve çok kolay arasında anlamlı farklılık bulunmamaktadır. Bifaktör model için ise RMSE değerleri ile yapılan analizlerde olduğu gibi güvenirlik için de anlamlı bulunan farklılıkların test güçlük düzeylerine göre belli bir eğilimde olmadığı sonucuna ulaşılmıştır (Ek 4).

Sınıflama doğruluğu ile yapılan analizlerde ÜDMTK için çoklu puanlanan madde yüzdesi %5 iken elde edilen sınıflama doğruluğu değerleri en büyükten en küçüğe doğru sırasıyla kolay, orta, çok kolay, zor ve çok zor şeklindedir. Diğer çoklu puanlanan madde yüzdeleri için sıralama aynıdır fakat %10 ve %25 için test güçlüğü orta ve çok kolay, %50 için orta, kolay ve çok kolay olduğunda elde edilen sınıflama doğruluğu değerleri arasındaki fark anlamlı değildir. ÇBMTK için çoklu puanlanan madde yüzdesinin bazı düzeylerinde bazı düzeylerde test güçlüğü kolay ve çok kolay olduğunda daha yüksek sınıflama doğruluğu değerleri elde edilmesine rağmen genel olarak tüm düzeyler için belli bir eğilimin olmadığı görülmüştür. Bifaktör model için RMSE ve güvenirlik analizlerinde olduğu gibi sınıflama doğruluğu değerleri açısından da test güçlük düzeylerine göre belli bir eğilimin olmadığı sonucuna ulaşılmıştır (Ek 4).

Özetle, RMSE değerleri açısından ÇBMTK ve ÜDMTK yöntemleri için en düşük değer test orta güçlükte elde edilirken en yüksek değerler test çok kolay ve çok zor olduğunda elde edilmiştir. Güvenirlik için bu durum tam tersi olup en yüksek güvenirlik test orta güçlükte olduğunda elde edilmiştir. Sınıflama doğruluğunda ise test güçlüğü zorlaştıkça elde edilen değerler azalmıştır. Bifaktör model için ise her üç değerlendirme kriterinde de test güçlük düzeylerine göre belli bir eğilimin olmadığı saptanmıştır.

Çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü arasındaki etkileşim etkisine ilişkin test güçlüğü basit ana etkilerinin incelenmesinin ardından modellerin basit ana etkileri ele alınmıştır. Bir başka ifadeyle test güçlüğüne ait düzeylerin her birinde (çok zor, zor, orta güçlük, kolay, çok kolay) yetenek kestiriminde kullanılan modellere (ÇBMTK, ÜDMTK ve Bifaktör) ait RMSE, güvenirlik ve sınıflama doğruluğu değerleri karşılaştırılmıştır. Bu amaçla gerçekleştirilen tekrarlı ölçümler ANOVA sonuçları Tablo 9'da verilmiştir.

Tablo 9.

Toplam Puan için Korelasyon 0,2 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması

		RMSE		Güvenirlik		Sınıflama doğruluğu	
ÇPMY	TG	F	η_p^2	F	η_p^2	F	η_p^2
%5	Çok zor	272263,964*	,999	33900,501*	,993	205691,907*	,999
	Zor	177468,677*	,999	14589,491*	,983	161032,015*	,998
	Orta güçlük	132676,541*	,998	16887,010*	,986	159265,584*	,998
	Kolay	158407,670*	,998	20533,749*	,988	162825,196*	,998
	Çok kolay	219544,825*	,999	21879,243*	,989	210632,514*	,999
%10	Çok zor	207384,128*	,999	24151,408*	,990	240393,887*	,999
	Zor	158704,600*	,998	17559,509*	,986	219852,935*	,999
	Orta güçlük	160485,539*	,998	38012,868*	,994	206937,295*	,999
	Kolay	155774,545*	,998	20677,840*	,988	244549,029*	,999
	Çok kolay	235470,544*	,999	30597,832*	,992	235028,166*	,999
%25	Çok zor	118174,060*	,998	22590,013*	,989	269936,621*	,999
	Zor	126164,779*	,998	41547,942*	,994	314918,464*	,999
	Orta güçlük	122873,348*	,998	42981,440*	,994	370148,886*	,999
	Kolay	104188,083*	,998	23470,323*	,990	393319,723*	,999
	Çok kolay	130970,354*	,998	18457,990*	,987	326707,498*	,999
%50	Çok zor	292633,750*	,999	49878,433*	,995	535487,928*	1,000
	Zor	227371,896*	,999	48442,476*	,995	580607,320*	1,000
	Orta güçlük	173301,318*	,999	41503,748*	,994	426876,073*	,999
	Kolay	225545,555*	,999	43334,990*	,994	440706,488*	,999
	Çok kolay	302668,760*	,999	44189,883*	,994	533440,613*	1,000

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi; TG: test güçlüğü

Tablo 9'da çoklu puanlanan madde yüzdesi ve test güçlüğünün tüm düzeylerinde RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her birinin kestirim modellerine göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. RMSE değerlerine yönelik elde edilen bu farkın kaynağını belirlemek amacıyla yapılan ikili karşılaştırmalar sonucunda, farkların hepsinin anlamlı olduğu ve RMSE değerlerinin düşükten yükseğe doğru sırasıyla ÇBMTK, ÜDMTK ve Bifaktör modellerinden elde edildiği görülmüştür (Ek 4).

Güvenirlik değerleri için gerçekleştirilen ikili karşılaştırmalarda ise bütün çoklu puanlanan madde yüzdelerinde ve bütün test güçlüklerinde yetenek kestirimine göre hesaplanan güvenirlik değerlerinin yüksekten düşüğe doğru sırasıyla ÇBMTK, ÜDMTK ve Bifaktör modellerden elde edildiği görülmüştür. Bu sıralama RMSE ile elde edilen sıralamanın tersi şeklindedir. Sınıflama doğruluğu değerleri dikkate alındığında ise benzer şekilde bütün ikili karşılaştırmaların anlamlı olduğu ve sınıflama doğruluğu değerlerinin yüksekten düşüğe doğru sırasıyla ÇBMTK, Bifaktör ve ÜDMTK modellerinden elde edildiği belirlenmiştir.

Korelasyon 0,5 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

Korelasyon 0,5 düzeyinde model, çoklu puanlanan madde yüzdesi ve test güçlüğü değişkenleri arasındaki üç yönlü etkileşim etkisinin incelenmesi amacıyla bu değişkenler ile RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için ayrı ayrı faktöriyel karma ANOVA yapılmıştır. Analiz sonuçları Tablo 10'da yer almaktadır.

Tablo 10.

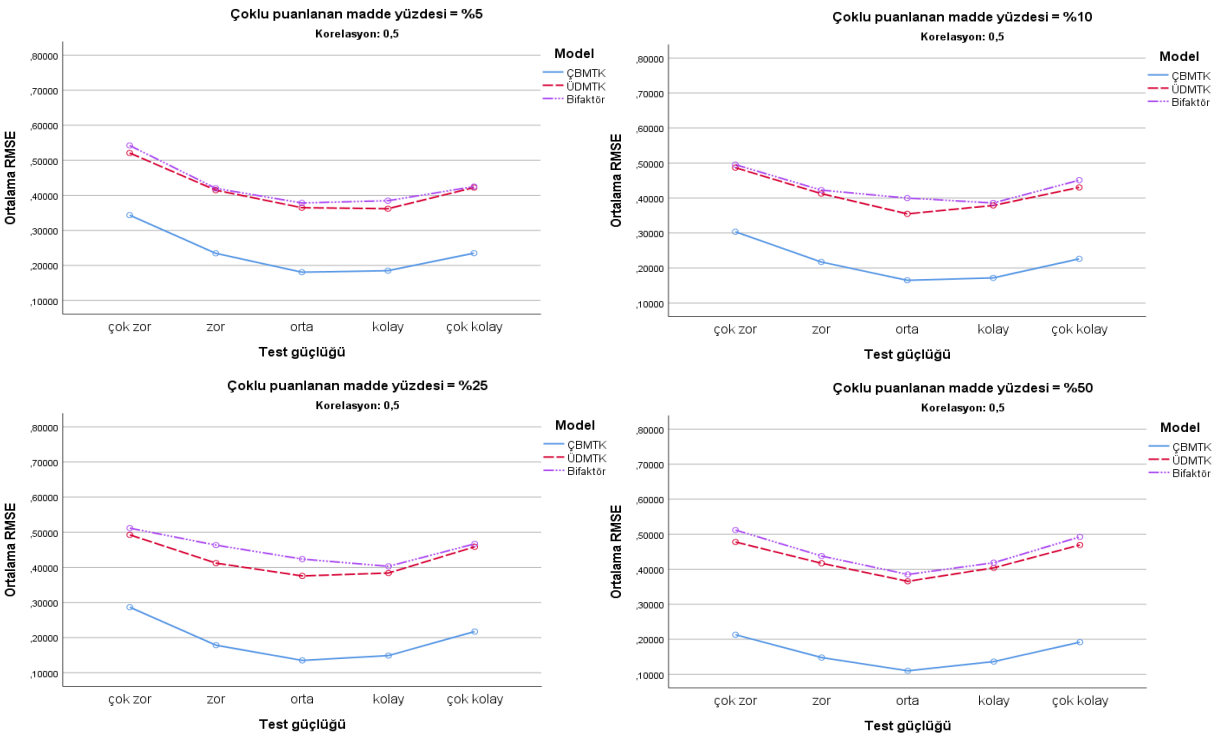
Toplam Puan için Korelasyon 0,5 Olduğunda Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Faktöriyel Karma ANOVA

	RMSE		Güvenirlik		Sınıflama doğruluğu	
Kaynak	F	η_p^2	F	η_p^2	F	η_p^2
Gruplar içi						
M	3370287,694*	,999	499040,415*	,996	2371362,611*	,999
M *x* ÇPMY	26988,837*	,976	20397,875*	,969	19171,994*	,967
M *x* TG	983,164*	,665	1664,402*	,771	4154,225*	,894
M *x* TG *x* ÇPMY	885,619*	,843	933,527*	,850	1095,803*	,869
Gruplar arası						
ÇPMY	2896,704*	,814	40,180*	,057	15929,172*	,960
TG	138076,955*	,996	225,029*	,313	24921,070*	,981
ÇPMY *x* TG	2917,410*	,946	17,166*	,094	404,384*	,710

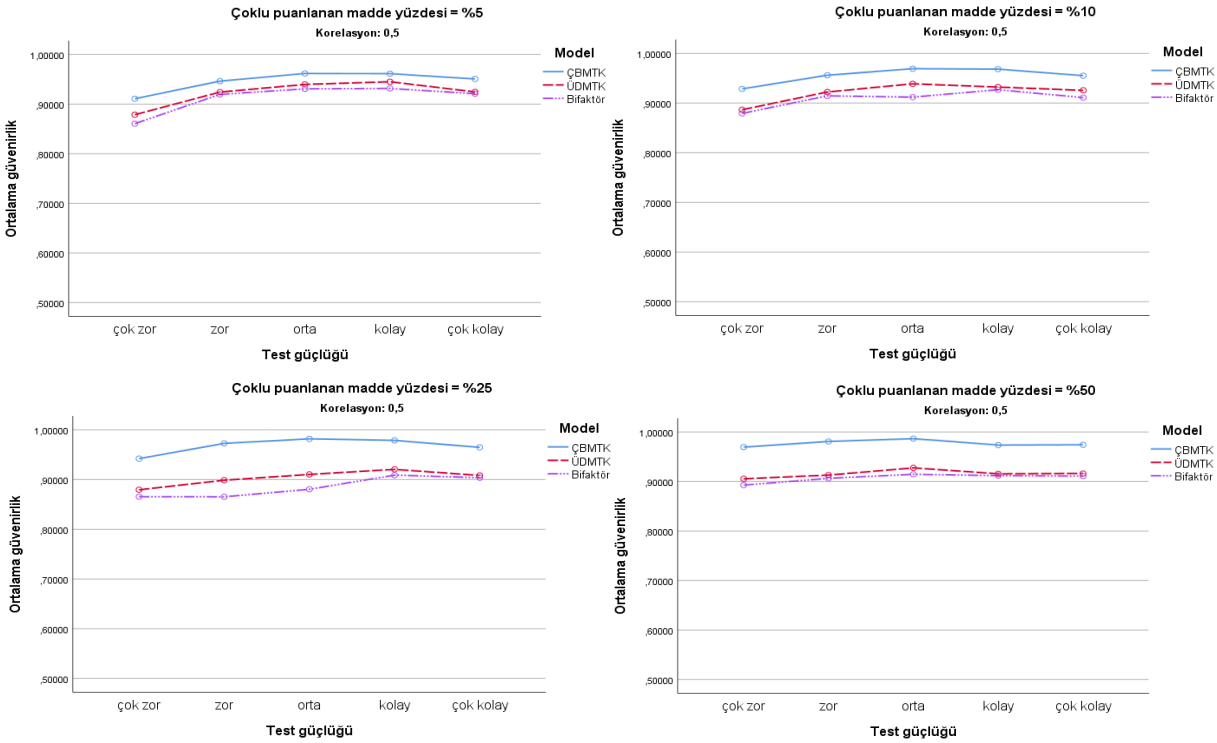
*p<,001; M: model; ÇPMY: çoklu puanlanan madde yüzdesi; TG: test güçlüğü

Boyutlar arası korelasyon 0,5 düzeyinde RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için gerçekleştirilen ANOVA sonucunda model, test güçlüğü ve çoklu puanlanan madde yüzdesi değişkenleri arasındaki üç yönlü etkileşimin her üç değerlendirme kriteri için de anlamlı olduğu tespit edilmiştir. Bu üç yönlü etkileşimleri incelemek amacıyla basit iki yönlü etkileşim yaklaşımı kullanılmıştır. Bunun için çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü arasındaki etkileşim etkisi ve ardından test güçlüğünün her bir düzeyinde (çok zor, zor, orta, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki etkileşim etkisi incelenmiştir.

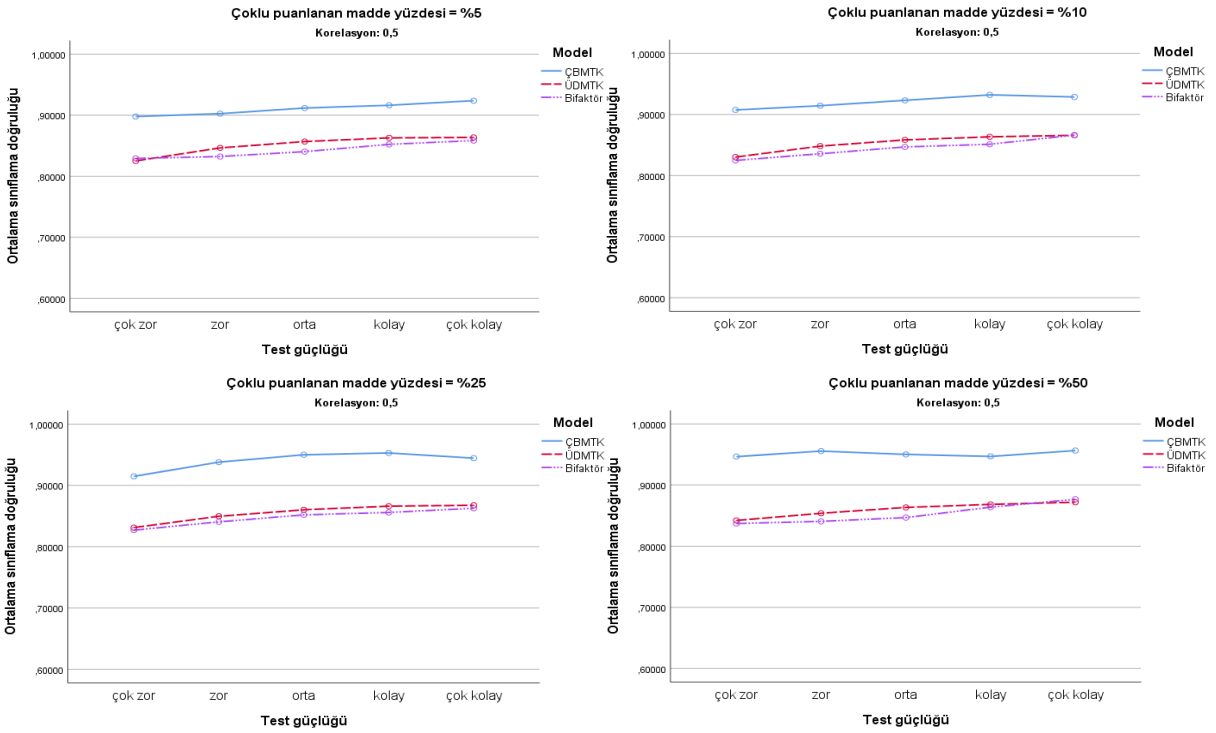
Öncelikle çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü değişkenleri ile RMSE, güvenirlik ve sınıflama doğruluğu değerleri için iki yönlü karma ANOVA yapılmış ve sonuçlar Tablo 11'de sunulmuştur. Çoklu puanlanan madde yüzdesinin her bir düzeyi için model ve test güçlüğü düzeylerinde elde edilen ortalama RMSE, güvenirlik ve sınıflama doğruluğu değerlerini içeren grafikler sırasıyla Şekil 6, Şekil 7 ve Şekil 8'de yer almaktadır.



Şekil 6. Toplam puan korelasyon 0,5 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE)



Şekil 7. Toplam puan korelasyon 0,5 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik)



Şekil 8. Toplam puan korelasyon 0,5 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (sınıflama doğruluğu)

Tablo 11.

Toplam Puan için Korelasyon 0,5 Düzeyinde Her Bir Çoklu Puanlanan Madde Yüzdesi için Model ve Test Güçlüğü Değişkenleri ile İki Yönlü Karma ANOVA Sonuçları

			RMSE		Güvenirlik		Sınıflama doğruluğu	
ÇPMY	Kaynak		F	η_p^2	F	η_p^2	F	η_p^2
%5	Gruplar içi	M	1011778,823*	1,000	81854,861*	,994	566488,803*	,999
		M *x* TG	403,514*	,765	1482,185*	,923	1929,745*	,940
	Gruplar arası	TG	71935,878*	,998	24775,141*	,995	14704,607*	,992
%10	Gruplar içi	M	822451,063*	,999	108985,231*	,995	675312,438*	,999
		M *x* TG	1223,141*	,908	910,587*	,880	1629,934*	,929
	Gruplar arası	TG	36126,969*	,997	10879,620*	,989	18676,029*	,993
%25	Gruplar içi	M	660825,759*	,999	159841,715*	,997	995304,253*	1,000
		M *x* TG	1382,985*	,918	1736,854*	,933	1086,003*	,898
	Gruplar arası	TG	21302,175*	,994	4510,973*	,973	20839,846*	,994
%50	Gruplar içi	M	1078523,741*	1,000	155279,568*	,997	443582,411*	,999
		M *x* TG	325,290*	,724	298,010*	,707	2227,413*	,947
	Gruplar arası	TG	33342,964*	,996	2,894*	,023	1468,848*	,922

*p<,001; M: model; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 11'de yer alan boyutlar arası korelasyon 0,5 düzeyi ve çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) gerçekleştirilen analiz sonuçları incelendiğinde model ve test güçlüğü arasındaki etkileşimin RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için anlamlı olduğu görülmektedir. Bu etkileşim etkilerine ait etki büyüklükleri yüksektir. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi için sırasıyla model ve test güçlüğü değişkenlerinin her bir düzeyinde diğer değişkene ait basit ana etkiler tek yönlü ANOVA ile incelenmiştir. İlk olarak test güçlüğü basit ana etkisi incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde test güçlük düzeyleri (çok zor, zor, orta güçlük, kolay, çok kolay) için hesaplanan RMSE, güvenirlik ve sınıflama doğruluğu değerleri ayrı ayrı karşılaştırılmıştır. Test güçlüğü basit ana etkisi için elde edilen sonuçlar Tablo 12'de verilmiştir.

Tablo 12.

Toplam Puan için Korelasyon 0,5 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması

		RMSE		Güvenirlik		Sınıflama doğruluğu	
ÇPMY	Model	F	η_p^2	F	η_p^2	F	η_p^2
%5	ÇBMTK	51632,965*	,998	24730,796*	,995	6081,382*	,980
	ÜDMTK	45933,058*	,997	21759,651*	,994	12666,436*	,990
	Bifaktör	36280,460*	,997	13362,228*	,991	8569,475*	,986
%10	ÇBMTK	40220,857*	,997	19197,884*	,994	6911,626*	,982
	ÜDMTK	23886,702*	,995	10065,516*	,988	11644,255*	,989
	Bifaktör	10006,535*	,988	3655,309*	,967	11982,815*	,990
%25	ÇBMTK	48339,461*	,997	29621,712*	,996	18058,789*	,993
	ÜDMTK	11929,195*	,990	2620,840*	,955	11649,080*	,989
	Bifaktör	4709,965*	,974	2403,204*	,951	8111,931*	,985
%50	ÇBMTK	20121,878*	,994	2,320	-	1515,767*	,925
	ÜDMTK	18504,492*	,993	3,673	-	1143,464*	,902
	Bifaktör	11773,944*	,990	4,367	-	2180,044*	,946

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 12'de çoklu puanlanan madde yüzdesi ve kestirim modellerinin tüm düzeylerinde RMSE ve sınıflama doğruluğu değerlerinin her birinin test güçlüklerine göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. Benzer biçimde güvenirlik değerleri için de çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğu durumlarda RMSE ve sınıflama doğruluğunda olduğu gibi anlamlı farklılık bulunmaktadır. Çoklu puanlanan madde yüzdesi %50 için ise elde edilen güvenirlik değerleri test güçlük düzeylerine göre anlamlı farklılık göstermemektedir.

Anlamlı bulunan farkların kaynağının belirlenmesi için ikili karşılaştırmalar yapılmıştır (Ek 5). Buna göre, ÇBMTK için elde edilen RMSE değerleri düşükten yükseğe doğru sırasıyla test

güçlüğünün orta, kolay, zor, çok kolay ve çok zor olması durumunda elde edilmiştir. Ancak çoklu puanlanan madde yüzdesi %5 iken test güçlüğü zor ve çok kolay olduğunda elde edilen RMSE değerleri arasındaki fark anlamlı değildir. ÜDMTK modeli ile elde edilen RMSE değerleri için de sıralama aynı olup sadece çoklu puanlanan madde yüzdesi %5 olduğunda en düşük hata test güçlüğü orta iken değil kolay olduğunda elde edilmiştir. Bifaktör model için çoklu puanlanan madde yüzdesi %5 ve %50 olduğunda RMSE değerlerinin test güçlüklerine göre sıralanışı ÇBMTK ve ÜDMTK modelleri ile benzerdir. Ancak çoklu puanlanan madde yüzdesi %10 ve %25 olduğundan en düşük RMSE test güçlüğü kolay olduğunda elde edilmiştir.

Güvenirlik değerleri açısından çoklu puanlanan madde yüzdesi %5, %10 ve %25 için elde edilen anlamlı farkın kaynağını belirlemek amacıyla ikili karşılaştırmalar yapılmıştır. Buna göre ÇBMTK için çoklu puanlanan madde yüzdesi %10 ve %25 olduğunda güvenirlik değerleri yüksekten düşüğe doğru sırasıyla test güçlüğü orta, kolay, zor, çok kolay ve çok zor olduğunda elde edilmiştir. Çoklu puanlanan madde yüzdesi %5 olduğunda ise orta ve kolay test güçlüklerinde elde edilen güvenirlikler arasındaki fark anlamlı değildir. Test güçlüğünün çok kolay, zor ve çok zor olduğu durumlarda güvenirlik daha düşüktür ve aralarındaki fark anlamlıdır (Ek 5).

ÜDMTK modeli ile yapılan kestirimler için en yüksek güvenirlik değerleri çoklu puanlanan madde yüzdesi %5 ve %25 olduğunda test kolay iken %10 olduğunda ise test orta güçlükte iken elde edilmiştir. Her üç çoklu puanlanan madde yüzdesi için de test çok zor olduğunda elde edilen güvenirlik diğerlerinden anlamlı olarak düşüktür. Bunu testin zor olduğu durum takip etmiştir. Çoklu puanlanan madde yüzdesi %5 düzeyinde test zor ve çok kolay olduğunda elde edilen güvenirlikler arasındaki fark anlamlı değildir. Bifaktör model için en yüksek güvenirlik değerleri çoklu puanlanan madde yüzdesi %10 ve %25 olduğunda test kolay iken %5 olduğunda ise test kolay ve orta güçlükte olduğunda elde edilmiştir. Diğer modeller ile benzer şekilde en düşük güvenirlik test güçlüğü çok zor olduğunda elde edilmiştir. Çoklu puanlanan madde yüzdesi %5 olduğunda test zor ve çok kolay olduğunda elde edilen güvenirlikler, %10 olduğunda test orta ve çok kolay olduğunda elde edilen güvenirlikler birbirinden anlamlı olarak farklılaşmamaktadır.

Sınıflama doğruluğu ele alındığında ÜDMTK ve Bifaktör modeller için bütün çoklu puanlanan madde yüzdelerinde sınıflama doğruluğu değerleri yüksekten düşüğe doğru sırasıyla test güçlüğü çok kolay, kolay, orta, zor ve çok zor olduğunda elde edilmiştir. Sadece çoklu puanlanan madde yüzdesi %5 düzeyinde ÜDMTK için test güçlüğü çok kolay ve kolay olduğunda elde edilen değerler arasındaki fark anlamlı değildir. ÇBMTK için de çoklu

puanlanan madde yüzdesi %5 iken test güçlüklerine göre elde edilen sınıflama doğruluğu değerlerinin sırası diğer modellerdeki gibidir. Çoklu puanlanan madde yüzdesi %10 ve %25 olduğunda ise en yüksek sınıflama test kolay olduğunda elde edilirken %50 olduğunda test zor ve çok kolay olduğunda elde edilmiştir. Genel anlamda en düşük sınıflama test güçlüğü çok zor iken hesaplanmıştır.

Özetle, korelasyon 0,5 düzeyinde, RMSE değerleri açısından bütün modeller için en düşük değer test orta güçlükte olduğunda elde edilirken en yüksek değerler test çok kolay ve çok zor olduğunda elde edilmiştir. Güvenirlik için bu durum tam tersi olup en yüksek güvenirlik test orta güçlükte olduğunda elde edilmiştir. Sınıflama doğruluğunda ise test güçlüğü zorlaştıkça elde edilen değerler azalmıştır. Çoğunlukla en yüksek sınıflama doğruluğu test çok kolay olduğunda elde edilmiştir.

Çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü arasındaki etkileşim etkisine ilişkin test güçlüğü basit ana etkilerinin incelenmesinin ardından modellerin basit ana etkileri ele alınmıştır. Bir başka ifadeyle test güçlüğüne ait düzeylerin her birinde (çok zor, zor, orta güçlük, kolay, çok kolay) yetenek kestiriminde kullanılan modellere (ÇBMTK, ÜDMTK ve Bifaktör) ait RMSE, güvenirlik ve sınıflama doğruluğu değerleri karşılaştırılmıştır. Bu amaçla gerçekleştirilen tekrarlı ANOVA sonuçları Tablo 13'te verilmiştir.

Tablo 13.

Toplam Puan için Korelasyon 0,5 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması

		RMSE		Güvenirlik		Sınıflama doğruluğu	
ÇPMY	TG	F	η_p^2	F	η_p^2	F	η_p^2
%5	Çok zor	292633,750*	,999	30340,653*	,992	125925,375*	,998
	Zor	227371,896*	,999	12722,366*	,981	115825,643*	,998
	Orta güçlük	173301,318*	,999	13308,824*	,982	118737,387*	,998
	Kolay	225545,555*	,999	9135,855*	,974	97244,560*	,997
	Çok kolay	302668,760*	,999	17610,465*	,986	104552,476*	,998
%10	Çok zor	136954,472*	,998	26007,856*	,991	152023,068*	,998
	Zor	156189,011*	,998	17010,593*	,986	127022,018*	,998
	Orta güçlük	163562,455*	,998	20609,028*	,988	120128,517*	,998
	Kolay	173996,188*	,999	19125,223*	,987	134846,004*	,998
	Çok kolay	174544,140*	,999	14710,466*	,983	93592,694*	,997
%25	Çok zor	94559,829*	,997	24096,728*	,990	157284,313*	,998
	Zor	129842,274*	,998	38765,831*	,994	185229,143*	,999
	Orta güçlük	135953,325*	,998	35301,795*	,993	189411,968*	,999
	Kolay	122567,628*	,998	20701,879*	,988	181705,970*	,999
	Çok kolay	127035,249*	,998	18669,044*	,987	134945,196*	,998
%50	Çok zor	222324,931*	,999	20731,584*	,988	67087,497*	,996
	Zor	222220,404*	,999	21367,602*	,989	72288,031*	,997
	Orta güçlük	200153,705*	,999	18088,750*	,987	58967,749*	,996
	Kolay	216923,840*	,999	15443,083*	,984	38317,777*	,994
	Çok kolay	237237,214*	,999	15611,138*	,984	39789,030*	,994

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi; TG: test güçlüğü

Tablo 13'te çoklu puanlanan madde yüzdesi ve test güçlüğünün tüm düzeylerinde RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her birinin kestirim modellerine göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. RMSE değerlerine yönelik elde edilen bu farkın kaynağını belirlemek amacıyla yapılan ikili karşılaştırmalar sonucunda, farkların hepsinin anlamlı olduğu ve RMSE değerlerinin düşükten yükseğe doğru sırasıyla ÇBMTK, ÜDMTK ve Bifaktör modellerinden elde edildiği görülmüştür (Ek 5).

Güvenirlik değerleri için gerçekleştirilen ikili karşılaştırmalarda ise bütün çoklu puanlanan madde yüzdelerinde ve bütün test güçlüklerinde yetenek kestirimine göre hesaplanan güvenirlik değerlerinin yüksekten düşüğe doğru sırasıyla ÇBMTK, ÜDMTK ve Bifaktör modellerden elde edildiği görülmüştür. Bu sıralama RMSE ile elde edilen sıralamanın tersi şeklindedir. Sınıflama doğruluğu değerleri dikkate alındığında ise benzer şekilde çoğu ikili karşılaştırmaların anlamlı olduğu ve sınıflama doğruluğu değerlerinin yüksekten düşüğe doğru sırasıyla ÇBMTK, ÜDMTK ve Bifaktör modellerinden elde edildiği belirlenmiştir. Çoklu puanlanan madde yüzdesi %10 ve test güçlüğü çok kolay olduğunda ÜDMTK ve Bifaktör modellerinden elde

edilen sınıflama doğruluğu değerleri arasındaki fark anlamlı değildir. Ayrıca çoklu puanlanan madde yüzdesi %5 ve test güçlüğü çok zor olduğunda en düşük sınıflama doğruluğu ÜDMTK model ile elde edilmiştir. Benzer şekilde, çoklu puanlanan madde yüzdesi %50 ve test güçlüğü çok kolay olduğunda da en düşük sınıflama doğruluğu ÜDMTK ile elde edilmiştir (Ek 5).

Korelasyon 0,8 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

Korelasyon 0,8 düzeyinde model, çoklu puanlanan madde yüzdesi ve test güçlüğü değişkenleri arasındaki üç yönlü etkileşim etkisinin incelenmesi amacıyla bu değişkenler ile RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için ayrı ayrı faktöriyel karma ANOVA yapılmıştır. Analiz sonuçları Tablo 14'te yer almaktadır.

Tablo 14.

Toplam Puan için Korelasyon 0,8 Düzeyinde Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Faktöriyel Karma ANOVA

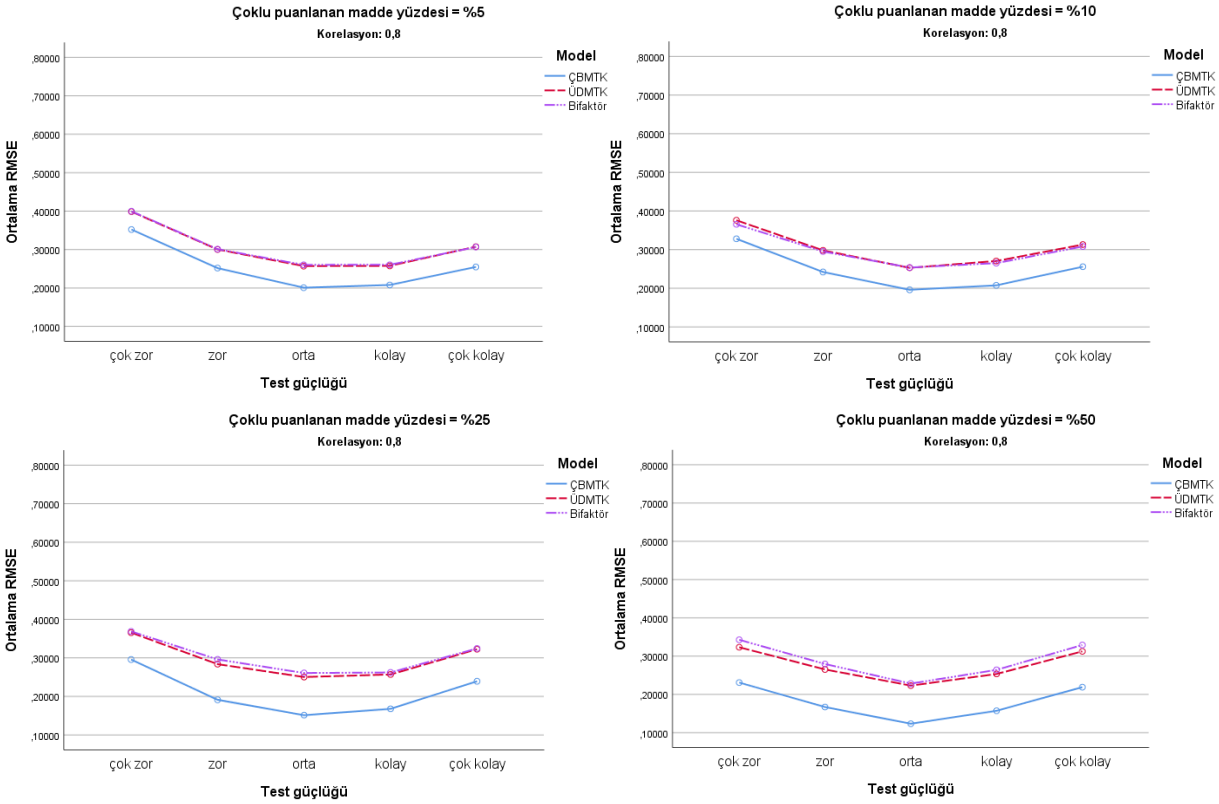
	RMSE		Güvenirlik		Sınıflama doğruluğu	
Kaynak	F	η_p^2	F	η_p^2	F	η_p^2
Gruplar içi						
M	585408,664*	,997	296768,835*	,993	2009835,895*	,999
M *x* ÇPMY	18530,848*	,966	30915,314*	,979	24202,784*	,973
M *x* TG	660,926*	,572	1138,315*	,697	2065,089*	,807
M *x* TG *x* ÇPMY	275,861*	,626	614,513*	,788	543,677*	,767
Gruplar arası						
ÇPMY	20448,422*	,969	40374,429*	,984	36010,824*	,982
TG	127442,201*	,996	77507,535*	,994	74365,321*	,993
ÇPMY *x* TG	2434,785*	,937	4161,608*	,962	1698,906*	,911

*p<,001; M: model; ÇPMY: çoklu puanlanan madde yüzdesi; TG: test güçlüğü

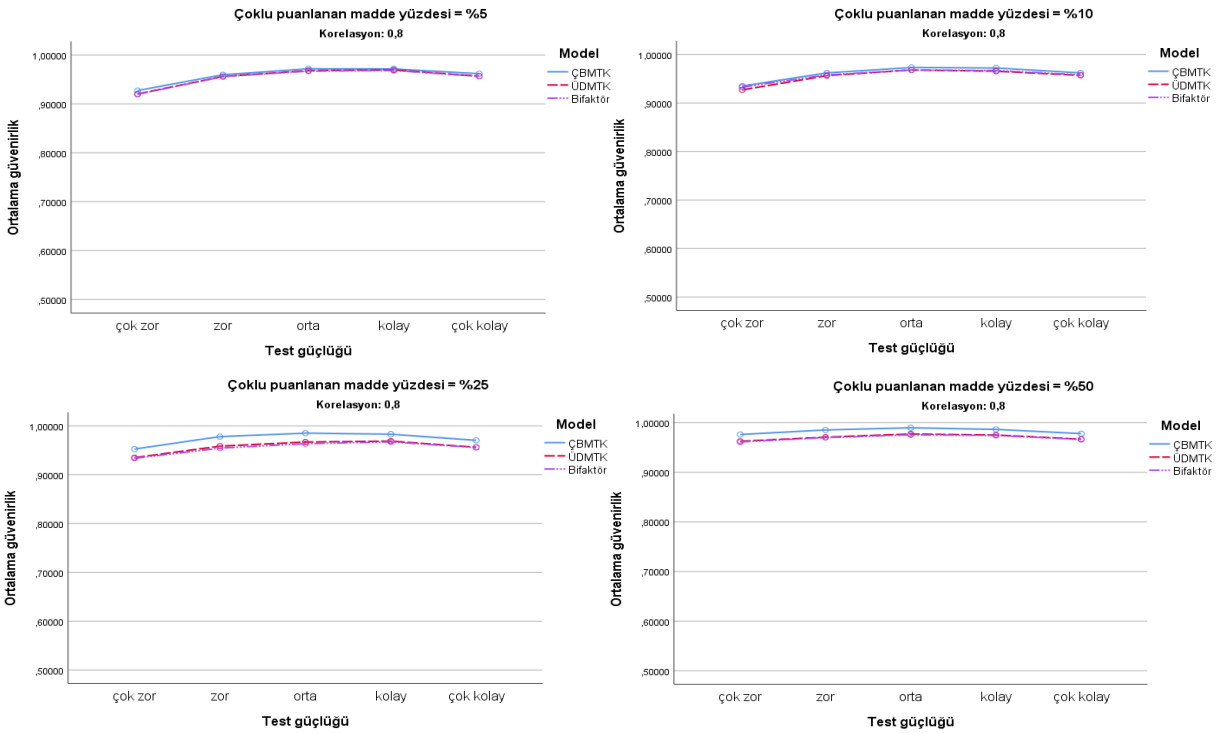
Boyutlar arası korelasyon 0,8 düzeyinde RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için gerçekleştirilen ANOVA sonucunda model, test güçlüğü ve çoklu puanlanan madde yüzdesi değişkenleri arasındaki üç yönlü etkileşimin her üç değerlendirme kriteri için de anlamlı olduğu tespit edilmiştir. Bu üç yönlü etkileşimleri incelemek amacıyla basit iki yönlü etkileşim yaklaşımı kullanılmıştır. Bunun için çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü arasındaki etkileşim etkisi ve ardından test güçlüğünün her bir düzeyinde (çok zor, zor, orta, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki etkileşim etkisi incelenmiştir.

Öncelikle çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü değişkenleri ile RMSE, güvenirlik ve sınıflama doğruluğu değerleri için iki yönlü karma ANOVA yapılmış ve sonuçlar Tablo 15'te sunulmuştur. Çoklu puanlanan madde

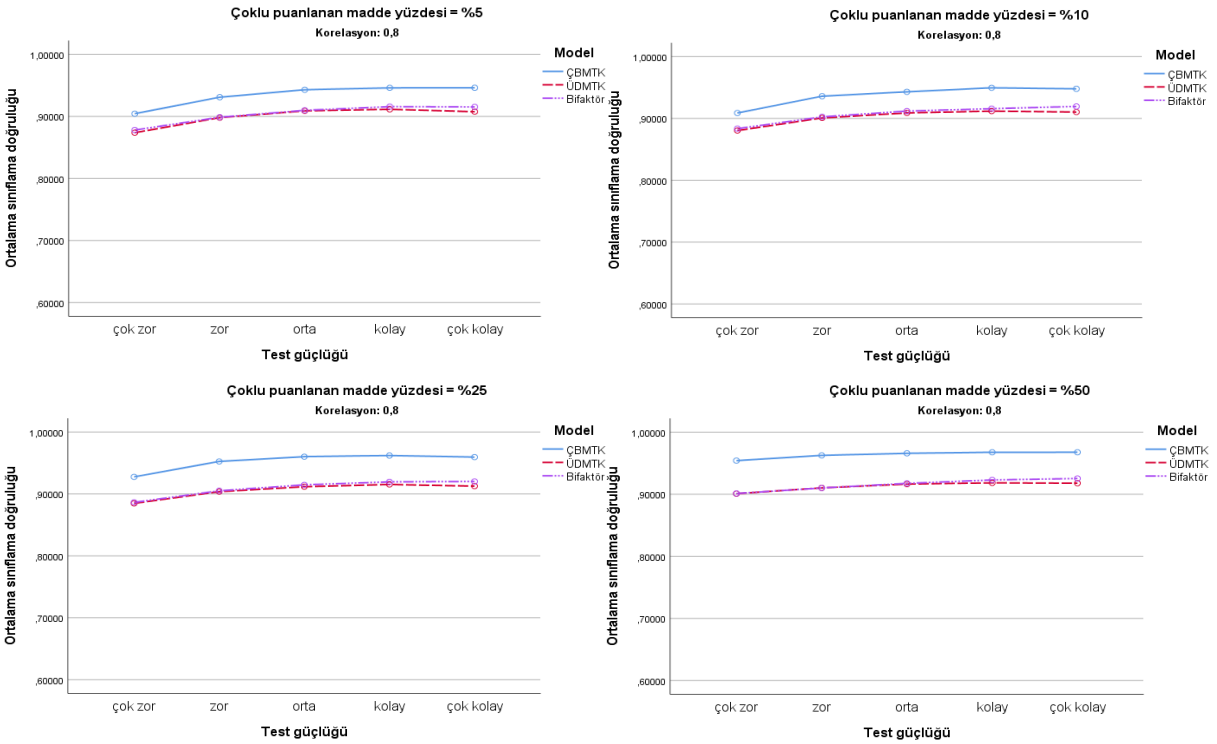
yüzdesinin her bir düzeyi için model ve test güçlüğü düzeylerinde elde edilen ortalama RMSE, güvenirlik ve sınıflama doğruluğu değerlerini içeren grafikler sırasıyla Şekil 9, Şekil 10 ve Şekil 11'de yer almaktadır.



Şekil 9. Toplam puan korelasyon 0,8 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE)



Şekil 10. Toplam puan korelasyon 0,8 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik)



Şekil 11. Toplam puan korelasyon 0,8 – çoklu puanlanan madde yüzdesi x model x test güçlüğü (sınıflama doğruluğu)

Tablo 15.

Toplam Puan için Korelasyon 0,8 Düzeyinde Çoklu Puanlanan Madde Yüzdesinin Her Bir Düzeyi için Model ve Test Güçlüğü Değişkenleri ile İki Yönlü Karma ANOVA Sonuçları

			RMSE		Güvenirlik		Sınıflama doğruluğu	
ÇPMY	Kaynak		F	η_p^2	F	η_p^2	F	η_p^2
%5	Gruplar içi	M	149969,189*	,997	32460,937*	,985	399566,673*	,999
		M *x* TG	217,813*	,638	958,528*	,886	889,488*	,878
	Gruplar arası	TG	50943,937*	,998	34663,376*	,996	21505,265*	,994
%10	Gruplar içi	M	128464,217*	,996	22975,547*	,979	354555,604*	,999
		M *x* TG	500,128*	,802	620,711*	,834	1160,643*	,904
	Gruplar arası	TG	28331,460*	,996	21474,969*	,994	21188,370*	,994
%25	Gruplar içi	M	192052,197*	,997	115410,159*	,996	578086,147*	,999
		M *x* TG	856,704*	,874	807,235*	,867	633,550*	,837
	Gruplar arası	TG	28599,632*	,996	16342,514*	,992	22279,792*	,994
%50	Gruplar içi	M	153862,382*	,997	197205,655*	,997	695914,217*	,999
		M *x* TG	72,863*	,371	565,323*	,820	1037,312*	,893
	Gruplar arası	TG	29725,863*	,996	8651,057*	,986	10261,769*	,988

*p<,001; M: model; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 15'te yer alan boyutlar arası korelasyon 0,8 düzeyi ve çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) gerçekleştirilen analiz sonuçları incelendiğinde model ve test güçlüğü arasındaki etkileşimin RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için anlamlı olduğu görülmektedir. Bu etkileşim etkilerine ait etki büyüklükleri yüksektir. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi için sırasıyla model ve test güçlüğü değişkenlerinin her bir düzeyinde diğer değişkene ait basit ana etkiler tek yönlü ANOVA ile incelenmiştir. İlk olarak test güçlüğü basit ana etkisi incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde test güçlük düzeyleri (çok zor, zor, orta güçlük, kolay, çok kolay) için hesaplanan RMSE, güvenirlik ve sınıflama doğruluğu değerleri ayrı ayrı karşılaştırılmıştır. Test güçlüğü basit ana etkisi için elde edilen sonuçlar Tablo 16'da verilmiştir.

Tablo 16.

Toplam Puan için Korelasyon 0,8 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması

		RMSE		Güvenirlik		Sınıflama doğruluğu	
ÇPMY	Model	F	η_p^2	F	η_p^2	F	η_p^2
%5	ÇBMTK	41206,279*	,997	29710,613*	,996	22322,645*	,994
	ÜDMTK	35861,972*	,997	35003,314*	,996	15590,656*	,992
	Bifaktör	41699,631*	,997	32650,869*	,996	15279,725*	,992
%10	ÇBMTK	29549,057*	,996	23916,605*	,995	21795,281*	,994
	ÜDMTK	17786,343*	,993	19049,586*	,994	12449,548*	,990
	Bifaktör	19406,071*	,994	15470,530*	,992	14761,995*	,992
%25	ÇBMTK	40059,261*	,997	29287,729*	,996	22748,476*	,995
	ÜDMTK	15025,914*	,992	10990,129*	,989	11758,069*	,990
	Bifaktör	12136,371*	,990	9183,880*	,987	13885,058*	,991
%50	ÇBMTK	18190,349*	,993	12559,740*	,990	4185,836*	,971
	ÜDMTK	12879,049*	,990	6233,332*	,981	5800,096*	,979
	Bifaktör	11987,638*	,990	5487,120*	,978	8471,220*	,986

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 16'da çoklu puanlanan madde yüzdesi ve kestirim modellerinin tüm düzeylerinde RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her birinin test güçlüklerine göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. Anlamlı bulunan bu farkların kaynağının belirlenmesi için ikili karşılaştırmalar yapılmıştır (Ek 6). RMSE değerleri açısından sonuçlar incelendiğinde genel olarak RMSE değerleri düşükten yükseğe doğru sırasıyla test güçlüğünün orta, kolay, zor, çok kolay ve çok zor olması durumunda elde edilmiştir. Sadece çoklu puanlanan madde yüzdesi %5 olduğunda ÜDMTK ve Bifaktör modellerinde ve çoklu puanlanan madde yüzdesi %25 olduğunda Bifaktör modelinde test güçlüğü orta ve kolay olduğunda elde edilen RMSE değerleri arasındaki fark anlamlı değildir.

Güvenirlik değerleri için yapılan ikili karşılaştırmalarda, çoklu puanlanan madde yüzdesi %50 olduğunda modellerin her üçünde de güvenirlik değerleri yüksekten düşüğe doğru test güçlüğünün orta, kolay, zor, çok kolay ve çok zor olması durumunda elde edilmiştir. Diğer çoklu puanlanan madde yüzdelerinde de sıralama benzer olmakla birlikte bazı farklılıklar yer almaktadır. Çoklu puanlanan madde yüzdesi %5 ve %25 olduğunda ÜDMTK ve Bifaktör modeller için en yüksek güvenirlik test kolay olduğunda elde edilmiştir. Çoklu puanlanan madde yüzdesi %10 olduğunda bütün modeller için sıralama aynı olup test güçlüğü zor ve çok kolay olduğunda elde edilen güvenirlikler arasındaki fark anlamlı değildir. Çoklu puanlanan madde yüzdesi %5 olduğunda ÜDMTK ve Bifaktör modeller için test güçlüğü zor ve çok kolay olduğunda elde edilen güvenirlikler arasındaki fark anlamlı değildir. Bütün çoklu puanlanan

madde yüzdelerinde ve bütün modeller için en düşük güvenirlik ise test çok zor olduğunda elde edilmiştir.

Sınıflama doğruluğu ile ilgili sonuçlar ise ÇBMTK için çoklu puanlanan madde yüzdesi %10 olduğunda sınıflama doğrulukları yüksekten düşüğe doğru test güçlüğünün kolay, çok kolay, orta, zor ve çok zor olması durumlarında elde edilmiştir. Çoklu puanlanan madde yüzdesi %5 ve %50 için de sıralama aynı fakat test güçlüğü kolay ve çok kolay ile elde edilen sınıflama doğrulukları arasındaki fark anlamlı değildir. ÜDMTK için de sonuçlar benzerlik göstermekte ve %10, %25 ve %50 çoklu puanlanan madde yüzdelerinde sınıflama doğruluklarının yüksekten düşüğe doğru elde edildikleri test güçlük düzeyleri sırasıyla kolay, çok kolay, orta, zor ve çok zor şeklindedir. ÇBMTK ile benzer şekilde ÜDMTK için de %50 çoklu puanlanan madde yüzdesi için kolay ve çok kolay test güçlüklerinde elde edilen değerler arasındaki fark anlamlı değildir. Bifaktör model için ise %10 ve %50 çoklu puanlanan madde yüzdelerinde elde edilen sınıflama doğrulukları test güçlük düzeylerine göre yüksekten düşüğe doğru sırasıyla çok kolay, kolay, orta, zor ve çok zor şeklindedir. Çoklu puanlanan madde yüzdesi %5 ve %25 için ise kolay ve çok kolay test güçlüklerinde elde edilen sınıflama doğruluğu değerleri arasındaki fark anlamlı değildir.

Özetle, genel olarak test güçlüğü orta olduğunda en düşük RMSE değerleri elde edilirken test kolaylaştıkça ve zorlaştıkça RMSE değerleri artış göstermiştir. Güvenirlik değerleri ise orta güçlükte en yüksek iken çok zor ve çok kolay test güçlüklerine doğru düşüş göstermiştir. Sınıflama doğruluğu değerleri ele alındığında test güçlüğü kolaylaştıkça sınıflamanın daha doğru olduğu ve zorlaştıkça sınıflama doğruluğu değerlerinin düşüş gösterdiği görülmektedir.

Boyutlar arası korelasyon 0,8 için çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü arasındaki etkileşim etkisine ilişkin test güçlüğü basit ana etkilerinin incelenmesinin ardından modellerin basit ana etkileri ele alınmıştır. Bir başka ifadeyle test güçlüğüne ait düzeylerin her birinde (çok zor, zor, orta güçlük, kolay, çok kolay) yetenek kestiriminde kullanılan modellere (ÇBMTK, ÜDMTK ve Bifaktör) ait RMSE, güvenirlik ve sınıflama doğruluğu değerleri karşılaştırılmıştır. Bu amaçla gerçekleştirilen tekrarlı ANOVA sonuçları Tablo 17'de verilmiştir.

Tablo 17.

Toplam Puan için Korelasyon 0,8 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması

		RMSE		Güvenirlik		Sınıflama doğruluğu	
ÇPMY	TG	F	η_p^2	F	η_p^2	F	η_p^2
%5	Çok zor	25724,769*	,990	12243,314*	,980	45369,945*	,995
	Zor	27766,170*	,991	3089,211*	,926	59022,509*	,996
	Orta güçlük	39476,650*	,994	4418,575*	,947	62776,015*	,996
	Kolay	30836,019*	,992	1943,368*	,887	58513,270*	,996
	Çok kolay	31455,411*	,992	6083,443*	,961	68691,493*	,996
%10	Çok zor	14625,785*	,983	11029,857*	,978	36151,349*	,993
	Zor	23716,496*	,990	3554,707*	,935	57490,990*	,996
	Orta güçlük	26395,353*	,991	2993,662*	,924	53004,693*	,995
	Kolay	28864,213*	,992	4828,725*	,951	64095,395*	,996
	Çok kolay	23825,939*	,990	2782,392*	,918	60629,396*	,995
%25	Çok zor	18207,923*	,987	14415,813*	,983	65213,761*	,996
	Zor	35339,669*	,993	21321,798*	,989	85075,344*	,997
	Orta güçlük	39221,832*	,994	17986,509*	,986	82841,860*	,997
	Kolay	30218,193*	,992	9773,981*	,975	76117,123*	,997
	Çok kolay	25366,781*	,990	9015,947*	,973	76196,803*	,997
%50	Çok zor	32339,479*	,992	28590,780*	,991	121234,476*	,998
	Zor	34478,391*	,993	31181,459*	,992	117685,456*	,998
	Orta güçlük	32810,522*	,993	24843,484*	,990	103760,841*	,998
	Kolay	32145,401*	,992	19243,949*	,987	97946,048*	,997
	Çok kolay	32240,835*	,992	18416,770*	,987	97833,320*	,997

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi; TG: test güçlüğü

Tablo 17'de çoklu puanlanan madde yüzdesi ve test güçlüğünün tüm düzeylerinde RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her birinin kestirim modellerine göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. RMSE değerlerine yönelik elde edilen bu farkın kaynağını belirlemek amacıyla yapılan ikili karşılaştırmalar sonucunda, farkların anlamlı olduğu ve analizlerde çoklu puanlanan madde yüzdesi %5, %25 ve %50 olduğunda RMSE değerlerinin düşükten yükseğe doğru sırasıyla ÇBMTK, ÜDMTK ve Bifaktör modellerinden elde edildiği görülmüştür. Çoklu puanlanan madde yüzdesi %10 için ise bu sıralama ÇBMTK, Bifaktör ve ÜDMTK şeklindedir. Çoklu puanlanan madde yüzdesi %5 iken test güçlüğü zor ve çoklu puanlanan madde yüzdesi %10 iken test güçlüğü orta güçlük olduğunda ÜDMTK ve Bifaktör model ile elde edilen RMSE değerleri arasındaki fark anlamlı değildir (Ek 6).

Güvenirlik değerleri ile yapılan analizlerde ise güvenirlik değerleri yüksekten düşüğe doğru kestirim modellerine göre ÇBMTK, ÜDMTK ve Bifaktör ile elde edilmiştir. Tek istisna, %10 çoklu puanlanan madde yüzdesi ve test güçlüğü orta iken ÜDMTK ve Bifaktör ile elde edilen güvenirlik değerleri arasındaki farkın anlamlı olmamasıdır.

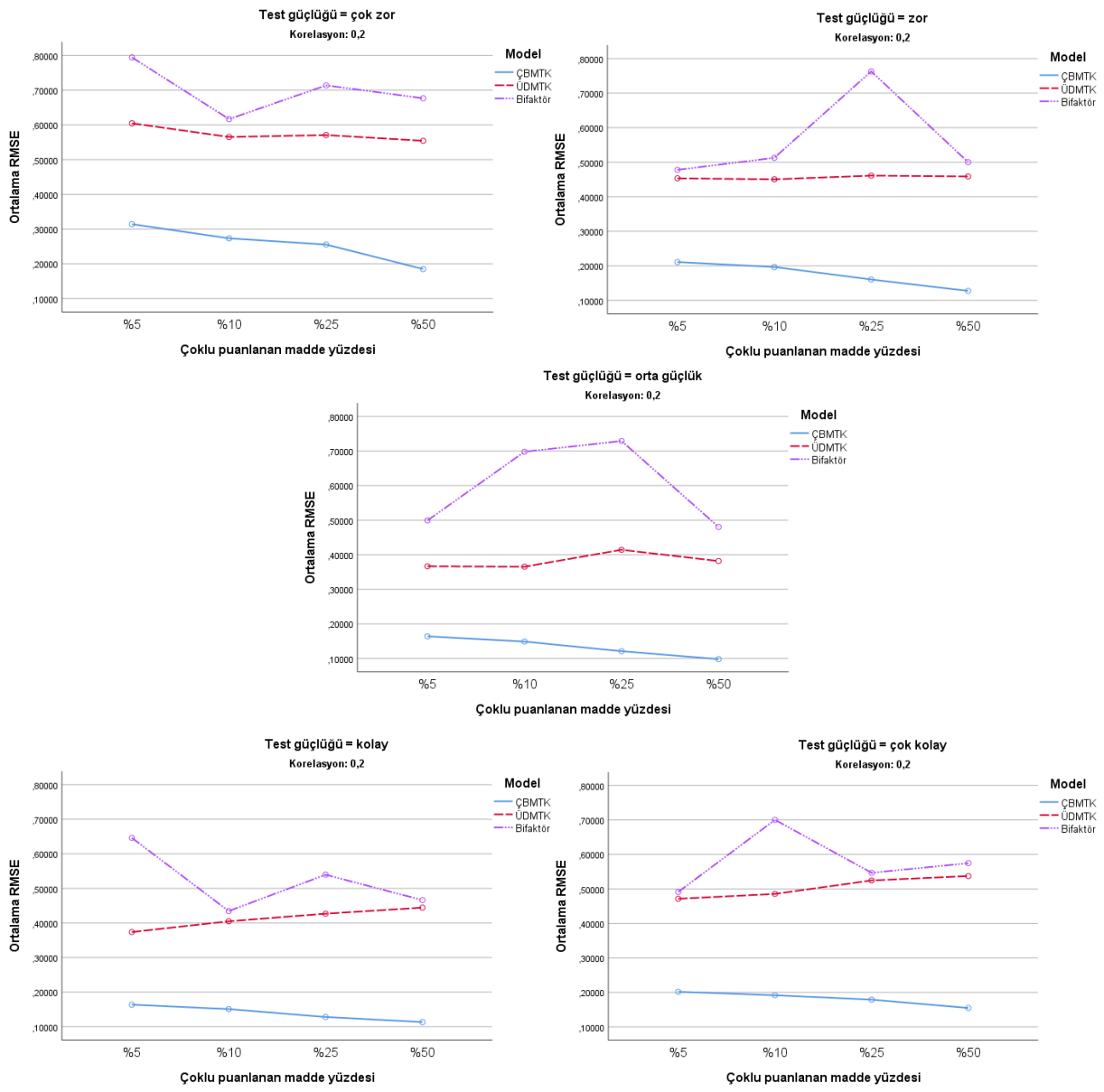
Sınıflama doğruluğu açısından da ikili karşılaştırmalara göre sınıflama doğruluğu değerleri yüksekten düşüğe doğru sırasıyla ÇBMTK, Bifaktör ve ÜDMTK modelleri ile elde edilmiştir. Sadece çoklu puanlanan madde yüzdesi %50 iken test güçlüğü çok zor ve zor olduğu durumlarda elde edilen sınıflama doğruluğu değerleri arasındaki fark anlamlı değildir.

4.1.2. Çoklu Puanlanan Madde Yüzdesi Değişkeni için İncelemeler

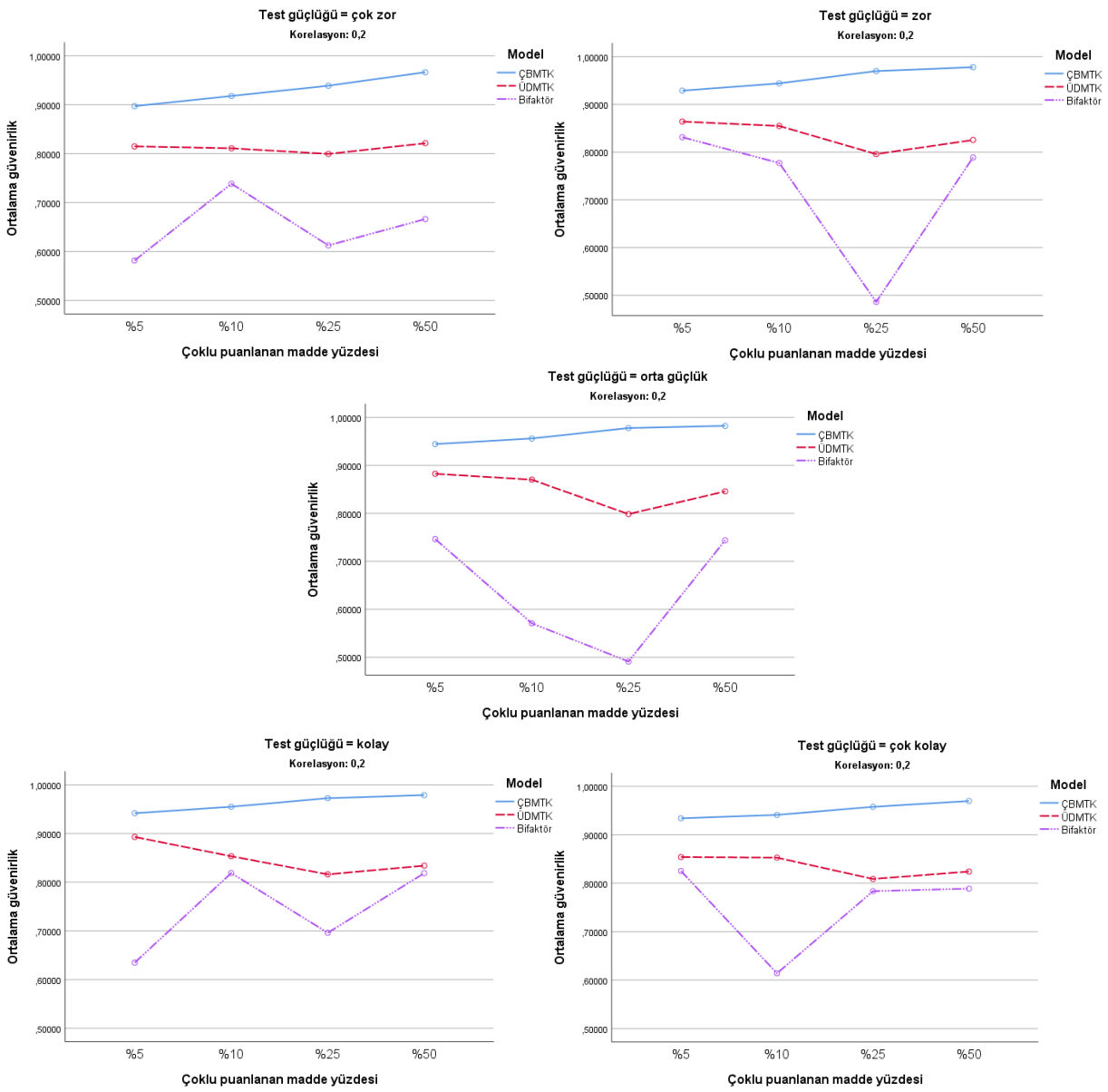
Korelasyonun bütün düzeylerinde çoklu puanlanan madde yüzdesi, test güçlüğü ve model arasındaki etkileşim etkileri anlamlı bulunmuştur. Bunun için öncelikle Bölüm 5.1.1'de çoklu puanlanan madde yüzdesinin her bir düzeyi için test güçlüğü ve model arasındaki etkileşim incelenmiştir. Test güçlüğü ve model basit ana etkileri yorumlanmıştır. Daha sonra ise yine bu üç yönlü etkileşimin incelenmesi için test güçlüğünün her bir düzeyinde (çok zor, zor, orta güçlük, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki iki yönlü etkileşim incelenmiştir. Burada amaç çoklu puanlanan madde yüzdesi düzeyleri arasındaki farkları ortaya koyabilmektir. Bir başka ifadeyle bu bölümde gerçekleştirilen analizlerde simülasyon veride toplam puan açısından çoklu puanlanan madde yüzdesinin ele alınan değerlendirme kriterleri açısından yetenek kestirimine olan etkisi ortaya konmuştur.

Korelasyon 0,2 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

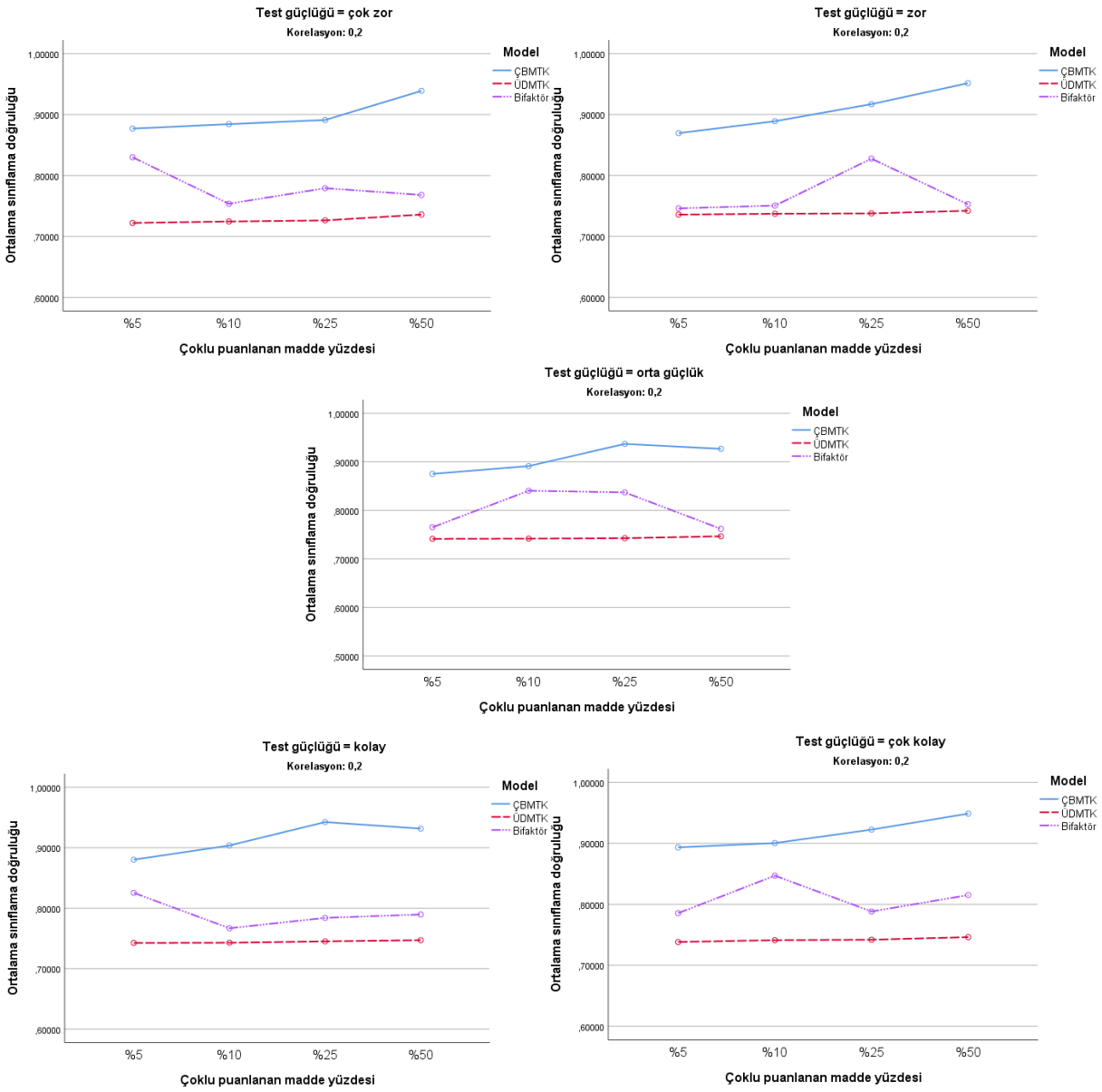
Boyutlar arası korelasyon 0,2 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki etkileşim etkisi incelenmiştir. Bu amaçla model ve test güçlüğü değişkenleri ile RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin karşılaştırılması için iki yönlü karma ANOVA yapılmış ve sonuçlar Tablo 18'de sunulmuştur. Test güçlüğünün her bir düzeyi için model ve çoklu puanlanan madde yüzdesi düzeylerinde elde edilen ortalama RMSE, güvenirlik ve sınıflama doğruluğu değerlerini içeren grafikler sırasıyla Şekil 12, Şekil 13 ve Şekil 14'te de yer almaktadır.



Şekil 12. Toplam puan korelasyon 0,2 – test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE)



Şekil 13. Toplam puan korelasyon 0,2 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik)



Şekil 14. Toplam puan korelasyon 0,2 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (sınıflama doğruluğu)

Tablo 18.

Toplam Puan için Korelasyon 0,2 iken Test Güçlüğünün Her Bir Düzeyi için Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile İki Yönlü Karma ANOVA

			RMSE		Güvenirlik		Sınıflama doğruluğu	
TG	Kaynak		F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	Gruplar içi	M	213309,796*	,998	50667,438*	,992	253625,237*	,998
		M *x* ÇPMY	1587,432*	,923	1077,895*	,891	6241,678*	,979
	Gruplar arası	TG	5780,578*	,978	1242,858*	,904	3136,690*	,960
Zor	Gruplar içi	M	329089,620*	,999	96823,955*	,996	476512,083*	,999
		M *x* ÇPMY	13795,323*	,991	14477,286*	,991	11628,086*	,989
	Gruplar arası	TG	8862,480*	,985	12312,690*	,989	11849,483*	,989
Orta güçlük	Gruplar içi	M	144030,191*	,997	74451,298*	,995	163377,199*	,998
		M *x* ÇPMY	3693,869*	,965	3617,419*	,965	4840,033*	,973
	Gruplar arası	TG	4603,402*	,972	4478,041*	,971	6411,685*	,980
Kolay	Gruplar içi	M	150538,507*	,997	46555,425*	,992	233510,146*	,998
		M *x* ÇPMY	3798,560*	,966	3484,344*	,963	4383,731*	,971
	Gruplar arası	TG	1975,089*	,937	1945,137*	,936	1583,534*	,923
Çok kolay	Gruplar içi	M	344863,055*	,999	66333,310*	,994	532638,709*	,999
		M *x* ÇPMY	6452,861*	,980	5822,398*	,978	7707,202*	,983
	Gruplar arası	TG	4064,980*	,969	3194,465*	,960	7899,700*	,984

*p<,001; M: model; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 18'de yer alan boyutlar arası korelasyon 0,2 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta güçlük, kolay, çok kolay) gerçekleştirilen analiz sonuçları incelendiğinde model ve çoklu puanlanan madde yüzdesi arasındaki etkileşimin RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için anlamlı ve etki büyüklüklerinin yüksek olduğu görülmektedir. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi için kestirim modellerinin her biri için çoklu puanlanan madde yüzdesi basit ana etkisi tek yönlü ANOVA ile incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde çoklu puanlanan madde yüzdeleri (%5, %10, %25 ve %50) için hesaplanan RMSE, güvenirlik ve sınıflama doğruluğu değerleri ayrı ayrı karşılaştırılmıştır.

Çoklu puanlanan madde yüzdesi basit ana etkisi için elde edilen sonuçlar Ek 7'de verilmiştir. ANOVA çıktıları, her bir test güçlük düzeyinde bütün modeller için çoklu puanlanan madde yüzdelerine göre elde edilen RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin birbirinden anlamlı olarak farklılaştığını göstermektedir. Farkın kaynağına ilişkin yapılan ikili karşılaştırmalar ve ortalama RMSE değerleri de Ek 7'de sunulmaktadır. Buna göre bütün test güçlüklerinde ÇBMTK için elde edilen RMSE değerleri yüksekten düşüğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. ÜDMTK için elde edilen RMSE değerleri arasındaki ortalama farklar genel olarak diğer modellere göre daha azdır. Test güçlüğü kolay ve çok kolay olduğunda ÜDMTK ile elde edilen RMSE değerleri çoklu puanlanan madde yüzdesi arttıkça artış göstermiştir. Bifaktör model ile elde

edilen RMSE değerleri için ise anlamlı bulunan farkların çoklu puanlanan madde yüzdesi düzeylerinin değişimine göre belli bir eğilime sahip olmadığı görülmüştür.

Güvenirlik değerleri açısından bütün test güçlüklerinde ÇBMTK için elde edilen güvenirlik değerleri en düşükten en yükseğe doğru çoklu puanlanan madde yüzdesinin sırasıyla %5, %10, %25 ve %50 olduğu durumda elde edilmiştir. ÜDMTK için ise elde edilen güvenirlik değerleri arasındaki ortalama farklar genel olarak diğer modellere göre daha azdır. ÜDMTK için test güçlüğü çok zor olduğunda güvenirlik değerleri en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdesine göre %50, %5, %10 ve %25 şeklinde iken diğer test güçlük düzeylerinde bu sıralama %5, %10, %50 ve %25 şeklindedir. Ancak test güçlüğü çok kolay iken çoklu puanlanan madde yüzdesi %5 ve % 10 için elde edilen güvenirlikler arasındaki fark anlamlı değildir. Bifaktör model ile elde edilen güvenirlik değerlerinin en yüksek ve en düşük olduğu çoklu puanlanan madde yüzdeleri ve sıralamaları test güçlüğünün her bir düzeyinde farklıdır. Test güçlüğü orta düzey iken çoklu puanlanan madde yüzdesinin %5 ile %50, kolay iken %10 ile %50 ve çok kolay iken %25 ile %50 olduğu düzeyler arasındaki farklar manidar değildir.

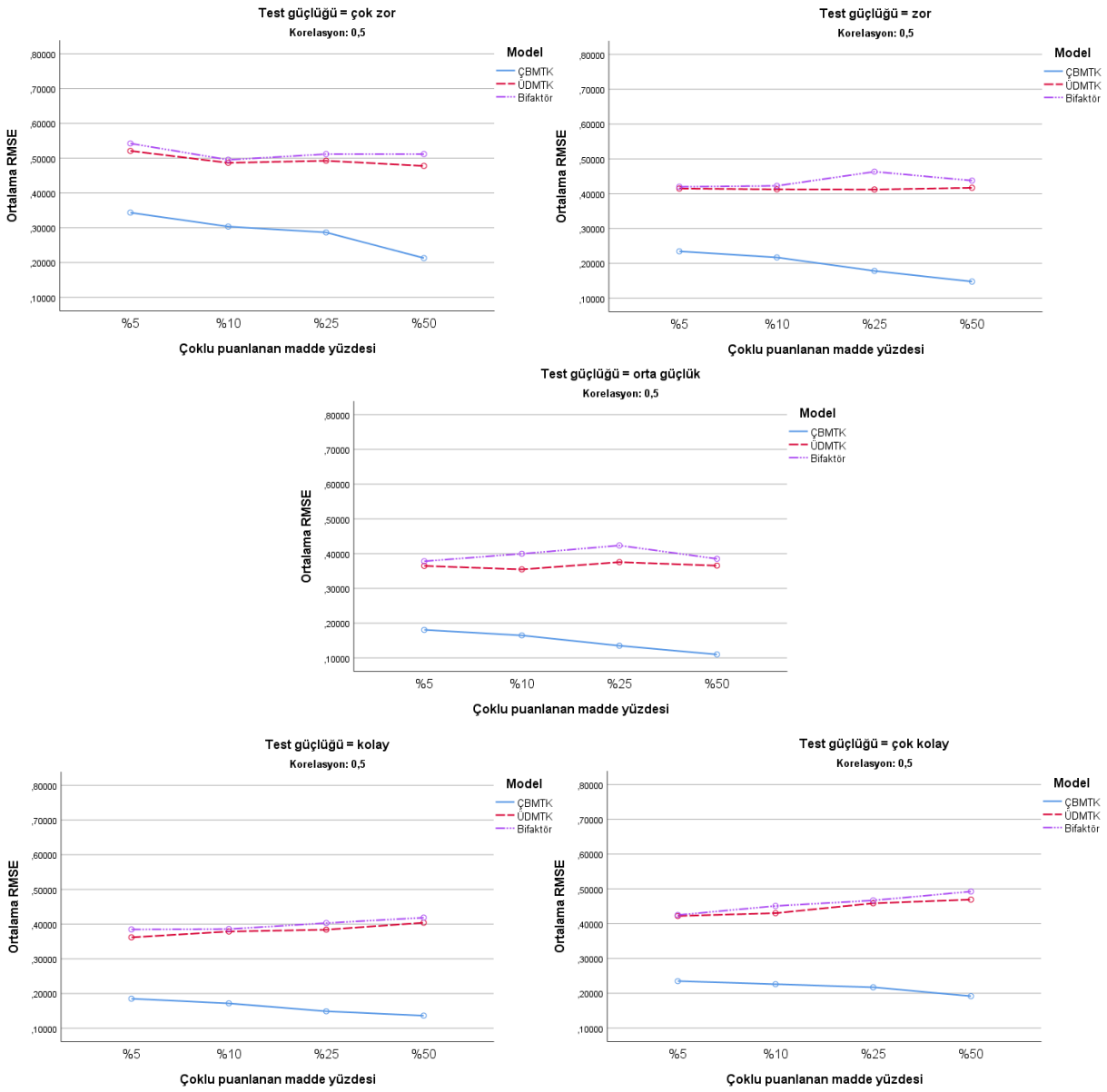
Sınıflama doğruluğu değerlerine yönelik gerçekleştirilen analizlerde ÇBMTK için yapılan bütün ikili karşılaştırmalar anlamlıdır. Test güçlüğünün çok zor, zor ve çok kolay olması durumunda elde edilen sınıflama doğrulukları çoklu puanlanan madde yüzdelerine göre en yüksekten en düşüğe doğru %50, %25, %10 ve %5 şeklinde sıralanırken test güçlüğü kolay ve orta düzeyde ise bu sıralama %25, %50, %10 ve %5 şeklindedir. ÜDMTK için elde edilen sınıflama doğruluğu değerleri çoklu puanlanan madde yüzdesi düzeylerine göre en yüksekten en düşüğe doğru sırasıyla %50, %25, %10 ve %5 olduğu durumlarda elde edilmektedir. Fakat ÜDMTK için bazı durumlarda ikili karşılaştırmalar anlamlı değildir. Test güçlüğü zor ve çok kolay olduğunda çoklu puanlanan madde yüzdesi %25 ve %10, test güçlüğü orta ve kolay olduğunda ise çoklu puanlanan madde yüzdesi %10 ve %5 için elde edilen sınıflama doğruluğu değerleri arasındaki fark anlamlı değildir. Güvenirlik ve RMSE değerleri ile benzer şekilde Bifaktör model için elde edilen sınıflama doğruluklarının çoklu puanlanan madde yüzdelerine göre belli bir eğilimde olmadığı sonucuna ulaşılmıştır.

Özetle, ÇBMTK modelinde test güçlüklerinin ne olduğu fark etmeksizin çoklu puanlanan madde yüzdesi arttıkça RMSE değerleri azalırken güvenirlik ve sınıflama doğruluğu değerleri artış göstermektedir. ÜDMTK modelinde ise çoklu puanlanan madde yüzdesinin değişimine göre elde edilen RMSE değerleri arasındaki farklılıklar belli bir eğilim göstermemektedir. Sınıflama doğruluğu, çoklu puanlanan madde yüzdesi arttıkça yükselmekte ve güvenirlik

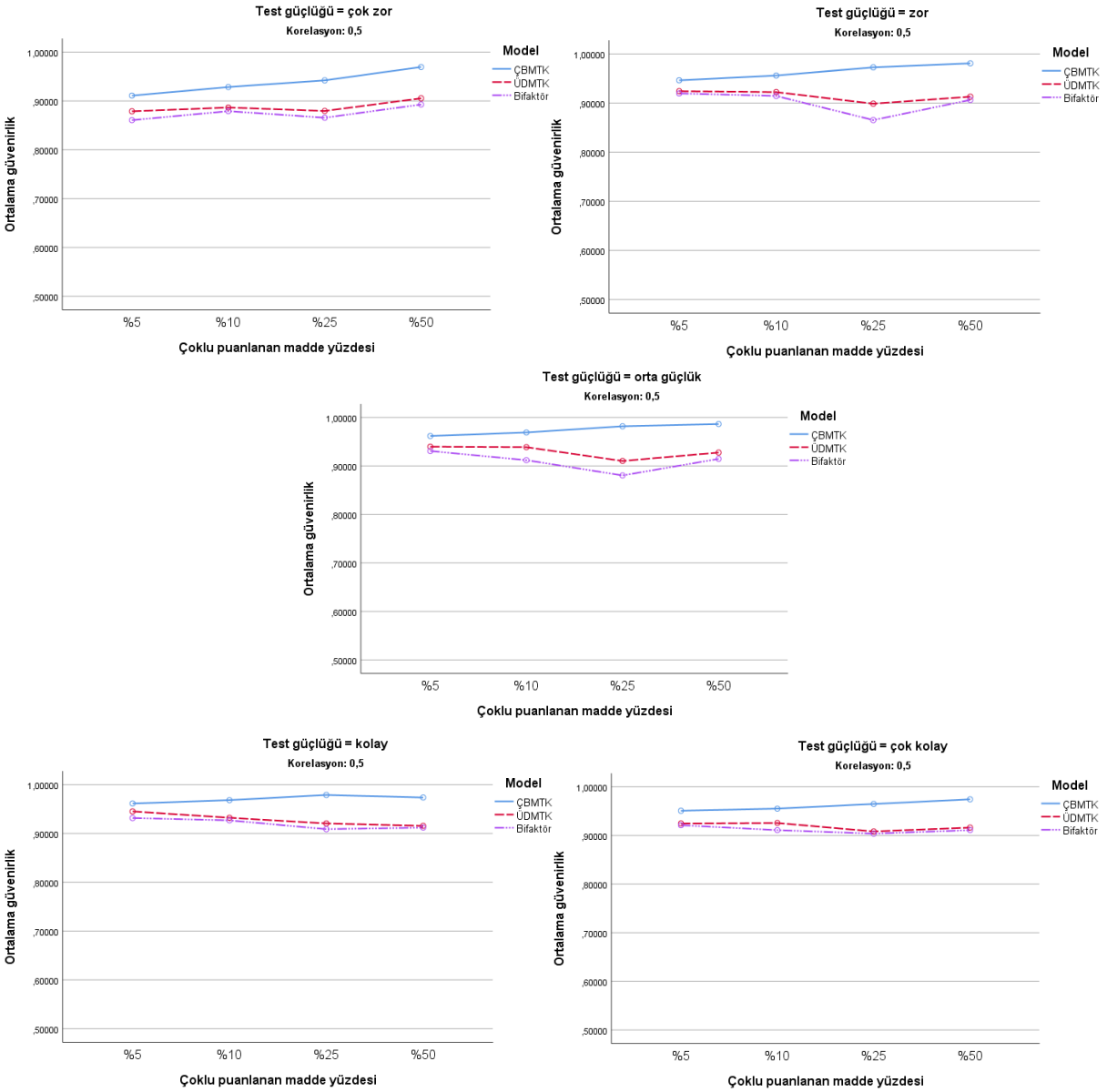
değerleri test güçlüğünün çok zor olduğu durum dışında çoklu puanlanan madde yüzdesi fazla olduğu durumlarda daha düşüktür. Öte yandan Bifaktör model ile elde edilen RMSE, güvenirlik ve sınıflama doğruluğu değerleri için çoklu puanlanan madde yüzdesinin değişimine göre elde edilen farklar arasında ortak bir eğilim bulunmamaktadır.

Korelasyon 0,5 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

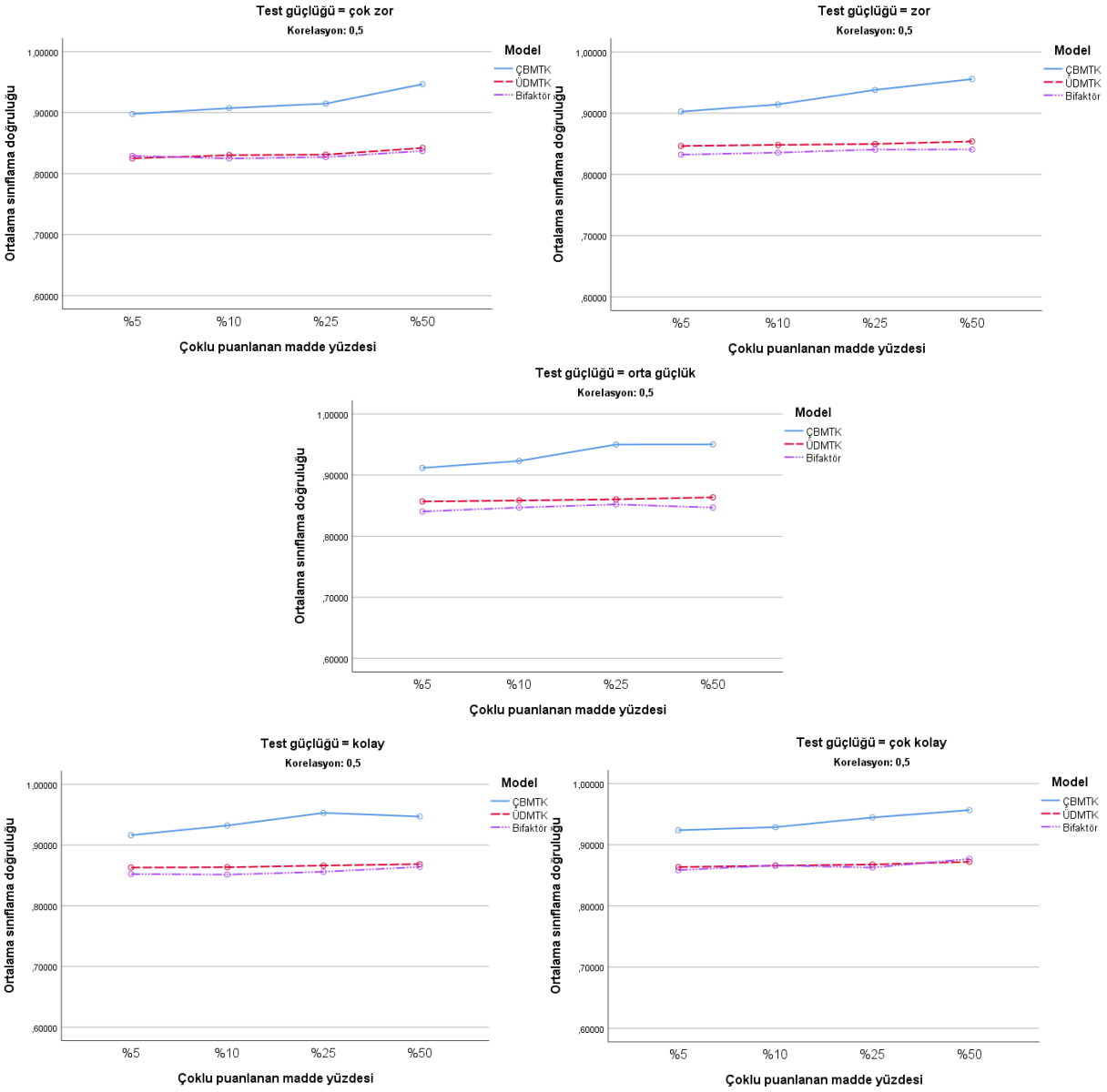
Boyutlar arası korelasyon 0,5 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki etkileşim etkisi incelenmiştir. Bu amaçla model ve test güçlüğü değişkenleri ile RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin karşılaştırılması için iki yönlü karma ANOVA yapılmış ve sonuçlar Tablo 19'da sunulmuştur. Test güçlüğünün her bir düzeyi için model ve çoklu puanlanan madde yüzdesi düzeylerinde elde edilen ortalama RMSE, güvenirlik ve sınıflama doğruluğu değerlerini içeren grafikler sırasıyla Şekil 15, Şekil 16 ve Şekil 17'de yer almaktadır.



Şekil 15. Toplam puan korelasyon 0,5 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE)



Şekil 16. Toplam puan korelasyon 0,5 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik)



Şekil 17. Toplam puan korelasyon 0,5 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (sınıflama doğruluğu)

Tablo 19.

Toplam Puan için Korelasyon 0,5 iken Test Güçlüğünün Her Bir Düzeyi İçin Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile İki Yönlü Karma ANOVA

			RMSE		Güvenirlik		Sınıflama doğruluğu	
TG	Kaynak		F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	Gruplar içi	M	523094,186*	,999	100972,214*	,996	516501,530*	,999
		M *x* ÇPMY	5589,420*	,977	1907,833*	,935	4447,898*	,971
	Gruplar arası	TG	7327,642*	,982	4783,589*	,973	10433,999*	,988
Zor	Gruplar içi	M	582178,121*	,999	117863,397*	,997	726196,084*	,999
		M *x* ÇPMY	6682,859*	,981	9593,375*	,986	10501,288*	,988
	Gruplar arası	TG	1159,888*	,898	2890,651*	,956	11077,439*	,988
Orta güçlük	Gruplar içi	M	721083,337*	,999	119852,047*	,997	690416,370*	,999
		M *x* ÇPMY	5371,151*	,976	6725,655*	,981	7194,156*	,982
	Gruplar arası	TG	1895,572*	,935	3262,144*	,961	10902,257*	,988
Kolay	Gruplar içi	M	932887,962*	1,000	58225,982*	,993	214315,674*	,998
		M *x* ÇPMY	6566,116*	,980	2525,069*	,950	1806,462*	,932
	Gruplar arası	TG	290,995*	,688	1,435	-	686,946*	,839
Çok kolay	Gruplar içi	M	731311,886*	,999	144311,912*	,997	836213,338*	1,000
		M *x* ÇPMY	6431,916*	,980	4727,165*	,973	5147,444*	,975
	Gruplar arası	TG	1505,214*	,919	470,267*	,781	7643,987*	,983

*p<,001; M: model; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 19'da yer alan boyutlar arası korelasyon 0,5 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta güçlük, kolay, çok kolay) gerçekleştirilen analiz sonuçları incelendiğinde model ve çoklu puanlanan madde yüzdesi arasındaki etkileşimin RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için anlamlı ve etki büyüklüğünün yüksek olduğu görülmektedir. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi amacıyla kestirim modellerinin her biri için çoklu puanlanan madde yüzdesi basit ana etkisi tek yönlü ANOVA ile incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde çoklu puanlanan madde yüzdeleri (%5, %10, %25 ve %50) için hesaplanan RMSE, güvenirlik ve sınıflama doğruluğu değerleri ayrı ayrı karşılaştırılmıştır.

Çoklu puanlanan madde yüzdesine ait basit ana etkiler için elde edilen sonuçlar Ek 8'de verilmiştir. ANOVA çıktıları, bütün test güçlük düzeyleri altında her bir model için çoklu puanlanan madde yüzdelerine göre elde edilen RMSE değerlerinin birbirinden anlamlı olarak farklılaştığını göstermektedir. Farkın kaynağına ilişkin yapılan ikili karşılaştırmalar ve ortalama RMSE değerleri Ek 8'de sunulmaktadır. Buna göre bütün test güçlüklerinde ÇBMTK için elde edilen RMSE değerleri çoklu puanlanan madde yüzdesi değişkenine göre en yüksekten en düşüğe doğru sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. ÜDMTK model için

test çok zor olduğunda çoklu puanlanan madde yüzdeleri arasında yapılan karşılaştırma sonucu ÇBMTK için elde edilen sonuçlarla aynı olup en yüksekten en düşüğe doğru RMSE değerlerinin sıralaması çoklu puanlanan madde yüzdesine göre %5, %10, %25 ve %50 şeklindedir. Test zor olduğunda ÜDMTK ile elde edilen RMSE değerleri çoklu puanlanan madde yüzdesinin her bir düzeyinde benzerlik göstermektedir. Test orta güçlükte olduğunda ÜDMTK ve Bifaktör modelleri için en yüksek RMSE değeri çoklu puanlanan madde yüzdesi %25 olduğunda elde edilmiştir. Test kolay ve çok kolay olduğunda ise ÜDMTK ve Bifaktör modellerden elde edilen ikili karşılaştırma sonuçları benzer olup hesaplanan RMSE değerleri en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdesi düzeylerine göre %50, %25, %10 ve %5 şeklindedir. Bu bulgu, ÇBMTK için elde edilen bulgunun tam tersidir.

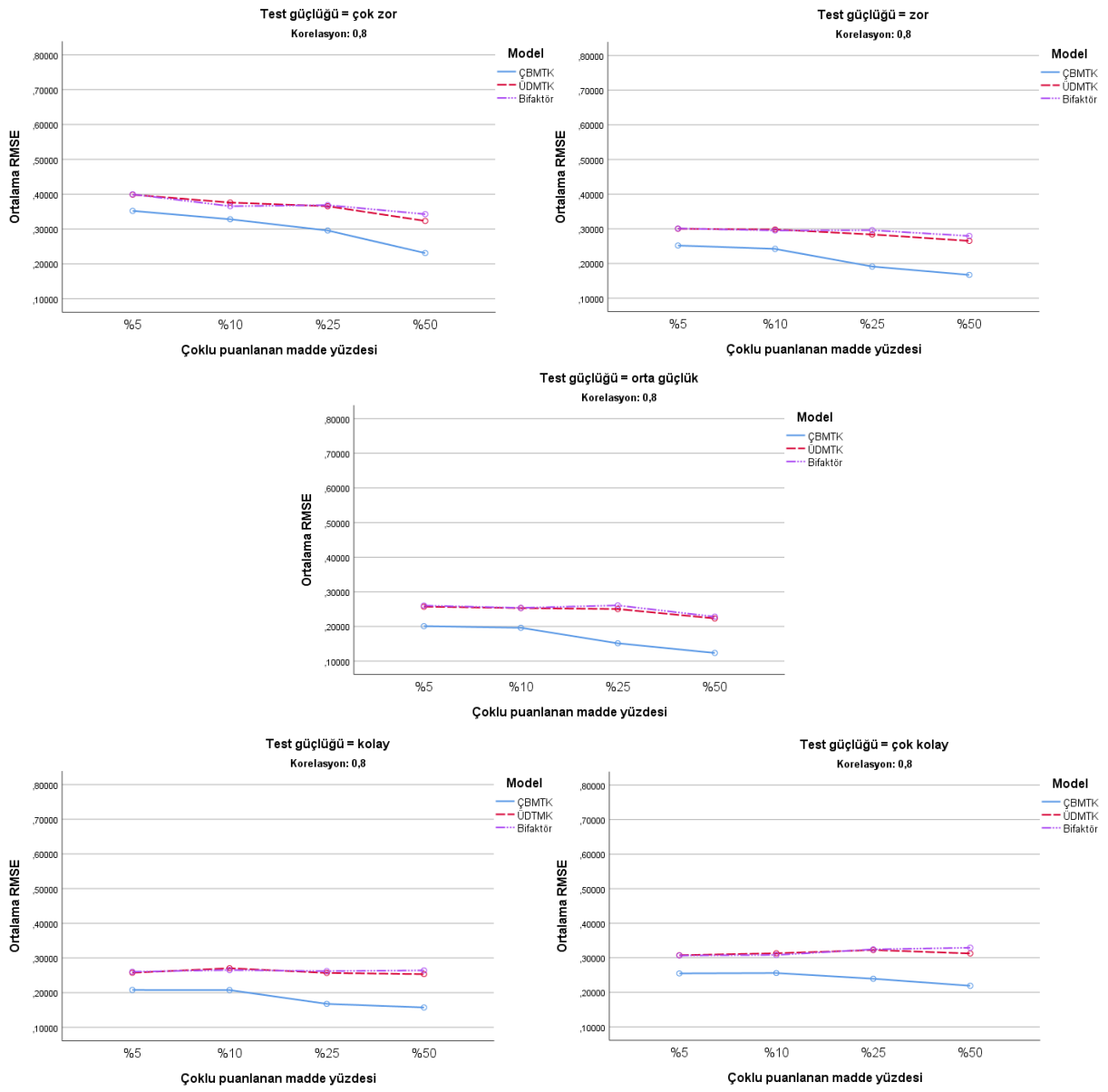
Güvenirlik değerleri için gerçekleştirilen analizler incelendiğinde test güçlük düzeyi kolay olduğunda ÇBMTK ve Bifaktör model için çoklu puanlanan madde yüzdelerine göre elde edilen güvenirlik değerlerinin birbirinden anlamlı olarak farklılaşmadığı görülmektedir. Diğer test güçlük düzeylerinde ise çoklu puanlanan madde yüzdesi düzeylerine göre güvenirlik değerleri farklılaşmaktadır. Farkın kaynağına ilişkin yapılan ikili karşılaştırmalar ve ortalama güvenirlik değerleri Ek 8'de sunulmaktadır. Buna göre kolay hariç bütün test güçlük düzeylerinde ÇBMTK için elde edilen güvenirlik değerleri en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. ÜDMTK için sonuçlar incelendiğinde test çok zor olduğu durumda elde edilen en yüksek güvenirlik değeri çoklu puanlanan madde yüzdesi %50 düzeyinde elde edilirken en düşük güvenirlik çoklu puanlanan madde yüzdesi %5 ve %25 düzeylerinde elde edilmiştir. Test güçlüğü zor olduğunda güvenirlik değerleri en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdesine göre sırasıyla %5, %10, % 50 ve %25 olduğunda elde edilmiştir. Test güçlüğü orta ve çok kolay olduğunda da sıralama aynıdır fakat çoklu puanlanan madde yüzdesi %5 ve %10 için hesaplanan güvenirlik değerleri arasındaki fark anlamlı değildir. Test güçlüğü kolay olduğunda güvenirlik değerleri arasındaki anlamlı fark sadece çoklu puanlanan madde yüzdesinin %5 ve %50 olması durumunda elde edilirken en yüksek güvenirlik değeri %5, en düşük güvenirlik değeri ise %50 düzeylerinde elde edilmiştir. Bifaktör model ÜDMTK ile elde edilen bulgularla benzerlik göstermektedir. Test çok zor olduğunda güvenirlik değerleri en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdesi düzeylerine göre sırasıyla %50, %10, %25 ve %5 olduğunda elde edilmiştir. Test güçlüğü zor, orta ve çok kolay olduğunda ise bu sıralama %5, %10, %50 ve %25 şeklindedir. Fakat test güçlüğü çok kolay olduğunda çoklu puanlanan madde yüzdesi %10 ve %50 arasındaki fark anlamlı değildir.

Sınıflama doğruluğu değerleri ile yapılan analizlerde ÇBMTK için test güçlüğü çok zor, zor, orta ve çok kolay olduğunda hesaplanan değerler en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdelerine göre sırasıyla %50, %25, %10 ve %5 olduğunda elde edilmiştir. Orta test güçlüğünde %50 ve %25 çoklu puanlanan madde yüzdelerinde hesaplanan sınıflama doğrululukları arasında anlamlı fark yoktur. Test güçlüğü kolay olduğunda ise diğerlerinden farklı olarak en yüksek sınıflama doğruluğu çoklu puanlanan madde yüzdesi %25 olduğunda elde edilmiştir. ÜDMTK model ile yapılan analizlerde de ÇBMTK ile benzer şekilde bütün test güçlük düzeylerinde, çoklu puanlanan madde yüzdesi arttıkça sınıflama doğruluğu artmıştır. Sadece test güçlüğü kolay olduğunda %5 ve %10 çoklu puanlanan madde yüzdesi düzeylerinde elde edilen değerler arasındaki fark anlamlı değildir. Bifaktör model ile yapılan analizler için de genel olarak benzer bir durum söz konusudur. Farklı olarak, test güçlüğü zor olduğunda %25 ve %50, test güçlüğü kolay olduğunda %5 ve %10 çoklu puanlanan madde yüzdelerinde hesaplanan değerler arasındaki fark anlamlı değildir.

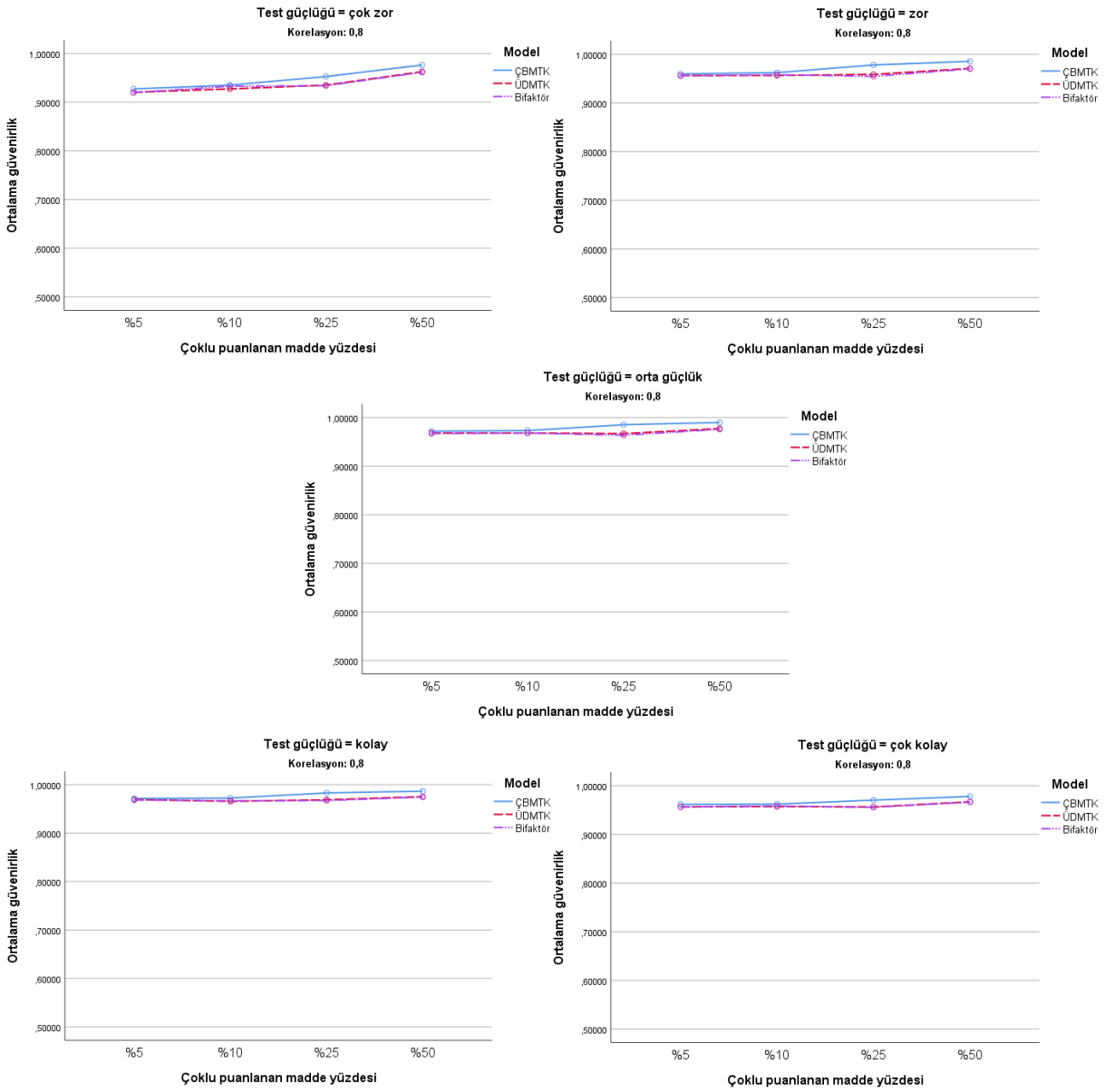
Özetle, bütün test güçlük düzeylerinde çoklu puanlanan madde yüzdesi arttıkça ÇBMTK modeli ile elde edilen RMSE değerleri azalırken güvenirlik ve sınıflama doğruluğu değerleri artış göstermiştir. ÜDMTK ve Bifaktör modellerinin her biri için hesaplanan RMSE ve güvenirlik değerleri çoklu puanlanan madde yüzdesinin düzeylerinde genel olarak benzerlik göstermiştir. Ancak test kolay ve çok kolay olduğunda çoklu puanlanan madde yüzdesindeki artış RMSE değerlerini arttırırken güvenirlik değerlerini düşürmüştür. Sınıflama doğruluğu açısından ise ÜDMTK ve Bifaktör modeller ile elde edilen ortalama değerler çoklu puanlanan madde yüzdesi düzeylerine göre birbirlerine yakın olup benzerlik göstermektedir.

Korelasyon 0,8 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

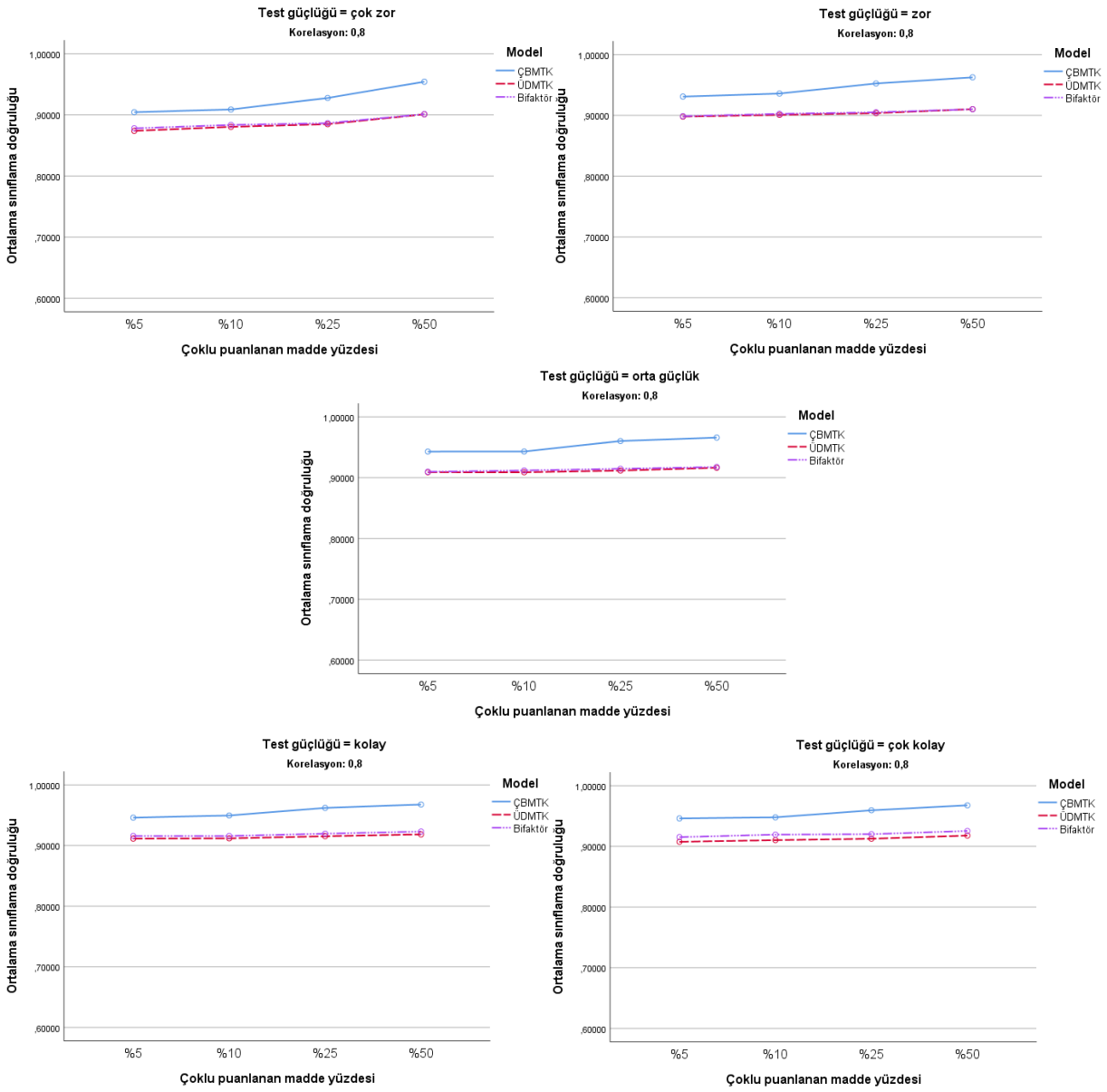
Boyutlar arası korelasyon 0,8 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki etkileşim etkisi incelenmiştir. Bu amaçla model ve test güçlüğü değişkenleri ile RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin karşılaştırılması için iki yönlü karma ANOVA yapılmış ve sonuçlar Tablo 20'de sunulmuştur. Test güçlüğünün her bir düzeyi için model ve çoklu puanlanan madde yüzdesi düzeylerinde elde edilen ortalama RMSE, güvenirlik ve sınıflama doğruluğu değerlerini içeren grafikler sırasıyla Şekil 18, Şekil 19 ve Şekil 20'de yer almaktadır.



Şekil 18. Toplam puan korelasyon 0,8 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE



Şekil 19. Toplam puan korelasyon 0,8 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik)



Şekil 20. Toplam puan korelasyon 0,8 — test güçlüğü x model x çoklu puanlanan madde yüzdesi (sınıflama doğruluğu)

Tablo 20.

Toplam Puan için Korelasyon 0,8 iken Test Güçlüğünün Her Bir Düzeyi için Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Karma ANOVA

			RMSE		Güvenirlik		Sınıflama doğruluğu	
TG	Kaynak		F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	Gruplar içi	M	79036,613*	,995	44911,831*	,991	307744,093*	,999
		M *x* ÇPMY	3903,002*	,967	4104,524*	,969	8406,932*	,985
	Gruplar arası	TG	10393,197*	,987	16276,468*	,992	13757,659*	,990
Zor	Gruplar içi	M	119447,197*	,997	69414,936*	,994	515684,155*	,999
		M *x* ÇPMY	4563,111*	,972	10331,932*	,987	7283,353*	,982
	Gruplar arası	TG	6885,250*	,981	9410,054*	,986	7409,482*	,982
Orta güçlük	Gruplar içi	M	195954,737*	,998	69222,558*	,994	544608,018*	,999
		M *x* ÇPMY	5364,060*	,976	8349,715*	,984	6266,381*	,979
	Gruplar arası	TG	8348,354*	,984	6521,573*	,980	5262,685*	,976
Kolay	Gruplar içi	M	125872,890*	,997	69407,694*	,994	436020,558*	,999
		M *x* ÇPMY	3361,963*	,962	7013,592*	,982	3278,162*	,961
	Gruplar arası	TG	1848,141*	,933	4967,343*	,974	4695,285*	,973
Çok kolay	Gruplar içi	M	99996,158*	,996	60979,445*	,994	309176,136*	,999
		M *x* ÇPMY	2965,753*	,957	4589,012*	,972	1985,567*	,938
	Gruplar arası	TG	176,853*	,573	4729,547*	,973	5514,131*	,977

*p<,001; M: model; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 20'de yer alan boyutlar arası korelasyon 0,8 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta güçlük, kolay, çok kolay) gerçekleştirilen analiz sonuçları incelendiğinde model ve çoklu puanlanan madde yüzdesi arasındaki etkileşimin RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için anlamlı etki büyüklüğünün yüksek olduğu görülmektedir. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi için kestirim modellerinin her biri için çoklu puanlanan madde yüzdesi basit ana etkisi tek yönlü ANOVA ile incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde çoklu puanlanan madde yüzdeleri (%5, %10, %25 ve %50) için hesaplanan RMSE, güvenirlik ve sınıflama doğruluğu değerleri ayrı ayrı karşılaştırılmıştır.

Çoklu puanlanan madde yüzdesine ait basit ana etkiler için elde edilen sonuçlar Ek 9'da verilmiştir. ANOVA çıktıları, bütün test güçlük düzeyleri altında her bir model için çoklu puanlanan madde yüzdelerine göre elde edilen RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin birbirinden anlamlı olarak farklılaştığını göstermektedir. Farkın kaynağına ilişkin yapılan ikili karşılaştırmalar ile ortalama RMSE, güvenirlik ve sınıflama doğruluğu değerleri Ek 9'da sunulmaktadır. Buna göre bütün test güçlüklerinde ÇBMTK için elde edilen RMSE değerleri çoklu puanlanan madde yüzdesi değişkenine göre en yüksekten en düşüğe doğru sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. ÜDMTK model için test çok zor, zor ve orta güçlükte olduğunda çoklu puanlanan maddeler arasındaki karşılaştırma sonucu ÇBMTK için elde edilen ile aynı olup en yüksekten en düşüğe doğru RMSE sıralaması %5,

%10, %25 ve %50 şeklindedir. ÜDMTK için test kolay olduğunda en düşük RMSE çoklu puanlanan madde yüzdesi %50 iken elde edilirken test çok kolay olduğunda en düşük RMSE %5 çoklu puanlanan madde yüzdesi ile elde edilmiştir. Bifaktör model için test güçlüğü çok zor, zor ve orta güçlükte olduğunda en düşük RMSE çoklu puanlanan madde yüzdesi %50 olduğunda ve en yüksek RMSE %5 olduğunda elde edilmiştir. Test güçlüğü kolay ve çok kolay olduğunda Bifaktör model ile en yüksek RMSE çoklu puanlanan madde yüzdesi düzeyi %50 olduğunda elde edilmiştir. ÜDMTK ve Bifaktör modeller için ortalama RMSE değerleri birbirine yakın olup bazı durumlarda anlamlı fark bulunmamaktadır.

Güvenirlik için sonuçlar incelendiğinde bütün test güçlük düzeylerinde ÇBMTK için elde edilen güvenirlik değerleri en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. ÜDMTK ve Bifaktör model için de en yüksek güvenirlik çoklu puanlanan madde yüzdesi %50 olduğunda elde edilmiştir. Diğer çoklu puanlanan madde yüzdeleri için hesaplanan güvenirliklerin birbirine yakın olduğu ve bazı durumlarda aralarında anlamlı fark olmadığı görülmüştür.

Sınıflama doğruluğu değerleri açısından ise tüm modellerde çoğu durumda çoklu puanlanan madde yüzdelerine göre değerlerin en yüksekten en düşüğe doğru %50, %25, %10 ve %5 şeklinde sıralandığı görülmektedir. Test güçlüğü orta olduğunda ÇBMTK ve ÜDMTK için yapılan analizlerde %5 ve %10 çoklu puanlanan madde yüzdelerinde hesaplanan sınıflama doğruluğu değerleri arasındaki fark anlamlı değildir. Benzer şekilde test güçlüğü kolay olduğunda da ÜDMTK ve Bifaktör için %5 ve %10 çoklu puanlanan madde yüzdelerinde hesaplanan sınıflama doğruluğu değerleri arasındaki fark anlamlı değildir.

Özetle, bütün test güçlük düzeylerinde çoklu puanlanan madde yüzdesi arttıkça ÇBMTK modeli ile elde edilen RMSE değerleri azalırken güvenirlik ve sınıflama doğruluğu değerleri artış göstermiştir. ÜDMTK ve Bifaktör modelleri için ilk üç çoklu puanlanan madde yüzdesi düzeylerinde birbirlerine yakın RMSE ve güvenirlik değerleri elde edilirken en yüksek güvenirlik çoklu puanlanan madde yüzdesi %50 olduğunda elde edilmiştir. Sınıflama doğruluğu açısından ise ÜDMTK ve Bifaktör modeller ile elde edilen ortalama değerler çoklu puanlanan madde yüzdesi düzeylerine göre birbirlerine yakındır.

4.1.3. Boyutlar Arası Korelasyon Değişkeni için İncelemeler

Bu bölümde gerçekleştirilen analizlerde simülasyon veride toplam puan açısından boyutlar arası korelasyonun ele alınan değerlendirme kriterleri açısından yetenek kestirimine olan etkisi ortaya konmuştur. Bu amaçla toplam test için elde edilen RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin karşılaştırılması için gerçekleştirilen faktöriyel karma ANOVA (Tablo 5) sonucunda model, çoklu puanlanan madde yüzdesi, test güçlüğü ve korelasyon değişkenleri arasında anlamlı çıkan dört yönlü etkileşimin yorumlanmasında bir inceleme daha yapılmıştır. Burada amaç korelasyon basit ana etkisinin incelenmesidir. Bu kez çoklu puanlanan madde yüzdesinin her bir düzeyinde model, test güçlüğü ve korelasyon arasındaki üç yönlü etkileşim etkisi incelenmiştir. Bu amaçla gerçekleştirilen faktöriyel karma ANOVA sonuçları Tablo 21'de sunulmuştur.

Tablo 21.

Toplam Puan için Çoklu Puanlanan Madde Yüzdesinin Her Bir Düzeyinde Model, Test Güçlüğü ve Korelasyon Değişkenleri ile Faktöriyel Karma ANOVA

		RMSE		Güvenirlik		Sınıflama doğruluğu	
ÇPMY	Kaynak	F	η_p^2	F	η_p^2	F	η_p^2
	Gruplar içi						
	M	436183,865*	,997	73190,951*	,980	435578,389*	,997
%5	M *x* K	83550,628*	,991	45273,998*	,984	73086,174*	,990
	M *x* TG	3773,692*	,910	4168,527*	,918	3944,031*	,914
	M *x* K *x* TG	3150,970*	,944	3453,801*	,949	2438,803*	,929
	Gruplar içi						
	M	723423,940*	,998	131653,321*	,989	671578,751*	,998
%10	M *x* K	140488,562*	,995	75709,708*	,990	113406,429*	,993
	M *x* TG	8330,910*	,957	6152,675*	,943	6301,232*	,944
	M *x* K *x* TG	5762,724*	,969	5066,849*	,965	4406,189*	,960
	Gruplar içi						
	M	656055,007*	,998	161456,680*	,991	992934,944*	,999
%25	M *x* K	104590,412*	,993	74003,249*	,990	140737,977*	,995
	M *x* TG	4993,853*	,931	4687,353*	,927	1817,310*	,830
	M *x* K *x* TG	2658,643*	,935	3086,423*	,943	2134,741*	,920
	Gruplar içi						
	M	1189254,799*	,999	125758,110*	,988	2449626,047*	,999
%50	M *x* K	119588,354*	,994	43798,038*	,983	325788,257*	,998
	M *x* TG	1782,689*	,828	1405,722*	,791	8971,506*	,960
	M *x* K *x* TG	1192,260*	,865	1063,843*	,851	2012,211*	,916

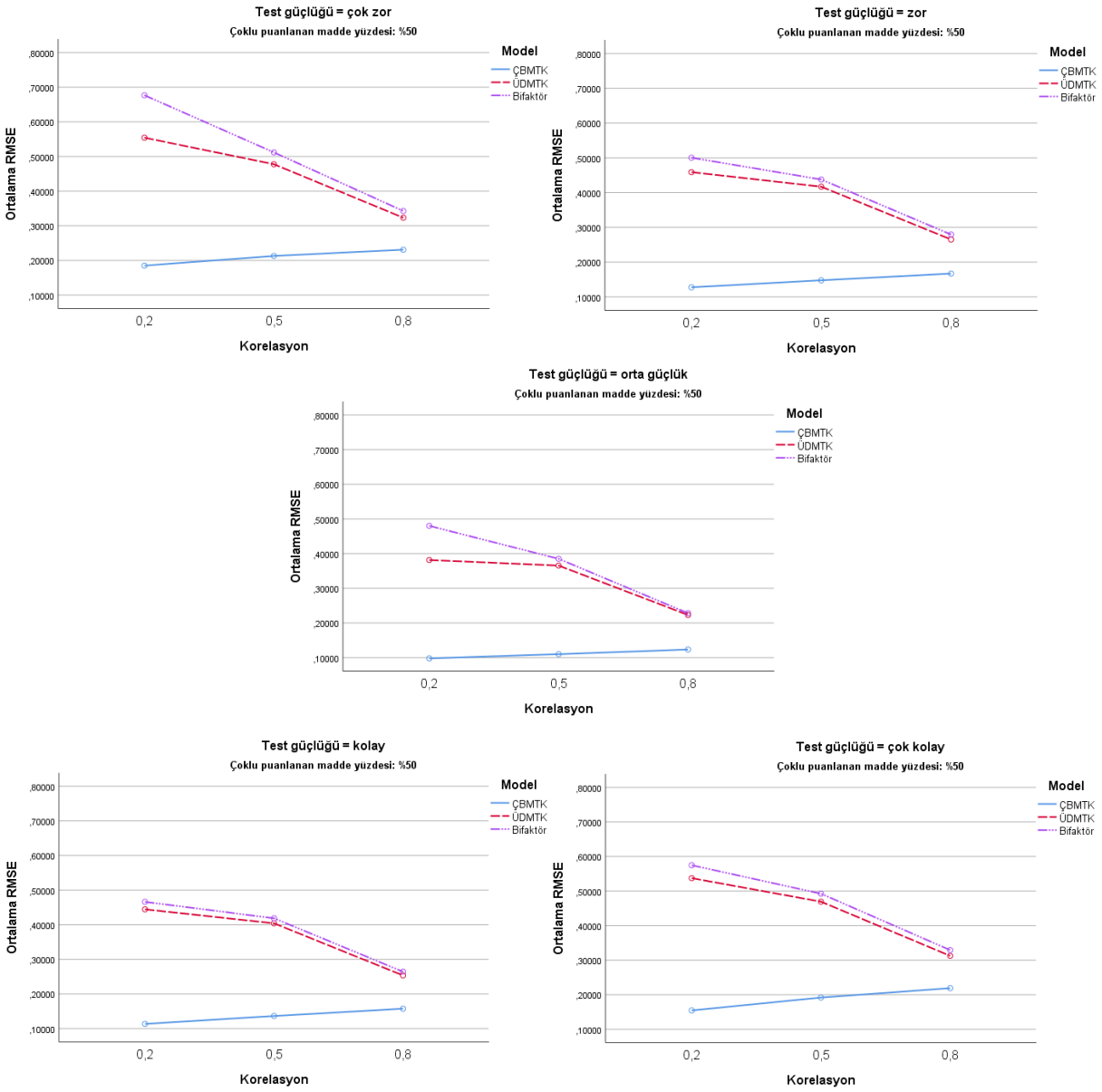
*p<,001; M: model; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi; K: korelasyon

Tablo 21 incelendiğinde çoklu puanlanan madde yüzdesinin bütün düzeylerinde RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin her biri için gerçekleştirilen ANOVA sonucunda model, korelasyon ve test güçlüğü arasındaki üç yönlü etkileşimlerin her üç değerlendirme kriteri için de anlamlı olduğu görülmektedir. Bu üç yönlü etkileşimi incelemek

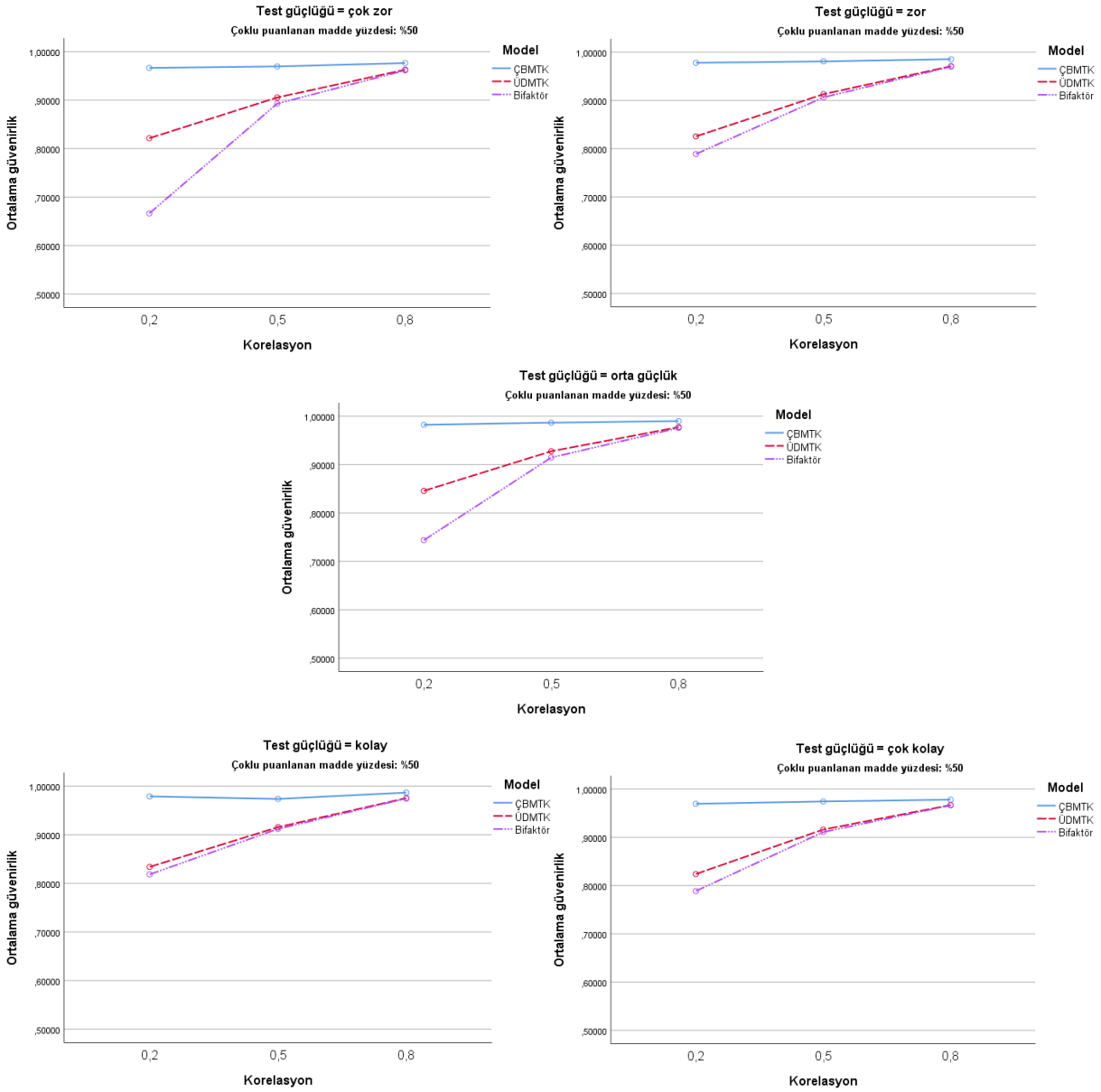
amacıyla basit iki yönlü etkileşim yaklaşımı kullanılmıştır. Bunun için çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) test güçlüğünün bütün düzeyleri (çok zor, zor, orta, kolay ve çok kolay) için model ve korelasyon arasındaki etkileşim etkisi incelenmiştir. Bu amaçla yapılan karma ANOVA sonuçları Ek 10'da yer almaktadır.

Buna göre çoklu puanlanan madde yüzdesinin ve test güçlüğünün her bir düzeyinde model ve korelasyon arasındaki iki yönlü etkileşim anlamlı bulunmuştur. Burada elde edilen iki yönlü etkileşim etkisinin değerlendirilmesi için korelasyon basit ana etkisi incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde boyutlar arası korelasyon düzeyleri (0,2, 0,5 ve 0,8) için hesaplanan RMSE, güvenirlik ve sınıflama doğruluğu değerleri karşılaştırılmıştır. Korelasyon basit ana etkisi için elde edilen sonuçlar Ek 10'da verilmiştir.

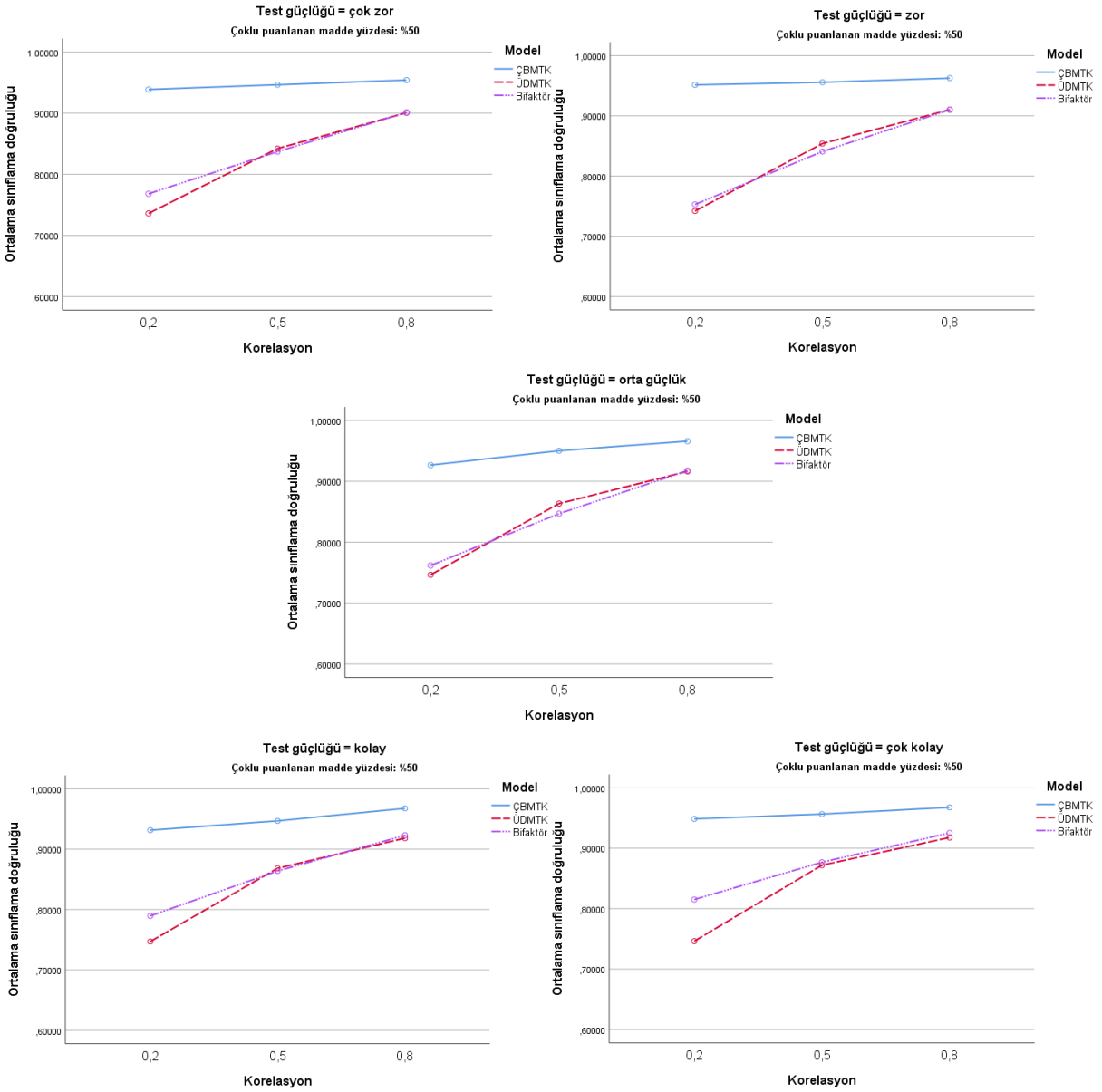
Çoklu puanlanan madde yüzdesi %50 için her bir test güçlüğü düzeyinde model ve korelasyon değişkenlerine ait ortalama RMSE, güvenirlik ve sınıflama doğruluğu değerlerini gösteren grafikler sırasıyla Şekil 21, Şekil 22 ve Şekil 23'te yer almaktadır. Çoklu puanlanan madde yüzdesinin diğer düzeyleri için grafiksel gösterimler Ek 10'da sunulmaktadır.



Şekil 21. Toplam puan çoklu puanlanan madde yüzdesi %50 –test güçlüğü x model x korelasyon (RMSE)



Şekil 22. Toplam puan çoklu puanlanan madde yüzdesi %50 – test güçlüğü x model x korelasyon (güvenirlik)



Şekil 23. Toplam puan çoklu puanlanan madde yüzdesi %50 – test güçlüğü x model x korelasyon (sınıflama doğruluğu)

Çoklu puanlanan madde yüzdesi ve test güçlüğü değişkenlerinin düzeylerinde anlamlı bulunan model ve korelasyon arasındaki iki yönlü etkileşimin incelendiği korelasyon basit ana etkisine ait analiz sonuçları bütün modeller için korelasyon düzeyleri arasındaki farkın her üç değerlendirme kriteri için de anlamlı olduğunu göstermektedir. Anlamlı bulunan bu farkların kaynağına ilişkin yapılan ikili karşılaştırma sonuçları Ek 10'da yer almaktadır. Buna göre genel olarak sonuçlar paralellik göstermektedir ve neredeyse bütün karşılaştırmalar anlamlıdır. ÇBMTK modeli ile yapılan analizlerde bütün çoklu puanlanan madde yüzdeleri için korelasyon arttıkça RMSE değerleri artış göstermiştir. Ortalama RMSE değerlerindeki farklılık ÇBMTK

modelinde diğer modellere göre daha azdır. ÜDMTK ve Bifaktör modelleri için ise korelasyon değeri arttıkça RMSE değeri azalmıştır. Sadece çoklu puanlanan madde yüzdesi %5 ve test güçlüğü orta düzey iken ÜDMTK için korelasyon 0,2 ve 0,5 olduğunda elde edilen RMSE değerleri arasında anlamlı farklılık yoktur.

Güvenirlik için yapılan analiz sonuçları incelendiğinde ise genel olarak kestirim modellerinin her üçü için de yapılan analizler korelasyon arttıkça güvenirlik değerlerinin arttığını göstermektedir. Sadece çoklu puanlanan madde yüzdesi %50 olduğunda test güçlüğü kolay iken ÇBMTK model ile elde edilen güvenirlikler korelasyon düzeyine göre farklılık göstermemiştir. Ortalama güvenirlik değerlerindeki farklılık ÇBMTK modelinde diğer modellere göre daha azdır.

Sınıflama doğruluğu açısından ise güvenirlik değerleri ile benzer şekilde bütün çoklu puanlanan madde yüzdelerinde ve bütün test güçlük düzeylerinde korelasyon arttıkça sınıflama doğruluğu değerleri artış göstermektedir. Sadece çoklu puanlanan madde yüzdesi %5 olduğunda Bifaktör model ile korelasyon 0,2 ve 0,5 iken elde edilen sınıflama doğruluğu değerleri arasındaki fark anlamlı değildir.

4.2. Simülasyon Çalışması Alt Puanlar için Bulgular

Bu bölümde simülasyon veri ile gerçekleştirilen analizlerde alt puan açısından incelemelere yer verilmiştir. ÇBMTK, ÜDMTK, Bifaktör kestirim modellerinin ve simülasyon koşullarının (test güçlüğü, boyutlar arası korelasyon ve çoklu puanlanan madde yüzdesi) yetenek kestirimlerine ait RMSE, güvenirlik ve sınıflama doğruluğu değerlerine olan etkisi faktöriyel karma ANOVA sonuçlarına göre alt başlıklar halinde incelenmiştir.

ÇBMTK, ÜDMTK ve Bifaktör modelleri ile yapılan analiz sonrasında elde edilen alt test puanları için RMSE, güvenirlik ve sınıflama doğruluğu değerleri hesaplanmıştır. Bu değerlerin analiz modellerine ve simülasyon koşullarına göre karşılaştırılması amacıyla her bir bağımlı değişken (RMSE, güvenirlik ve sınıflama doğruluğu) için ayrı ayrı olmak üzere faktöriyel karma ANOVA yapılmıştır. Sonuçlar Tablo 22'de sunulmuştur.

Tablo 22.

Alt Puanlar için RMSE, Güvenirlik ve Sınıflama Doğruluğu Değerlerine Ait Faktöriyel Karma ANOVA Sonuçları

	RMSE		Güvenirlik		Sınıflama doğruluğu	
Kaynak	F	η_p^2	F	η_p^2	F	η_p^2
Gruplar içi						
M	1738402,316*	,986	2834606,142*	,992	557229,479*	,959
M *x* AT	2153,173*	,213	7048,176*	,469	2004,185*	,201
M *x* ÇPMY	1965,636*	,198	47,737*	,006	93,212*	,012
M *x* TG	2649,089*	,307	40,585*	,007	363,901*	,057
M *x* K	290111,066*	,960	738622,457*	,984	51201,524*	,811
M *x* TG *x* ÇPMY	47,750*	,023	66,149*	,032	4,137*	,002
M *x* ÇPMY *x* K	176,968*	,042	31,072*	,008	29,379*	,007
M *x* K *x* TG	187,642*	,059	23,545*	,008	17,397*	,006
M *x* TG *x* ÇPMY *x* K	13,755*	,014	22,510*	,022	3,944*	,004
Gruplar arası						
AT	14326,977*	,642	11503,622*	,590	13762,010*	,633
ÇPMY	3022,887*	,275	1692,542*	,175	3324,894*	,294
TG	5061,846*	,458	4477,189*	,428	13751,030*	,697
K	14325,613*	,545	60767,098*	,835	3112,229*	,206
ÇPMY *x* TG	42,187*	,021	58,964*	,029	84,081*	,040
ÇPMY *x* K	59,992*	,015	24,538*	,006	1,511*	,000
TG *x* K	120,417*	,039	110,901*	,036	120,121*	,039
ÇPMY *x* TG *x* K	1,782*	,002	4,077*	,004	1,577*	,002

*p<,001 M: model; AT: alt test; ÇPMY: çoklu puanlanan madde yüzdesi; TG: test güçlüğü; K: korelasyon

Tablo 22'de yer alan RMSE ve güvenirlik değerleri ile gerçekleştirilen analizler sonucunda model, çoklu puanlanan madde yüzdesi, test güçlüğü ve korelasyon değişkenleri arasındaki dört yönlü etkileşimin anlamlı ve düşük etki büyüklüğüne sahip olduğu görülmektedir. Bu dört yönlü etkileşimin yorumlanması için öncelikle boyutlar arası korelasyon değişkeninin her bir düzeyinde (0,2, 0,5 ve 0,8) çoklu puanlanan madde yüzdesi, model ve test güçlüğü değişkenleri arasındaki üç yönlü etkileşim etkisi incelenmiştir. Ardından çoklu puanlanan madde yüzdesi değişkeninin her bir düzeyinde model, korelasyon ve test güçlüğü değişkenleri arasındaki üç yönlü etkileşim etkisi incelenmiştir. Üç yönlü etkileşim etkisinin incelenmesine yönelik analizler aşağıda ayrı ayrı sunulmaktadır. Sınıflama doğrulu değerleri ile yapılan analiz sonucunda ise üç ve dört yönlü etkileşimlere dair istatistik değeri anlamlı olsa da kısmi eta kare değerlerinin 0,01'in altında olduğu, bu etkileşime dair etkinin oldukça düşük olduğu görülmüştür. Bu sebeple sınıflama doğruluğu değerleri için iki yönlü etkileşim etkileri incelenmiştir. Anlamlı bulunan iki yönlü etkileşim etkileri için değişkenler sırayla sabit tutularak diğer değişkenin düzeyleri arasındaki farklar incelenmiştir. Sınıflama doğruluğuna ait iki yönlü etkileşimlerin incelenmesine en sonda yer verilmiştir.

4.2.1. Kestirim Modelleri ve Test Güçlüğü Değişkeni için İncelemeler

Bu bölümde gerçekleştirilen analizlerde simülasyon veride toplam puan açısından yetenek kestirim modellerinin ve test güçlüğünün ele alınan değerlendirme kriterleri açısından yetenek kestirimine olan etkisi ortaya konmuştur.

Korelasyon 0,2 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

Korelasyon 0,2 düzeyinde model, çoklu puanlanan madde yüzdesi ve test güçlüğü değişkenleri arasındaki üç yönlü etkileşim etkisinin incelenmesi amacıyla bu değişkenler ile RMSE ve güvenirlik değerlerinin her biri için ayrı ayrı faktöriyel karma ANOVA yapılmıştır. Analiz sonuçları Tablo 23'te yer almaktadır.

Tablo 23.

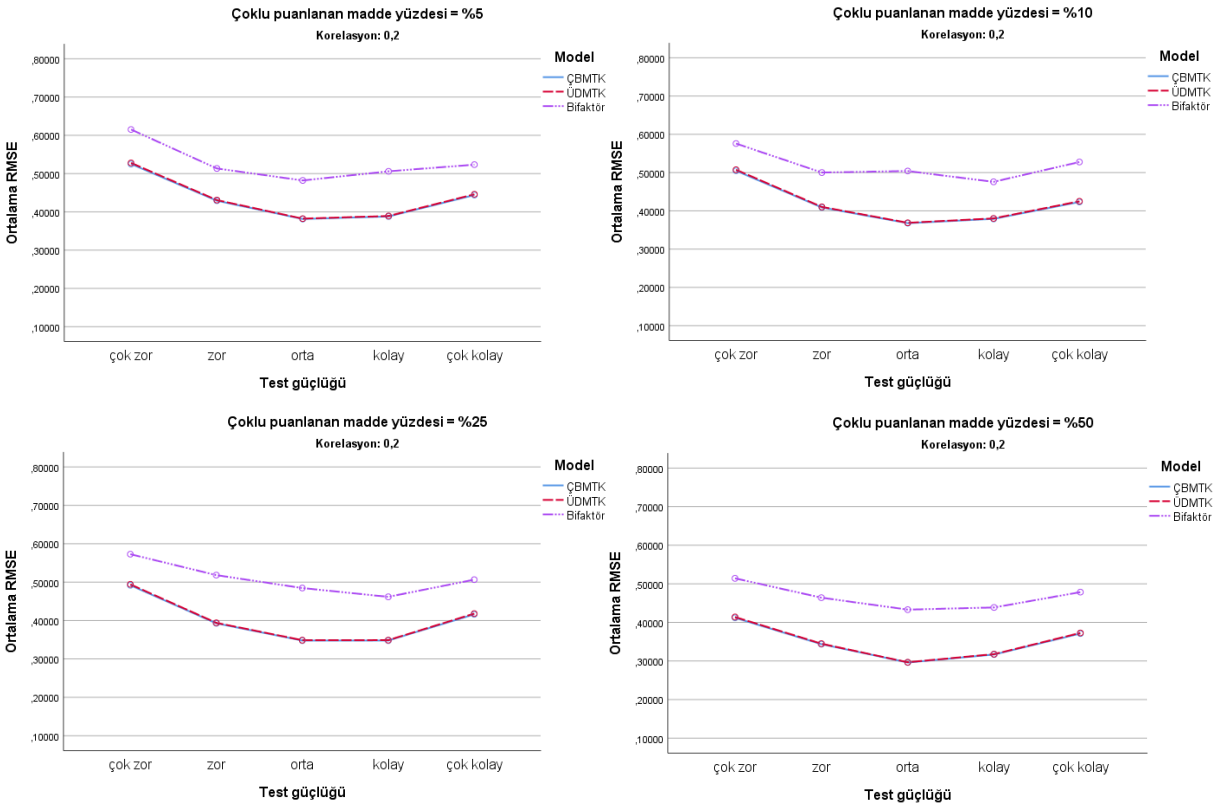
Alt Puanlar için Korelasyon 0,2 Düzeyinde Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Faktöriyel Karma ANOVA (RMSE, Güvenirlik)

	RMSE		Güvenirlik	
Kaynak	F	η_p^2	F	η_p^2
Gruplar içi				
M	38653,784*	,829	25732,585*	,763
M *x* AT	1710,604*	,391	2395,366*	,474
M *x* ÇPMY	95,600*	,035	16,297*	,006
M *x* TG	187,742*	,086	33,666*	,017
M *x* TG *x* ÇPMY	30,281*	,044	53,637*	,075
Gruplar arası				
AT	7281,227*	,733	5569,088*	,677
ÇPMY	1252,367*	,320	911,427*	,255
TG	2124,465*	,516	2438,550*	,550
ÇPMY *x* TG	18,542*	,027	35,054*	,050

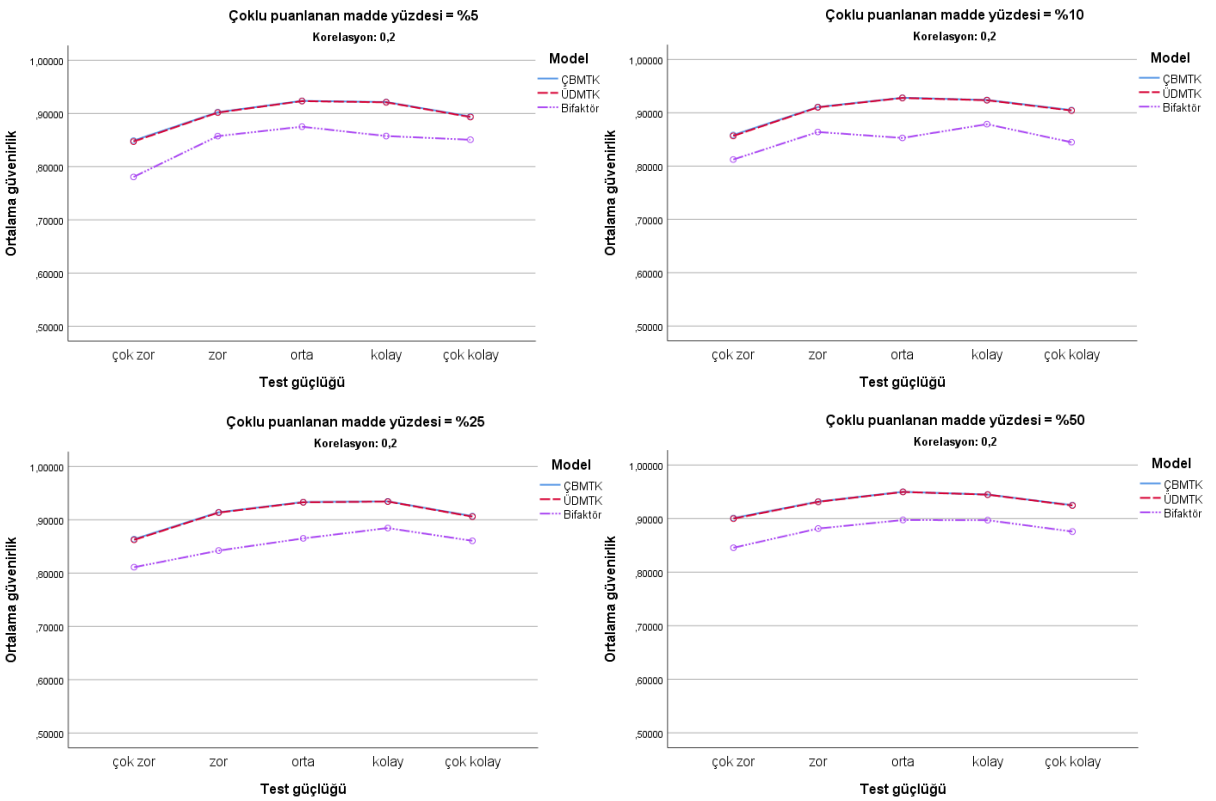
*p<,001 M: model; AT: alt test; ÇPMY: çoklu puanlanan madde yüzdesi; TG: test güçlüğü

Boyutlar arası korelasyon 0,2 düzeyinde RMSE ve güvenirlik değerlerinin her biri için gerçekleştirilen ANOVA sonucunda model, test güçlüğü ve çoklu puanlanan madde yüzdesi değişkenleri arasındaki üç yönlü etkileşimin anlamlı olduğu tespit edilmiştir. Etkileşimlere ait etki büyüklükleri RMSE için düşük ve güvenirlik için orta düzeydedir. Bu üç yönlü etkileşimleri incelemek amacıyla basit iki yönlü etkileşim yaklaşımı kullanılmıştır. Bunun için çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model-test güçlüğü ve model-alt test arasındaki etkileşim etkileri incelenmiştir. Daha sonra test güçlüğünün her bir düzeyinde (çok zor, zor, orta, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki etkileşim etkisi incelenmiştir.

Öncelikle çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model, test güçlüğü ve alt test (blok) değişkenleri ile RMSE ve güvenirlik değerleri için karma ANOVA yapılmış ve sonuçlar Tablo 24'te sunulmuştur. Çoklu puanlanan madde yüzdesinin her bir düzeyi için model ve test güçlüğü düzeylerinde elde edilen ortalama RMSE ve güvenirlik değerlerini içeren grafikler sırasıyla Şekil 24 ve Şekil 25'te yer almaktadır.



Şekil 24. Alt puanlar korelasyon 0,2 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE)



Şekil 25. Alt puanlar korelasyon 0,2 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik)

Tablo 24.

Alt Puanlar için Korelasyon 0,2 Düzeyinde Her Bir Çoklu Puanlanan Madde Yüzdesi için Model, Test Güçlüğü ve Alt Test Değişkenleri ile ANOVA Sonuçları (RMSE, Güvenirlik)

			RMSE		Güvenirlik	
ÇPMY	Kaynak		F	η_p^2	F	η_p^2
%5	Gruplar içi	M	9797,227*	,831	5794,259*	,744
		M *x* AT	805,632*	,548	686,213*	,508
		M *x* TG	54,188*	,098	51,473*	,094
	Gruplar arası	AT	1909,284*	,742	1018,498*	,605
		TG	1672,908*	,771	1267,463*	,718
%10	Gruplar içi	M	7533,267*	,791	5306,652*	,727
		M *x* AT	619,137*	,483	665,978*	,501
		M *x* TG	88,546*	,151	62,278*	,111
	Gruplar arası	AT	5303,915*	,889	2484,693*	,789
		TG	1775,180*	,781	1168,059*	,701
%25	Gruplar içi	M	12540,084*	,863	7765,299*	,796
		M *x* AT	1158,362*	,636	1254,042*	,654
		M *x* TG	119,985*	,194	62,672*	,112
	Gruplar arası	AT	13405,800*	,953	5912,737*	,899
		TG	2689,878*	,844	1769,209*	,780
%50	Gruplar içi	M	106935,127*	,982	38383,718*	,951
		M *x* AT	1256,708*	,654	404,344*	,378
		M *x* TG	301,397*	,377	23,530*	,045
	Gruplar arası	AT	10846,732*	,942	4458,214*	,870
		TG	1102,198*	,689	626,212*	,557

*p<,001; M: model; AT: alt test; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 24'te yer alan boyutlar arası korelasyon 0,2 düzeyi ve çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) gerçekleştirilen analiz sonuçları incelendiğinde model-test güçlüğü ve model-alt test arasındaki etkileşimin RMSE ve güvenirlik değerleri için anlamlı olduğu görülmektedir. Model-alt test etkileşim etkisi yüksek etki büyüklüklerine sahiptir. Model-test güçlüğü etkileşim etkisi ise genel olarak düşük, orta ya da ortaya yakın etki büyüklüklerine sahiptir. Öncelikle anlamlı bulunan iki yönlü etkileşimler arasında model-test güçlüğü etkileşimi ele alınmıştır. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi için sırasıyla model ve test güçlüğü değişkenlerinin her bir düzeyinde diğer değişkene ait basit ana etkiler tek yönlü ANOVA ile incelenmiştir. İlk olarak test güçlüğü basit ana etkisi incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde test güçlük düzeyleri (çok zor, zor, orta güçlük, kolay, çok kolay) için hesaplanan RMSE ve güvenirlik değerleri karşılaştırılmıştır. Test güçlüğü basit ana etkisi için elde edilen sonuçlar Tablo 25'te verilmiştir.

Tablo 25.

Alt Puanlar için Korelasyon 0,2 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması (RMSE, Güvenirlik)

		RMSE		Güvenirlik	
ÇPMY	Model	F	η_p^2	F	η_p^2
%5	ÇBMTK	3529,470*	,876	2765,980*	,847
	ÜDMTK	3523,304*	,876	2755,634*	,847
	Bifaktör	397,368*	,444	392,333*	,441
%10	ÇBMTK	3237,767*	,867	2284,932*	,821
	ÜDMTK	3237,767*	,867	2281,318*	,821
	Bifaktör	222,018*	,308	209,231*	,296
%25	ÇBMTK	4781,170*	,906	2915,615*	,854
	ÜDMTK	4787,740*	,906	2910,103*	,854
	Bifaktör	362,497*	,421	353,489*	,415
%50	ÇBMTK	1622,895*	,765	856,310*	,632
	ÜDMTK	1635,513*	,767	852,329*	,631
	Bifaktör	426,979*	,462	357,032*	,418

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 25'te çoklu puanlanan madde yüzdesi ve kestirim modellerinin tüm düzeylerinde RMSE ve güvenirlik değerlerinin her birinin test güçlüklerine göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. Anlamlı bulunan bu farkların kaynağının belirlenmesi için ikili karşılaştırmalar yapılmıştır (Ek 11). Buna göre çoklu puanlanan madde yüzdesi %5, %10 ve %50 olduğunda ÇBMTK ve ÜDMTK modelleri ile test güçlük düzeylerine göre elde edilen bütün RMSE değerleri anlamlı olarak birbirinden farklılaşmaktadır. RMSE değerleri test güçlük düzeylerinde en düşükten en yükseğe doğru orta, kolay, zor, çok kolay ve çok zor şeklinde sıralanmaktadır. ÇBMTK ve ÜDMTK modelleri ile

çoklu puanlanan madde yüzdesi %25 olduğunda ise sıralama aynı olup tek fark orta güçlük ve kolay test güçlüğü düzeylerinde elde edilen RMSE değerleri arasındaki farkın anlamlı olmamasıdır.

Bifaktör model için en düşük RMSE çoklu puanlanan madde yüzdesi %5 ve %10 olduğunda orta güçlük düzeyi iken elde edilirken çoklu puanlanan madde yüzdesi %25 olduğunda test güçlük düzeyi kolay iken elde edilmiştir. Çoklu puanlanan madde yüzdesi %50 olduğunda orta ve kolay test güçlük düzeyleri arasında anlamlı fark yoktur ve en düşük hataya sahiptir. Çoklu puanlanan madde yüzdesi %5 iken test güçlüğü zor ve kolay arasında, çoklu puanlanan madde yüzdesi %10 iken zor ve orta güçlük arasında anlamlı fark yoktur. Çoklu puanlanan madde yüzdesi %25 düzeyinde ise zor ve çok kolay test güçlük düzeyleri arasında anlamlı fark yoktur. ÇBMTK ve ÜDMTK modellerinde olduğu gibi Bifaktör model için de bütün çoklu puanlanan madde yüzdelerinde en yüksek RMSE test güçlüğü çok zor olduğunda elde edilmiştir.

Güvenirlik değerleri açısından ise ÇBMTK ve ÜDMTK modelleri altında bütün çoklu puanlanan madde yüzdelerinde güvenirlik değerleri en yüksekten en düşüğe doğru sırasıyla test güçlüğü orta, kolay, zor, çok kolay ve çok zor olduğunda elde edilmiştir. Çoklu puanlanan madde yüzdesi %10 ve %50 olduğunda güvenirlik değerleri arasındaki bütün farklar anlamlı iken %5 ve %25 olduğunda orta ve kolay test güçlükleri için elde edilen güvenirlik değerleri arasındaki fark anlamlı değildir. Bifaktör model için çoklu puanlanan madde yüzdesi %5 olduğunda en yüksek güvenirlik orta güçlük düzeyi ile elde edilirken test güçlüğü zor, kolay ve çok kolay ile elde edilen güvenirlikler arasındaki fark anlamlı değildir. Çoklu puanlanan madde yüzdesi %10 ve %25 için en yüksek güvenirlik kolay test güçlüğü ile elde edilmiştir. Ancak orta ve çok kolay test güçlük düzeyleri ile elde edilen güvenirlikler arasındaki fark anlamlı değildir. Çoklu puanlanan madde yüzdesi %50 düzeyinde ise zor ve çok kolay test güçlüğü düzeyleri arasında anlamlı fark yoktur. Benzer şekilde orta ve kolay test güçlükleri ile elde edilen güvenirlikler arasındaki fark da anlamlı değildir ve bunlar en yüksek güvenirliğe sahiptir. ÇBMTK ve ÜDMTK modellerinde olduğu gibi Bifaktör model için de bütün çoklu puanlanan madde yüzdelerinde en yüksek güvenirlik test güçlüğü çok zor olduğunda elde edilmiştir.

Genel olarak, boyutlar arası korelasyon 0,2 olduğunda çoklu puanlanan madde yüzdelerinin hepsinde, test güçlüğü orta ve kolay olduğunda daha düşük RMSE ve daha yüksek güvenirlik elde edilirken test güçlüğü zor, çok kolay ve çok zor olduğunda daha yüksek RMSE ve daha düşük güvenirlikler elde edilmiştir.

Çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü arasındaki etkileşim etkisine ilişkin test güçlüğü basit ana etkilerinin incelenmesinin ardından modellerin basit ana etkileri ele alınmıştır. Bir başka ifadeyle test güçlüğüne ait düzeylerin her birinde (çok zor, zor, orta güçlük, kolay, çok kolay) yetenek kestiriminde kullanılan modellere (ÇBMTK, ÜDMTK ve Bifaktör) ait RMSE ve güvenirlik değerleri karşılaştırılmıştır. Bu amaçla gerçekleştirilen tekrarlı ANOVA sonuçları Tablo 26'da verilmiştir.

Tablo 26.

Alt Puanlar için Korelasyon 0,2 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması (RMSE, Güvenirlik)

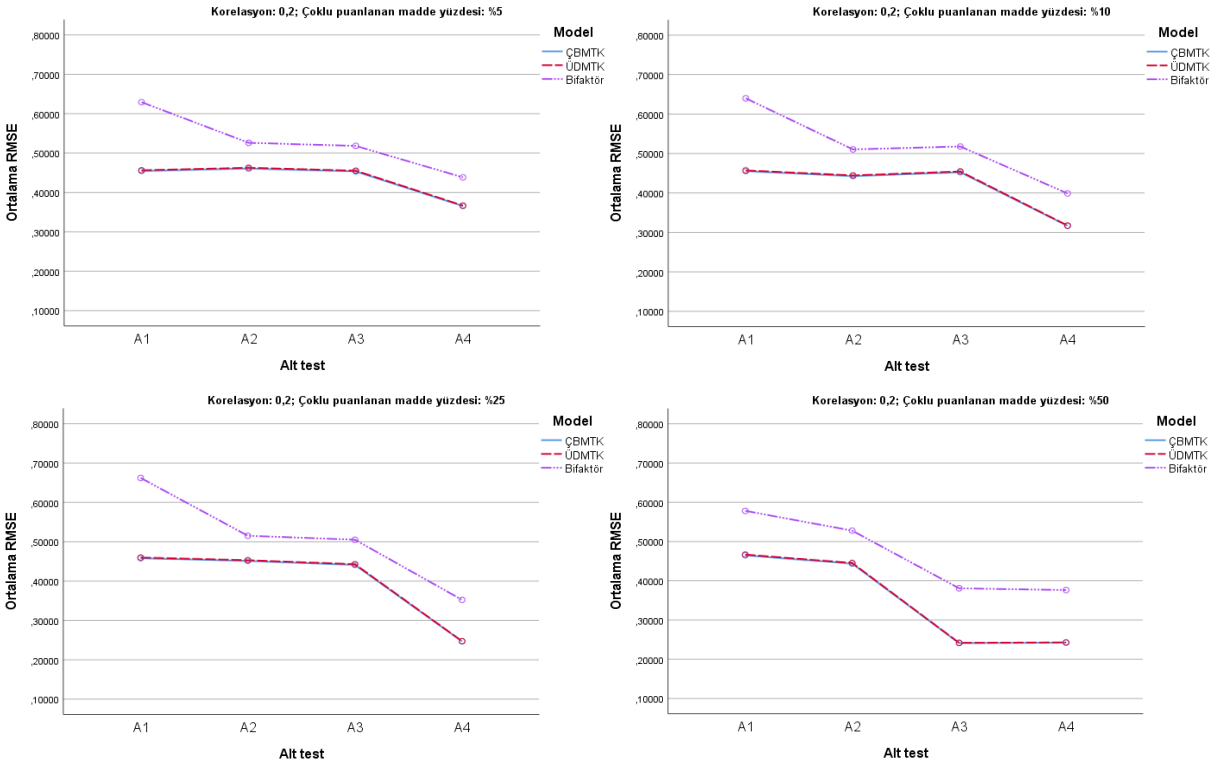
		RMSE		Güvenirlik	
ÇPMY	TG	F	η_p^2	F	η_p^2
%5	Çok zor	2129,147*	,681	2571,968*	,721
	Zor	1145,456*	,535	704,825*	,415
	Orta güçlük	1223,630*	,551	530,490*	,348
	Kolay	1694,500*	,630	897,398*	,474
	Çok kolay	1299,470*	,566	929,062*	,483
%10	Çok zor	2558,721*	,720	3030,671*	,753
	Zor	1196,326*	,546	771,861*	,437
	Orta güçlük	1717,217*	,633	1182,196*	,543
	Kolay	1188,425*	,544	636,996*	,390
	Çok kolay	1827,851*	,647	1348,247*	,575
%25	Çok zor	2315,872*	,699	3011,544*	,752
	Zor	2252,752*	,694	1805,811*	,645
	Orta güçlük	2309,194*	,699	1372,304*	,580
	Kolay	1753,089*	,638	863,964*	,465
	Çok kolay	1996,270*	,667	1483,064*	,598
%50	Çok zor	8346,290*	,893	4580,242*	,821
	Zor	11247,988*	,919	3744,888*	,790
	Orta güçlük	14614,512*	,936	4264,904*	,811
	Kolay	11527,193*	,921	3475,626*	,777
	Çok kolay	9047,672*	,901	3580,056*	,782

*p<,001; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

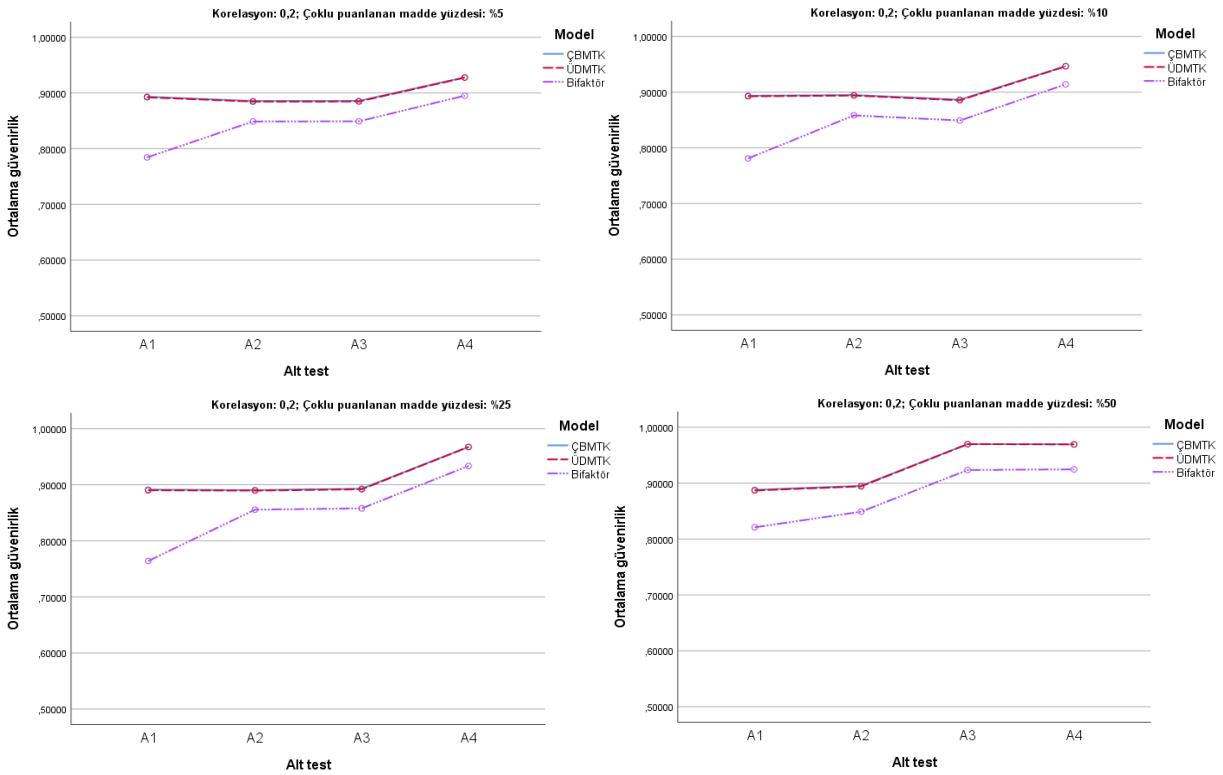
Tablo 26'da çoklu puanlanan madde yüzdesi ve test güçlüğünün tüm düzeylerinde RMSE ve güvenirlik değerlerinin her birinin kestirim modellerine göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. RMSE değerlerine yönelik elde edilen bu farkın kaynağını belirlemek amacıyla yapılan ikili karşılaştırmalar sonucunda, farkların hepsinin anlamlı olduğu ve RMSE değerlerinin düşükten yükseğe doğru sırasıyla ÇBMTK, ÜDMTK ve Bifaktör modellerinden elde edildiği görülmüştür (Ek 11). ÇBMTK ve ÜDMTK arasındaki fark anlamlı olsa da ortalama değerleri birbirlerine oldukça yakındır. Güvenirlik değerleri için durum bunun tam tersi olup en düşükten en yükseğe doğru sırasıyla

Bifaktör, ÜDMTK ve ÇBMTK modelleri ile elde edilmiştir. Güvenirlik açısından da ÇBMTK ve ÜDMTK için elde edilen değerlerin birbirine çok yakın olduğu görülmüştür.

Boyutlar arası korelasyon 0,2 düzeyi ve çoklu puanlanan madde yüzdesinin her bir düzeyi (%5, %10, %25 ve %50) için gerçekleştirilen analiz ile (Tablo 24) elde edilen model-test güçlüğü arasındaki anlamlı etkileşimin incelenmesinin ardından model-alt test arasındaki etkileşimin incelenmesine geçilmiştir. Tablo 24'te yer alan analiz sonuçları incelendiğinde bütün çoklu puanlanan madde yüzdesi düzeylerinde model-alt test arasındaki etkileşimin anlamlı ve büyük etkiye sahip olduğu tespit edilmiştir. Model-alt test arasındaki iki yönlü etkileşimlerin değerlendirilmesi için alt test basit ana etkisi tek yönlü ANOVA ile incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde alt testlerin birbiri arasındaki farkları görmek için dört alt test için hesaplanan RMSE ve güvenirlik değerleri ayrı ayrı karşılaştırılmıştır. Alt test basit ana etkisi için elde edilen sonuçlar Tablo 27'de verilmiştir. Çoklu puanlanan madde yüzdesinin her bir düzeyi için model ve alt test düzeylerinde elde edilen ortalama RMSE ve güvenirlik değerlerini içeren grafikler sırasıyla Şekil 26 ve Şekil 27'de yer almaktadır.



Şekil 26. Alt puanlar korelasyon 0,2 - çoklu puanlanan madde yüzdesi x model *x alt test (RMSE)*



Şekil 27. Alt puanlar korelasyon 0,2 - çoklu puanlanan madde yüzdesi x model x alt test (güvenirlik)

Tablo 27.

Alt Puanlar için Korelasyon 0,2 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Alt Testlerin Karşılaştırılması (RMSE, Güvenirlik)

		RMSE		Güvenirlik	
ÇPMY	Model	F	η_p^2	F	η_p^2
%5	ÇBMTK	2744,731*	,805	1550,837*	,700
	ÜDMTK	2712,423*	,803	1531,076*	,698
	Bifaktör	1169,030*	,638	756,954*	,533
%10	ÇBMTK	2744,731*	,805	2829,238*	,810
	ÜDMTK	2712,423*	,803	2794,428*	,808
	Bifaktör	1169,030*	,638	1262,755*	,655
%25	ÇBMTK	6240,435*	,904	6361,315*	,905
	ÜDMTK	6189,174*	,903	6299,112*	,905
	Bifaktör	1882,303*	,739	2765,548*	,806
%50	ÇBMTK	17503,993*	,963	856,310*	,632
	ÜDMTK	17409,015*	,963	852,329*	,631
	Bifaktör	4157,608*	,862	357,032*	,418

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 27'de çoklu puanlanan madde yüzdesi ve kestirim modellerinin tüm düzeylerinde RMSE ve güvenirlik değerlerinin her birinin alt testlere göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. Anlamlı bulunan bu farkların kaynağının

belirlenmesi için ikili karşılaştırmalar yapılmıştır (Ek 12). Buna göre çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda her bir alt test için ÇBMTK ve ÜDMTK modelleri ile elde edilen RMSE değerleri ilk üç alt testte birbirine yakın ve alt test 4'te en düşüktür. Çoklu puanlanan madde yüzdesi %50 olduğunda ise tamamı çoklu puanlanan maddelerden oluşan alt test 3 ve alt test 4 için elde edilen RMSE değerleri arasındaki fark anlamlı değildir ve diğer alt testlerden daha düşük RMSE değerine sahiptirler. Bifaktör model için çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda en yüksek RMSE alt test 1 ile hesaplanırken alt test 2 ve alt test 3 arasındaki fark anlamlı değildir. En düşük RMSE alt test 4 ile elde edilmiştir. Çoklu puanlanan madde yüzdesi %50 olduğunda ise alt test 3 ve alt test 4 ile elde edilen RMSE değerleri arasındaki fark anlamlı değil ve diğerlerinden anlamlı olarak daha düşüktür.

Güvenirlik değerleri ile yapılan analizlerde çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda her bir alt test için ÇBMTK ve ÜDMTK modelleri ile elde edilen güvenirlik değerleri ilk üç alt testte birbirine yakın ve alt test 4'te en yüksektir. Çoklu puanlanan madde yüzdesi %50 olduğunda ise alt test 3 ve alt test 4 için elde edilen güvenirlik değerleri arasındaki fark anlamlı değildir ve diğer alt testlerden daha yüksek güvenirlik değerine sahiptir. Bifaktör model için çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda en düşük güvenirlik alt test 1 ve en yüksek güvenirlik alt test 4 ile hesaplanırken alt test 2 ve alt test 3 arasındaki fark anlamlı değildir. Çoklu puanlanan madde yüzdesi %50 olduğunda ise diğer modeller ile benzer şekilde alt test 3 ve alt test 4 ile elde edilen güvenirlik değerleri arasındaki fark anlamlı değil ve diğerlerinden anlamlı olarak daha yüksektir.

Özetle, ÇBMTK ve ÜDMTK modelleri ile elde edilen hem RMSE hem güvenirlik değerleri için çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda tamamı ikili puanlanan maddelerden oluşan ilk 3 alt test birbirine yakın sonuçlar üretirken alt test 4 daha düşük RMSE ve daha yüksek güvenirlik değerlerine sahiptir. Çoklu puanlanan madde yüzdesi %50 olduğunda tamamı çoklu puanlanan maddelerden oluşan alt test 3 ve alt test 4 için elde edilen değerler birbirine yakındır ve diğer alt testlerden daha düşük RMSE ve daha yüksek güvenirlik değerlerine sahiptir. Bifaktör model için de sonuçlar benzer olup tek fark alt test 1'e ait RMSE değerlerinin diğer alt testlerden yüksek ve güvenirlik değerlerinin düşük olmasıdır.

Korelasyon 0,5 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

Korelasyon 0,5 düzeyinde model, çoklu puanlanan madde yüzdesi ve test güçlüğü değişkenleri arasındaki üç yönlü etkileşim etkisinin incelenmesi amacıyla bu değişkenler ile RMSE ve

güvenirlik değerlerinin her biri için ayrı ayrı faktöriyel karma ANOVA yapılmıştır. Analiz sonuçları Tablo 28'de yer almaktadır.

Tablo 28.

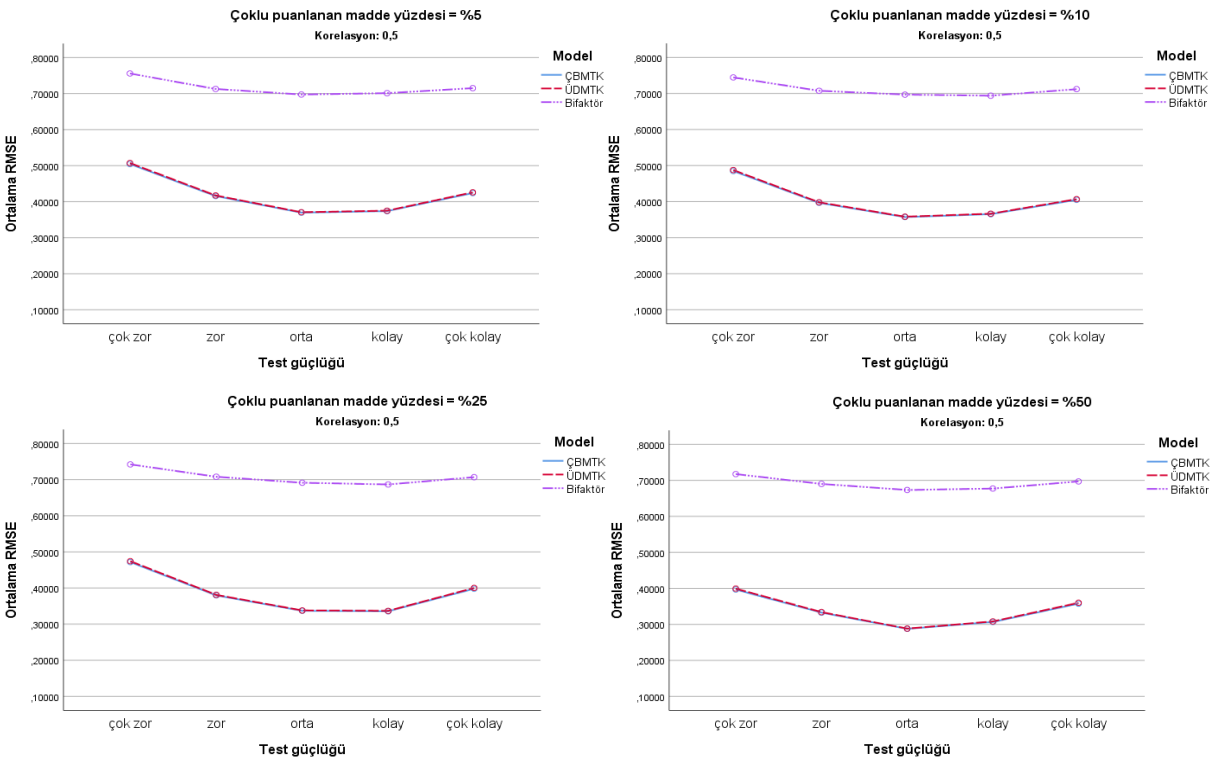
Alt Puanlar için Korelasyon 0,5 Düzeyinde Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Faktöriyel Karma ANOVA (RMSE, Güvenirlik)

	RMSE		Güvenirlik	
Kaynak	F	η_p^2	F	η_p^2
Gruplar içi				
M	1559487,970*	,995	2062797,824*	,996
M *x* AT	3106,999*	,539	4514,426*	,629
M *x* ÇPMY	2233,909*	,457	54,388*	,020
M *x* TG	2839,651*	,587	12,412*	,006
M *x* TG *x* ÇPMY	29,796*	,043	27,314*	,039
Gruplar arası				
AT	5550,349*	,676	2658,788*	,500
ÇPMY	1194,637*	,310	348,468*	,116
TG	2166,115*	,521	1156,247*	,367
ÇPMY *x* TG	15,515*	,023	16,593*	,024

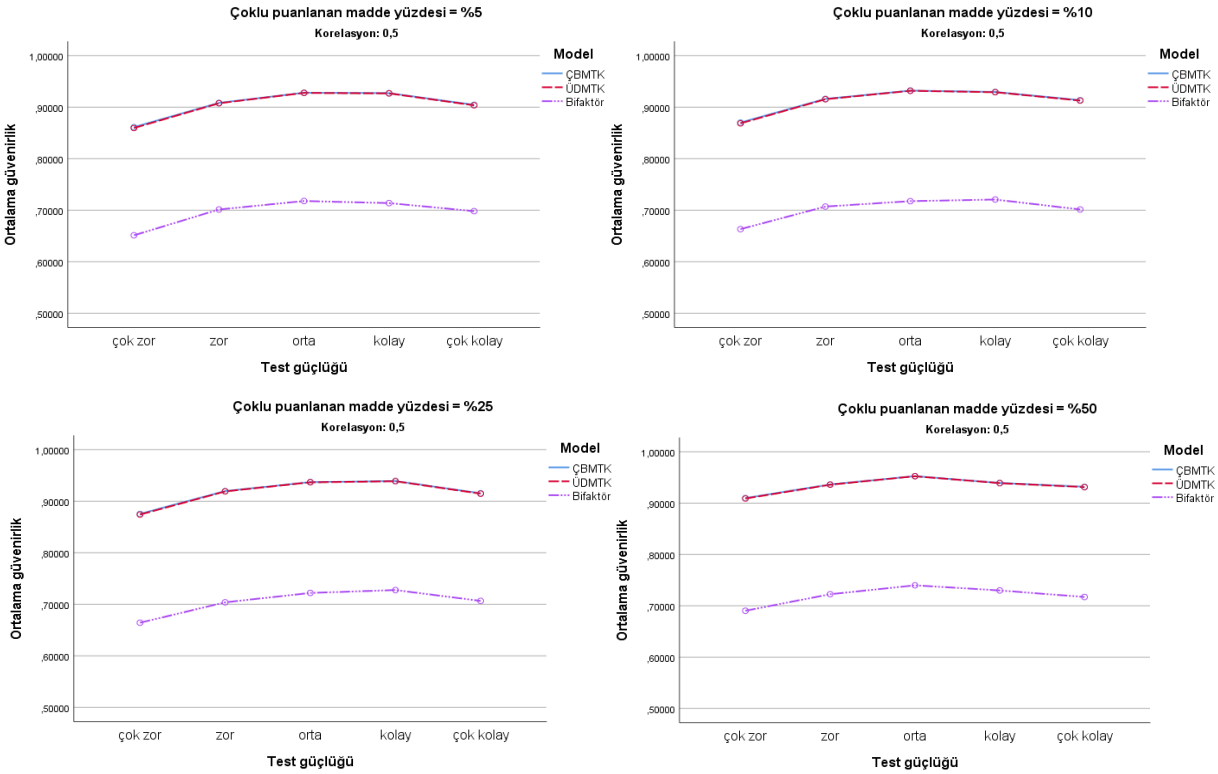
*p<,001; M: model; AT: alt test; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Boyutlar arası korelasyon 0,5 düzeyinde RMSE ve güvenirlik değerlerinin her biri için gerçekleştirilen ANOVA sonucunda model, test güçlüğü ve çoklu puanlanan madde yüzdesi değişkenleri arasındaki üç yönlü etkileşimin anlamlı ve etki büyüklüklerinin düşük düzeyde olduğu tespit edilmiştir. Bu üç yönlü etkileşimleri incelemek amacıyla basit iki yönlü etkileşim yaklaşımı kullanılmıştır. Bunun için çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model-test güçlüğü ve model-alt test arasındaki etkileşim etkileri incelenmiştir.

Öncelikle çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model, test güçlüğü ve alt test (blok) değişkenleri ile RMSE ve güvenirlik değerleri için karma ANOVA yapılmış ve sonuçlar Tablo 29'da sunulmuştur. Çoklu puanlanan madde yüzdesinin her bir düzeyi için model ve test güçlüğü düzeylerinde elde edilen ortalama RMSE ve güvenirlik değerlerini içeren grafikler sırasıyla Şekil 28 ve Şekil 29'da yer almaktadır.



Şekil 28. Alt puanlar korelasyon 0,5 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE)



Şekil 29. Alt puanlar korelasyon 0,5 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik)

Tablo 29.

Alt Puanlar için Korelasyon 0,5 Düzeyinde Her Bir Çoklu Puanlanan Madde Yüzdesi için Model, Test Güçlüğü ve Alt Test Değişkenleri ile ANOVA Sonuçları (RMSE, Güvenirlik)

			RMSE		Güvenirlik	
ÇPMY	Kaynak		F	η_p^2	F	η_p^2
%5	Gruplar içi	M	2038116,421*	,999	511080,261*	,996
		M *x* AT	2113,212*	,761	836,545*	,557
		M *x* TG	4655,637*	,903	20,355*	,039
	Gruplar arası	AT	2348,348*	,780	1608,132*	,708
		TG	3280,292*	,868	2058,792*	,805
%10	Gruplar içi	M	1266432,191*	,998	636484,748*	,997
		M *x* AT	2347,489*	,780	1446,827*	,685
		M *x* TG	2572,927*	,838	30,550*	,058
	Gruplar arası	AT	5830,710*	,898	3441,629*	,838
		TG	2971,191*	,856	1854,597*	,788
%25	Gruplar içi	M	1321958,133*	,998	750594,062*	,997
		M *x* AT	6030,220*	,901	2665,926*	,801
		M *x* TG	3095,813*	,861	30,331*	,057
	Gruplar arası	AT	15995,150*	,960	7883,747*	,922
		TG	4114,173*	,892	2473,524*	,832
%50	Gruplar içi	M	3777600,974*	,999	526384,800*	,996
		M *x* AT	18881,078*	,966	1023,064*	,606
		M *x* TG	4018,820*	,890	28,663*	,054
	Gruplar arası	AT	13791,360*	,954	631,803*	,488
		TG	1477,454*	,748	64,284*	,114

*p<,001; M: model; AT: alt test; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 29'da yer alan boyutlar arası korelasyon 0,5 düzeyi ve çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) gerçekleştirilen analiz sonuçları incelendiğinde model-test güçlüğü ve model-alt test arasındaki etkileşimin RMSE ve güvenirlik değerleri için anlamlı olduğu görülmektedir. RMSE değerleri için yapılan analizde bu etkileşimler, büyük etki büyüklüğüne sahipken güvenirlik için düşük etki büyüklüğüne sahiptir.

Öncelikle anlamlı bulunan iki yönlü etkileşimler arasında model-test güçlüğü etkileşimi ele alınmıştır. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi için sırasıyla model ve test güçlüğü değişkenlerinin her bir düzeyinde diğer değişkene ait basit ana etkiler tek yönlü ANOVA ile incelenmiştir. İlk olarak test güçlüğü basit ana etkisi incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde test güçlük düzeyleri (çok zor, zor, orta güçlük, kolay, çok kolay) için hesaplanan RMSE ve güvenirlik değerleri ayrı ayrı karşılaştırılmıştır. Test güçlüğü basit ana etkisi için elde edilen sonuçlar Tablo 30'da verilmiştir.

Tablo 30.

Alt Puanlar için Korelasyon 0,5 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması (RMSE, güvenirlik)

		RMSE		Güvenirlik	
ÇPMY	Model	F	η_p^2	F	η_p^2
%5	ÇBMTK	5002,674*	,909	3813,948*	,885
	ÜDMTK	4963,390*	,909	3774,628*	,883
	Bifaktör	742,614*	,599	736,495*	,597
%10	ÇBMTK	3981,457*	,889	2740,405*	,846
	ÜDMTK	3965,478*	,888	2730,874*	,846
	Bifaktör	664,371*	,572	723,377*	,592
%25	ÇBMTK	5365,066*	,915	3317,311*	,869
	ÜDMTK	5308,216*	,914	3307,338*	,869
	Bifaktör	893,867*	,642	1057,382*	,680
%50	ÇBMTK	1992,333*	,800	50,418*	,092
	ÜDMTK	2030,319*	,803	52,114*	,095
	Bifaktör	449,265*	,474	96,234*	,162

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 30'da çoklu puanlanan madde yüzdesi ve kestirim modellerinin tüm düzeylerinde RMSE ve güvenirlik değerlerinin her birinin test güçlüklerine göre anlamlı farklılık gösterdiği görülmektedir. RMSE için etki büyükleri yüksektir. Güvenirlik için yapılan analizlerin çoğunda etki büyüklüğü yüksekken çoklu puanlanan madde yüzdesi %50 olduğunda ÇBMTK ve ÜDMTK ile yapılan analizlerde test güçlükleri arasındaki farka ait etki orta düzeydedir. Genel olarak Bifaktör model için yapılan analizlerde etki büyüklüğünün diğer modellere göre daha düşük olduğu görülmektedir. Anlamlı bulunan bu farkların kaynağının belirlenmesi için ikili karşılaştırmalar yapılmıştır (Ek 13). Buna göre çoklu puanlanan madde yüzdesi %10 ve %50 olduğunda ÇBMTK ve ÜDMTK modelleri ile test güçlük düzeylerine göre elde edilen bütün RMSE değerleri anlamlı olarak birbirinden farklılaşmaktadır. RMSE değerleri test güçlük düzeylerinde en düşükten en yükseğe doğru orta, kolay, zor, çok kolay ve çok zor şeklinde sıralanmaktadır. ÇBMTK ve ÜDMTK modelleri ile çoklu puanlanan madde yüzdesi %5 ve %25 olduğunda ise sıralama aynı olup tek fark orta güçlük ve kolay test güçlük düzeylerinde elde edilen RMSE değerleri arasındaki farkın anlamlı olmamasıdır. Bifaktör model için en düşük RMSE çoklu puanlanan madde yüzdesi %25 olduğunda test güçlük düzeyi kolay ile elde edilirken diğer çoklu puanlanan madde yüzdeleri için test güçlüğü orta ve kolay olduğunda elde edilmiş ve bu güçlük düzeyleri arasındaki fark anlamlı bulunmamıştır. Çoklu puanlanan madde yüzdesi %5, %10 ve %25 için test güçlüğü zor ve çok kolay olduğunda elde edilen RMSE değerleri arasındaki fark da anlamlı değildir. Bifaktör model için de diğer modeller ile benzer şekilde en yüksek RMSE test güçlüğü çok zor olduğunda elde edilmiştir.

Güvenirlik değerleri, ÇBMTK ve ÜDMTK modelleri için test güçlük düzeylerinde en düşükten en yükseğe doğru orta, kolay, zor, çok kolay ve çok zor şeklinde sıralanmaktadır. Fakat çoklu puanlanan madde yüzdesi %5 ve %25 olduğunda orta güçlük ve kolay test güçlüğü için elde edilen güvenirlik değerleri arasındaki fark ve çoklu puanlanan madde yüzdesi %10 olduğunda test güçlüğü zor ve çok kolay için elde edilen güvenirlikler arasındaki fark anlamlı değildir. Çoklu puanlanan madde yüzdesi %50 için test güçlüğü zor, kolay ve çok kolay iken elde edilen güvenirlikler arasındaki fark anlamlı değildir. Bütün durumlarda en düşük güvenirlik test güçlüğü çok zor olduğunda elde edilmiştir. Bifaktör model için en yüksek güvenirlik çoklu puanlanan madde yüzdesi %25 olduğunda test güçlük düzeyi kolay ile elde edilirken diğer çoklu puanlanan madde yüzdeleri için test güçlüğü orta ve kolay olduğunda elde edilmiş ve bu güçlük düzeyleri arasındaki fark anlamlı bulunmamıştır. Çoklu puanlanan madde yüzdesi %5, %25 ve %50 için test güçlüğü zor ve çok kolay olduğunda elde edilen güvenirlik değerleri arasındaki fark anlamlı değildir. Bifaktör model için de diğer modeller ile benzer şekilde en düşük güvenirlik test güçlüğü çok zor olduğunda elde edilmiştir. Ancak bütün modeller için çoklu puanlanan madde yüzdesi %50 düzeyinde olduğunda test güçlüklerine göre elde edilen ortalama güvenirlik değerleri birbirine yakındır.

Genel olarak, boyutlar arası korelasyon 0,5 olduğunda, bütün çoklu puanlanan madde yüzdeleri için hesaplanan RMSE değerlerinin test kolaylaştıkça ve zorlaştıkça yani orta düzeyden uzaklaştıkça arttığı söylenebilir. Benzer şekilde test güçlüğü arttıkça ve kolaylaştıkça elde edilen güvenirlikler artış göstermiştir.

Çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü arasındaki etkileşim etkisine ilişkin test güçlüğü basit ana etkilerinin incelenmesinin ardından modellerin basit ana etkileri ele alınmıştır. Bir başka ifadeyle test güçlüğüne ait düzeylerin her birinde (çok zor, zor, orta güçlük, kolay, çok kolay) yetenek kestiriminde kullanılan modellere (ÇBMTK, ÜDMTK ve Bifaktör) ait RMSE ve güvenirlik değerleri karşılaştırılmıştır. Bu amaçla gerçekleştirilen tekrarlı ANOVA sonuçları Tablo 31'de verilmiştir.

Tablo 31.

Alt Puanlar için Korelasyon 0,5 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması (RMSE, Güvenirlik)

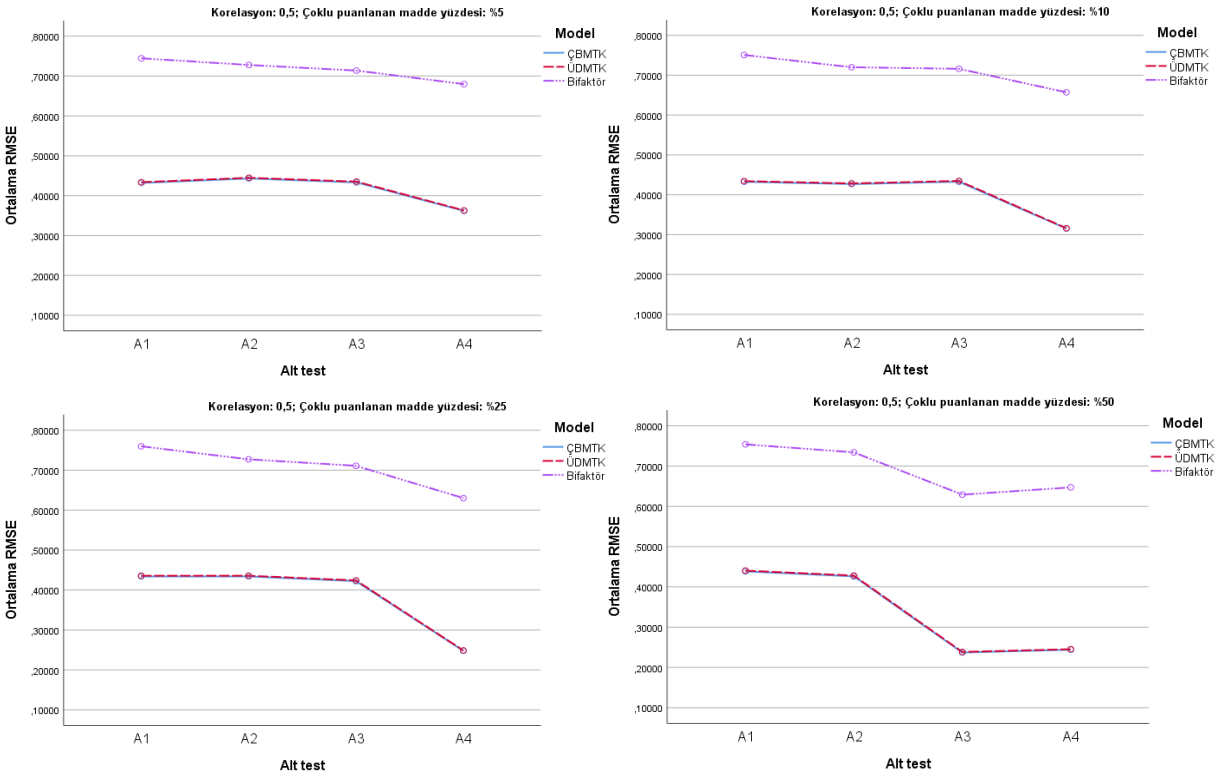
		RMSE		Güvenirlik	
ÇPMY	TG	F	η_p^2	F	η_p^2
%5	Çok zor	146416,241*	,993	51301,779*	,981
	Zor	203611,674*	,995	51408,698*	,981
	Orta güçlük	247685,457*	,996	54183,129*	,982
	Kolay	247102,103*	,996	55536,812*	,982
	Çok kolay	195462,080*	,995	50709,025*	,981
%10	Çok zor	93548,380*	,989	61606,999*	,984
	Zor	131215,710*	,992	63207,102*	,984
	Orta güçlük	156005,567*	,994	67004,613*	,985
	Kolay	146233,304*	,993	63284,457*	,985
	Çok kolay	128443,281*	,992	64942,347*	,985
%25	Çok zor	96841,565*	,990	75362,989*	,987
	Zor	140121,384*	,993	77583,848*	,987
	Orta güçlük	162699,335*	,994	77059,967*	,987
	Kolay	159813,799*	,994	74508,988*	,987
	Çok kolay	125003,154*	,992	72795,894*	,987
%50	Çok zor	315589,692*	,997	57454,472*	,983
	Zor	391671,392*	,997	56530,435*	,983
	Orta güçlük	456126,991*	,998	57177,338*	,983
	Kolay	420714,776*	,998	55037,552*	,982
	Çok kolay	353451,404*	,997	56341,580*	,983

*p<,001; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

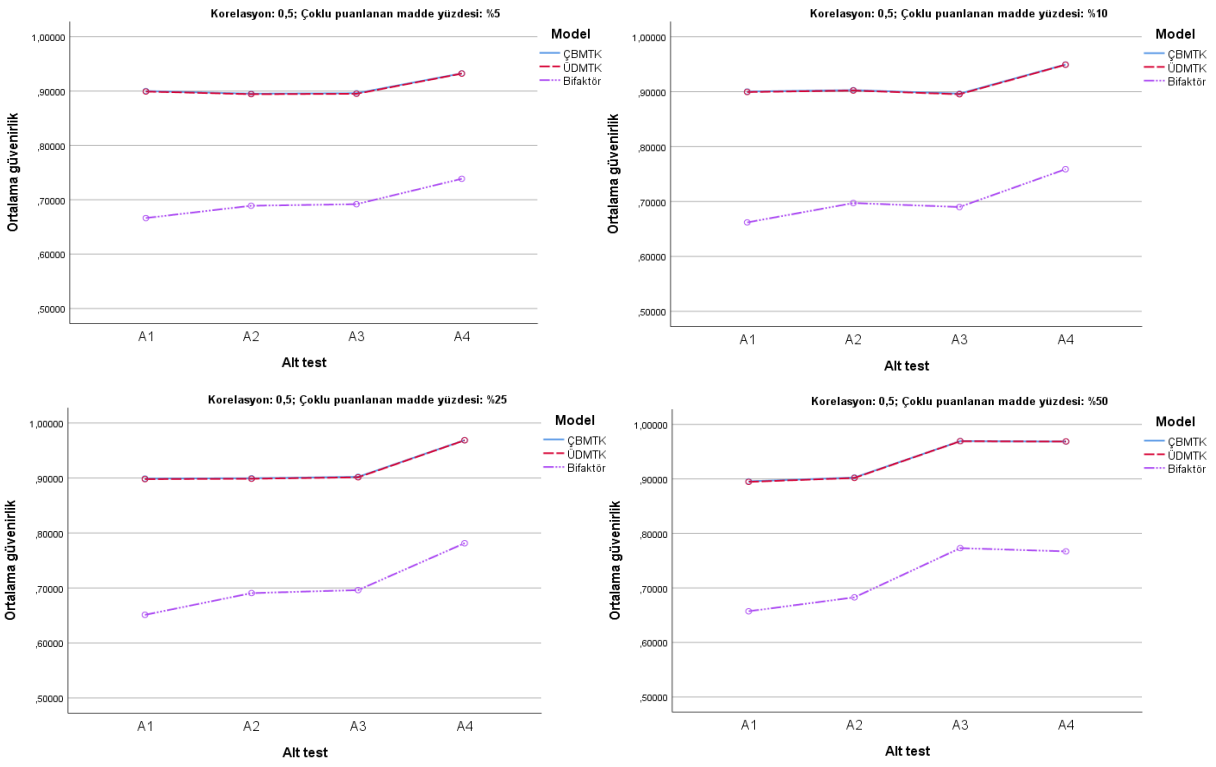
Tablo 31'de çoklu puanlanan madde yüzdesi ve test güçlüğünün tüm düzeylerinde RMSE ve güvenirlik değerlerinin her birinin kestirim modellerine göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. Anlamlı bulunan bu farkların hangi modeller arasında olduğunun incelenmesi için ikili karşılaştırmalar yapılmıştır (Ek 13). Buna göre korelasyon 0,2 ile benzer şekilde korelasyon düzeyi 0,5 olduğunda da bütün çoklu puanlanan madde yüzdelerinde ve test güçlük düzeylerinde modeller arasındaki bütün ikili karşılaştırmalar anlamlıdır ve RMSE değerleri düşükten yükseğe doğru sırasıyla ÇBMTK, ÜDMTK ve Bifaktör modelleri ile elde edilmiştir. Güvenirlik değerleri için durum bunun tam tersi olup modellerden elde edilen değerler en düşükten en yükseğe doğru Bifaktör, ÜDMTK ve ÇBMTK şeklinde sıralanmaktadır. Hem güvenirlik hem de RMSE değerleri açısından ÇBMTK ve ÜDMTK modelleri arasındaki fark anlamlı olsa da ortalama değerleri birbirine oldukça yakındır.

Boyutlar arası korelasyon 0,5 düzeyi ve çoklu puanlanan madde yüzdesinin her bir düzeyi (%5, %10, %25 ve %50) için gerçekleştirilen analiz ile (Tablo 29) elde edilen model-test güçlüğü arasındaki anlamlı etkileşimin incelenmesinin ardından model-alt test arasındaki etkileşimin

incelenmesine geçilmiştir. Tablo 29'da yer alan analiz sonuçları incelendiğinde bütün çoklu puanlanan madde yüzdesi düzeylerinde model-alt test arasındaki etkileşimin anlamlı ve büyük etkiye sahip olduğu tespit edilmiştir. Model-alt test arasındaki iki yönlü etkileşimlerin değerlendirilmesi için alt test basit ana etkisi tek yönlü ANOVA ile incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde alt testlerin birbiri arasındaki farkları görmek için dört alt test için hesaplanan RMSE ve güvenirlik değerleri ayrı ayrı karşılaştırılmıştır. Alt test basit ana etkisi için elde edilen sonuçlar Tablo 32'de verilmiştir. Çoklu puanlanan madde yüzdesinin her bir düzeyi için model ve alt test düzeylerinde elde edilen ortalama RMSE ve güvenirlik değerlerini içeren grafikler sırasıyla Şekil 30 ve Şekil 31'de yer almaktadır.



Şekil 30. Alt puanlar korelasyon 0,5 - çoklu puanlanan madde yüzdesi x model x alt test (*RMSE)*



Şekil 31. Alt puanlar korelasyon 0,5 - çoklu puanlanan madde yüzdesi x model x alt test (güvenirlik)

Tablo 32.

Alt Puanlar için Korelasyon 0,5 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Alt Testlerin Karşılaştırılması (RMSE, Güvenirlik)

ÇPMY	Model	F	η_p^2	F	η_p^2
%5	ÇBMTK	2975,505*	,818	2065,409*	,757
	ÜDMTK	2930,246*	,815	2028,037*	,753
	Bifaktör	1302,990*	,662	1197,929*	,643
%10	ÇBMTK	6482,774*	,907	3422,638*	,838
	ÜDMTK	6380,495*	,906	3363,969*	,835
	Bifaktör	3123,204*	,825	2844,653*	,811
%25	ÇBMTK	17967,695*	,964	7433,084*	,918
	ÜDMTK	17563,554*	,964	7323,750*	,917
	Bifaktör	7157,107*	,915	6446,944*	,907
%50	ÇBMTK	16713,619*	,962	426,722*	,391
	ÜDMTK	16535,336*	,961	435,682*	,396
	Bifaktör	6995,565*	,913	1199,816*	,644

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 32'de çoklu puanlanan madde yüzdesi ve kestirim modellerinin tüm düzeylerinde RMSE ve güvenirlik değerlerinin her birinin alt testlere göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. Anlamlı bulunan bu farkların kaynağının belirlenmesi için ikili karşılaştırmalar yapılmıştır (Ek 14). Buna göre korelasyon 0,5 iken çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda her bir alt test için ÇBMTK ve ÜDMTK

modelleri ile elde edilen RMSE değerleri ilk üç alt testte birbirine yakın ve alt test 4'te en düşüktür. Çoklu puanlanan madde yüzdesi %50 olduğunda ise tamamı çoklu puanlanan maddelerden oluşan alt test 3 ve alt test 4 için elde edilen RMSE değerleri arasındaki fark anlamlı değildir ve diğer alt testlerden daha düşük RMSE değerine sahiptirler. Bifaktör model için çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda en yüksek RMSE alt test 1 ile hesaplanırken alt test 2, alt test 3 ve alt test 4 sırasıyla bunu takip etmektedir. En düşük RMSE alt test 4 ile elde edilmiştir. Çoklu puanlanan madde yüzdesi %50 olduğunda ise alt test 3 ve alt test 4 ile elde edilen RMSE değerleri arasındaki fark anlamlı değil ve diğerlerinden anlamlı olarak daha düşüktür.

Güvenirlik değerleri açısından da benzer sonuçlar elde edilmiştir. Çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda ÇBMTK ve ÜDMTK modelleri ile alt testlere göre elde edilen güvenirlik değerleri ilk üç alt testte birbirine yakınken alt test 4 en yüksek güvenirlik değerine sahiptir. Çoklu puanlanan madde yüzdesi %50 olduğunda ise alt test 3 ve alt test 4 için elde edilen güvenirlik değerleri arasındaki fark anlamlı değildir ve diğer alt testlerden daha yüksek güvenirlik değerine sahiptirler. Bifaktör model için çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda en düşük güvenirlik alt test 1 ile hesaplanırken alt test 2, alt test 3 ve alt test 4 sırasıyla bunu takip etmektedir. En yüksek güvenirlik alt test 4 ile elde edilmiştir. Çoklu puanlanan madde yüzdesi %50 olduğunda ise alt test 3 ve alt test 4 ile elde edilen güvenirlik değerleri arasındaki fark anlamlı değil ve diğerlerinden anlamlı olarak daha yüksektir.

Özetle, ÇBMTK ve ÜDMTK modelleri ile elde edilen hem RMSE hem güvenirlik için çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda tamamı ikili puanlanan maddelerden oluşan ilk 3 test birbirine yakın sonuçlar üretirken alt test 4 daha düşük RMSE ve daha yüksek güvenirlik değerlerine sahiptir. Çoklu puanlanan madde yüzdesi %50 olduğunda tamamı çoklu puanlanan maddelerden oluşan alt test 3 ve alt test 4 için elde edilen değerler birbirine yakındır ve diğer alt testlerden daha düşük RMSE ve daha yüksek güvenirlik değerlerine sahiptir.

Korelasyon 0,8 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

Korelasyon 0,8 düzeyinde model, çoklu puanlanan madde yüzdesi ve test güçlüğü değişkenleri arasındaki üç yönlü etkileşim etkisinin incelenmesi amacıyla bu değişkenler ile RMSE ve güvenirlik değerlerinin her biri için ayrı ayrı faktöriyel karma ANOVA yapılmıştır. Analiz sonuçları Tablo 33'de yer almaktadır.

Tablo 33.

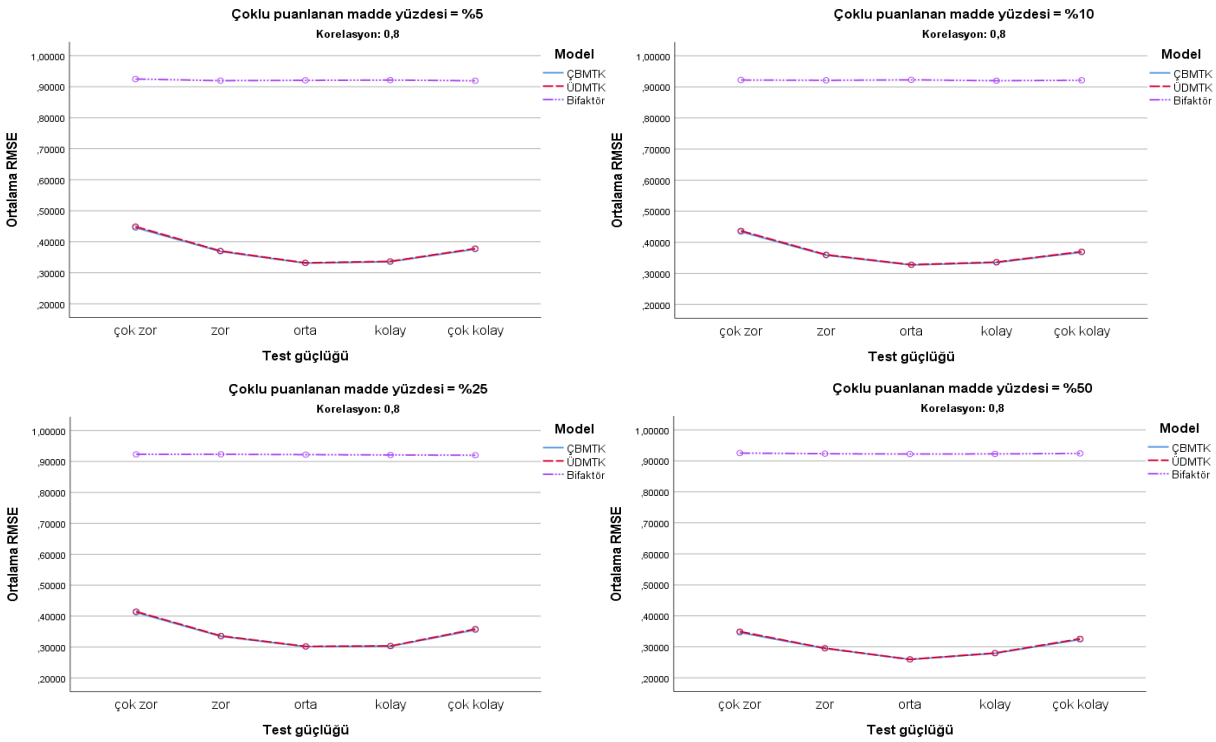
Alt Puanlar için Korelasyon 0,8 Düzeyinde Model, Test Güçlüğü ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Faktöriyel Karma ANOVA (RMSE, Güvenirlik)

	RMSE		Güvenirlik	
Kaynak	F	η_p^2	F	η_p^2
Gruplar içi				
M	5985047,862*	,999	7003728,404*	,999
M *x* AT	3310,325*	,555	5805,684*	,686
M *x* ÇPMY	4898,303*	,648	127,103*	,046
M *x* TG	6151,378*	,755	53,999*	,026
M *x* TG *x* ÇPMY	70,888*	,096	26,193*	,038
Gruplar arası				
AT	4849,723*	,646	5423,908*	,671
ÇPMY	1415,735*	,347	779,879*	,227
TG	2059,656*	,508	1488,814*	,427
ÇPMY *x* TG	23,385*	,034	20,616*	,030

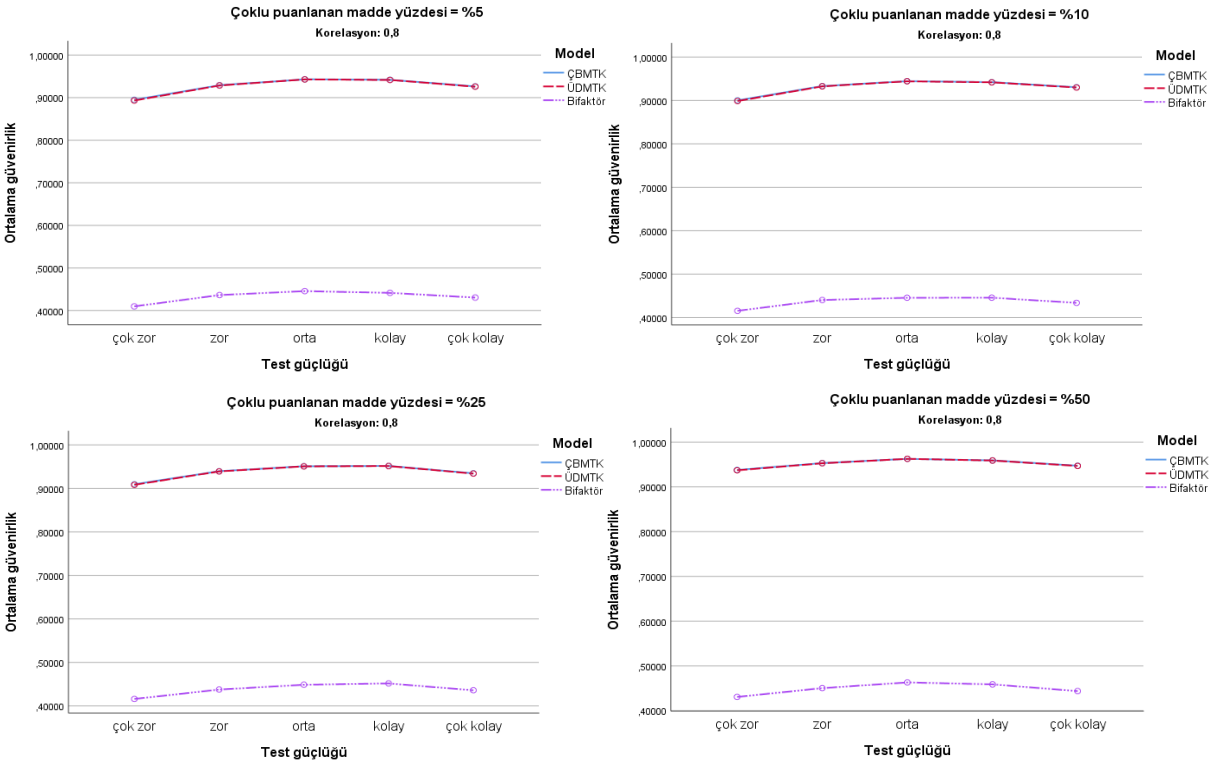
*p<,001; M: model; AT: alt test; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Boyutlar arası korelasyon 0,8 düzeyinde RMSE ve güvenirlik değerlerinin her biri için gerçekleştirilen ANOVA sonucunda model, test güçlüğü ve çoklu puanlanan madde yüzdesi değişkenleri arasındaki üç yönlü etkileşimin anlamlı olduğu tespit edilmiştir. RMSE için etki büyüklüğü orta düzeyde iken güvenirlik için düşük düzeydedir. Bu üç yönlü etkileşimleri incelemek amacıyla basit iki yönlü etkileşim yaklaşımı kullanılmıştır. Bunun için çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model-test güçlüğü ve model-alt test arasındaki etkileşim etkileri incelenmiştir.

Öncelikle çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model, test güçlüğü ve alt test (blok) değişkenleri ile RMSE ve güvenirlik değerleri için karma ANOVA yapılmış ve sonuçlar Tablo 34'te sunulmuştur. Çoklu puanlanan madde yüzdesinin her bir düzeyi için model ve test güçlüğü düzeylerinde elde edilen ortalama RMSE ve güvenirlik değerlerini içeren grafikler sırasıyla Şekil 32 ve Şekil 33'te yer almaktadır.



Şekil 32. Alt puanlar korelasyon 0,8 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (RMSE)



Şekil 33. Alt puanlar korelasyon 0,8 –çoklu puanlanan madde yüzdesi x model x test güçlüğü (güvenirlik)

Tablo 34.

Alt Puanlar için Korelasyon 0,8 Düzeyinde Her Bir Çoklu Puanlanan Madde Yüzdesi için Model, Test Güçlüğü ve Alt Test Değişkenleri İle ANOVA Sonuçları (RMSE)

			RMSE		Güvenirlik	
ÇPMY	Kaynak		F	η_p^2	F	η_p^2
%5	Gruplar içi	M	14649085,590*	1,000	3166511,790*	,999
		M *x* AT	2482,108*	,789	1305,890*	,663
		M *x* TG	19676,636*	,975	97,899*	,164
	Gruplar arası	AT	2820,144*	,809	1876,712*	,739
		TG	3812,432*	,884	1502,493*	,751
%10	Gruplar içi	M	18100862,331*	1,000	3852100,825*	,999
		M *x* AT	2006,020*	,751	3731,369*	,849
		M *x* TG	21064,272*	,977	106,685*	,176
	Gruplar arası	AT	4458,171*	,870	3540,671*	,842
		TG	3274,442*	,868	1393,991*	,737
%25	Gruplar içi	M	8792910,081*	1,000	4547864,747*	1,000
		M *x* AT	10827,418*	,942	6340,202*	,905
		M *x* TG	10935,287*	,956	51,200*	,093
	Gruplar arası	AT	17155,606*	,963	9251,191*	,933
		TG	3882,989*	,886	1543,928*	,756
%50	Gruplar içi	M	15001377,312*	1,000	4270663,752*	1,000
		M *x* AT	29464,324*	,978	5799,265*	,897
		M *x* TG	8992,492*	,948	27,527*	,052
	Gruplar arası	AT	22184,612*	,971	10554,679*	,941
		TG	2518,646*	,835	673,984*	,575

*p<,001; M: model; AT: alt test; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 34'te yer alan boyutlar arası korelasyon 0,8 düzeyi ve çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) gerçekleştirilen analiz sonuçları incelendiğinde model-test güçlüğü ve model-alt test arasındaki etkileşimin RMSE ve güvenirlik değerleri için anlamlı olduğu görülmektedir. RMSE değerleri için yapılan analizde bu etkileşimler, büyük etki büyüklüğüne sahiptir. Güvenirlik için ise model-test güçlüğü etkileşimine ait etki büyüklüğünün düşük, orta ya da orta düzeye çok yakın olduğu görülmektedir.

Öncelikle model-test güçlüğü etkileşimi ele alınmıştır. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi için sırasıyla model ve test güçlüğü değişkenlerinin her bir düzeyinde diğer değişkene ait basit ana etkiler tek yönlü ANOVA ile incelenmiştir. İlk olarak test güçlüğü basit ana etkisi incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde test güçlük düzeyleri (çok zor, zor, orta güçlük, kolay, çok kolay) için hesaplanan RMSE ve güvenirlik değerleri ayrı ayrı karşılaştırılmıştır. Test güçlüğü basit ana etkisi için elde edilen sonuçlar Tablo 35'te verilmiştir.

Tablo 35.

Alt Puanlar için Korelasyon 0,8 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Test Güçlüklerinin Karşılaştırılması (RMSE, Güvenirlik)

		RMSE		Güvenirlik	
ÇPMY	Model	F	η_p^2	F	η_p^2
%5	ÇBMTK	6413,520*	,928	5204,558*	,913
	ÜDMTK	6524,344*	,929	5252,998*	,913
	Bifaktör	27,106*	,052	265,799*	,348
%10	ÇBMTK	6091,50*7	,924	4468,633*	,900
	ÜDMTK	6151,069*	,925	4522,864*	,901
	Bifaktör	4,952*	,010	251,514*	,336
%25	ÇBMTK	6600,642*	,930	4507,629*	,901
	ÜDMTK	6615,058*	,930	4510,311*	,901
	Bifaktör	7,218*	,014	352,059*	,414
%50	ÇBMTK	4024,136*	,890	1590,612*	,762
	ÜDMTK	4145,961*	,893	1579,111*	,760
	Bifaktör	9,512*	,019	272,488*	,354

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 35'te çoklu puanlanan madde yüzdesi ve kestirim modellerinin tüm düzeylerinde RMSE ve güvenirlik değerlerinin her birinin test güçlüklerine göre anlamlı farklılık gösterdiği görülmektedir. Genel olarak etki büyüklükleri yüksektir fakat RMSE değerlerinde Bifaktör model için bütün çoklu puanlanan madde yüzdesi düzeylerinde test güçlükleri arasındaki farka ait etki büyüklükleri düşük düzeydedir. Anlamlı bulunan bu farkların kaynağının belirlenmesi için ikili karşılaştırmalar yapılmıştır (Ek 15). Buna göre çoklu puanlanan madde yüzdesi %5, %10 ve %50 olduğunda ÇBMTK ve ÜDMTK modelleri ile test güçlük düzeylerine göre elde edilen bütün RMSE değerleri anlamlı olarak birbirinden farklılaşmaktadır. RMSE değerleri test güçlük düzeylerinde en düşükten en yükseğe doğru orta güçlük, kolay, zor, çok kolay ve çok zor şeklinde sıralanmaktadır. ÇBMTK ve ÜDMTK modelleri ile çoklu puanlanan madde yüzdesi %25 olduğunda ise sıralama aynı olup tek fark orta güçlük ve kolay test güçlük düzeylerinde elde edilen RMSE değerleri arasındaki farkın anlamlı olmamasıdır. Bifaktör model için çoklu puanlanan madde yüzdelerinin hepsinde test güçlükleri arasındaki farka ait etki büyüklüğü düşüktür. İkili karşılaştırmalarda bazıları arasındaki farklar anlamlı olsa da ortalama RMSE değerleri birbirlerine oldukça yakındır.

Güvenirlik değerleri açısından ise çoklu puanlanan madde yüzdesi %10 ve %50 olduğunda ÇBMTK ve ÜDMTK modelleri ile test güçlük düzeylerine göre elde edilen bütün güvenirlik değerleri anlamlı olarak birbirinden farklılaşmaktadır. Güvenirlik değerleri test güçlük düzeylerinde en düşükten en yükseğe doğru orta, kolay, zor, çok kolay ve çok zor şeklinde sıralanmaktadır. ÇBMTK ve ÜDMTK modelleri ile çoklu puanlanan madde yüzdesi %5 ve

%25 olduğunda ise sıralama aynı olup tek fark orta güçlük ve kolay test güçlük düzeylerinde elde edilen güvenirlik değerleri arasındaki farkın anlamlı olmamasıdır. Bifaktör model için çoklu puanlanan madde yüzdelerinin hepsinde test güçlükleri arasındaki farka ait etki büyüklüğü diğer modellere göre daha düşüktür. İkili karşılaştırmalarda bazıları arasındaki farklar anlamlı olsa da ortalama güvenirlik değerleri birbirlerine oldukça yakındır.

Genel olarak, boyutlar arası korelasyon 0,8 olduğunda, bütün çoklu puanlanan madde yüzdeleri için hesaplanan RMSE değerlerinin test kolaylaştıkça ve zorlaştıkça arttığı söylenebilir. Benzer şekilde test güçlüğü arttıkça ve kolaylaştıkça elde edilen güvenirlikler artış göstermiştir. Fakat ortalama değerler diğer korelasyon düzeylerine göre birbirine daha yakındır.

Çoklu puanlanan madde yüzdesinin her bir düzeyinde (%5, %10, %25 ve %50) model ve test güçlüğü arasındaki etkileşim etkisine ilişkin test güçlüğü basit ana etkilerinin incelenmesinin ardından modellerin basit ana etkileri ele alınmıştır. Bir başka ifadeyle test güçlüğüne ait düzeylerin her birinde (çok zor, zor, orta güçlük, kolay, çok kolay) yetenek kestiriminde kullanılan modellere (ÇBMTK, ÜDMTK ve Bifaktör) ait RMSE ve güvenirlik değerleri karşılaştırılmıştır. Bu amaçla gerçekleştirilen tekrarlı ANOVA sonuçları Tablo 36'da verilmiştir.

Tablo 36.

Alt Puanlar için Korelasyon 0,8 Düzeyindeki Her Bir Çoklu Puanlanan Madde Yüzdesi ve Test Güçlüklerinde Modellerin Karşılaştırılması (RMSE, Güvenirlik)

		RMSE		Güvenirlik	
ÇPMY	TG	F	η_p^2	F	η_p^2
%5	Çok zor	1134813,451*	,999	313405,601*	,997
	Zor	1503110,090*	,999	332541,417*	,997
	Orta güçlük	1724107,912*	,999	342126,580*	,997
	Kolay	1702348,027*	,999	345189,099*	,997
	Çok kolay	1460787,805*	,999	336094,720*	,997
%10	Çok zor	1429530,911*	,999	376361,780*	,997
	Zor	1898827,773*	,999	396712,485*	,997
	Orta güçlük	2127665,602*	,999	409001,421*	,998
	Kolay	2052262,616*	,999	404181,101*	,998
	Çok kolay	1836358,879*	,999	402181,683*	,998
%25	Çok zor	687309,104*	,999	465676,536*	,998
	Zor	914244,120*	,999	494622,722*	,998
	Orta güçlük	1018663,161*	,999	498543,385*	,998
	Kolay	1010967,719*	,999	494198,787*	,998
	Çok kolay	839467,452*	,999	485710,493*	,998
%50	Çok zor	1374504,283*	,999	494438,856*	,998
	Zor	1622859,776*	,999	497116,157*	,998
	Orta güçlük	1805254,836*	,999	498882,997*	,998
	Kolay	1701219,008*	,999	495855,201*	,998
	Çok kolay	1479597,638*	,999	496989,091*	,998

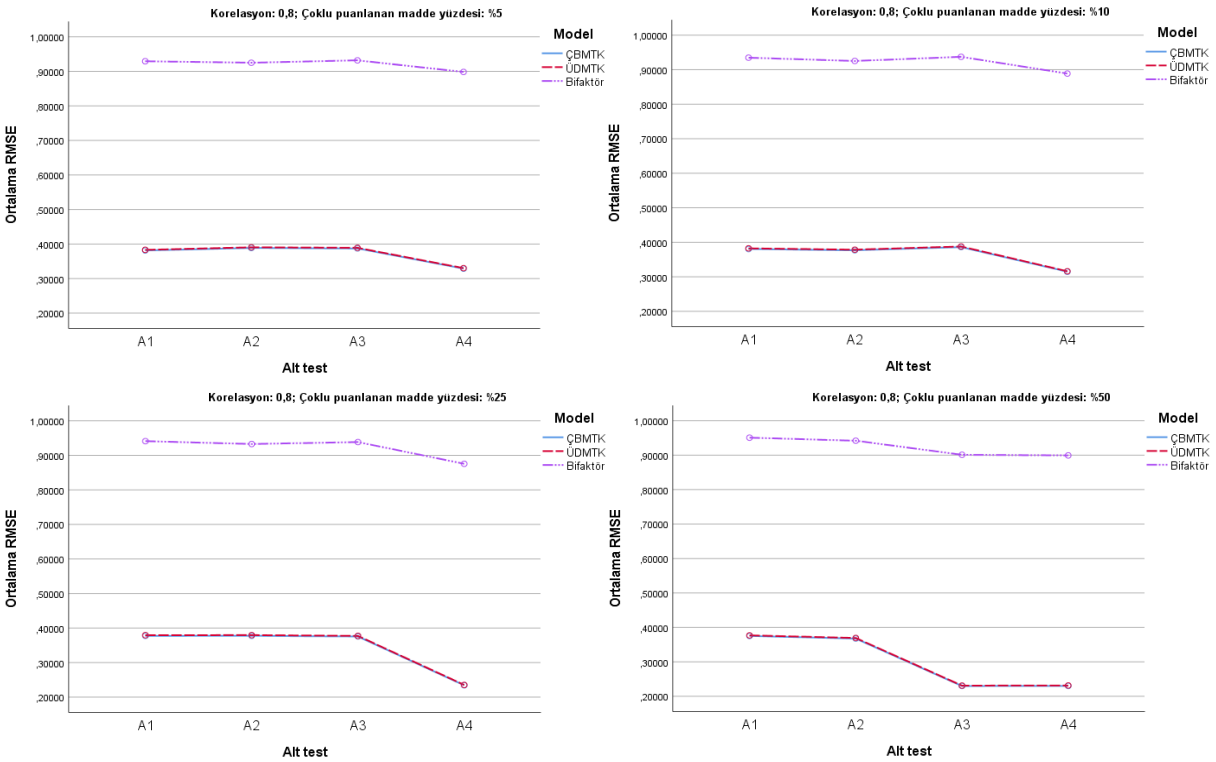
*p<,001; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 36'da çoklu puanlanan madde yüzdesi ve test güçlüğünün tüm düzeylerinde RMSE ve güvenirlik değerlerinin her birinin kestirim modellerine göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. Anlamlı bulunan bu farkların hangi modeller arasında olduğunun incelenmesi için ikili karşılaştırmalar yapılmıştır (Ek 15). Buna göre korelasyon 0,2 ve 0,5 düzeyleri ile benzer şekilde korelasyon 0,8 olduğunda da bütün çoklu puanlanan madde yüzdelerinde ve test güçlük düzeylerinde modeller arasındaki bütün ikili karşılaştırmalar anlamlıdır. RMSE değerleri en düşükten en yükseğe doğru sırasıyla ÇBMTK, ÜDMTK ve Bifaktör modelleri ile elde edilmiştir.

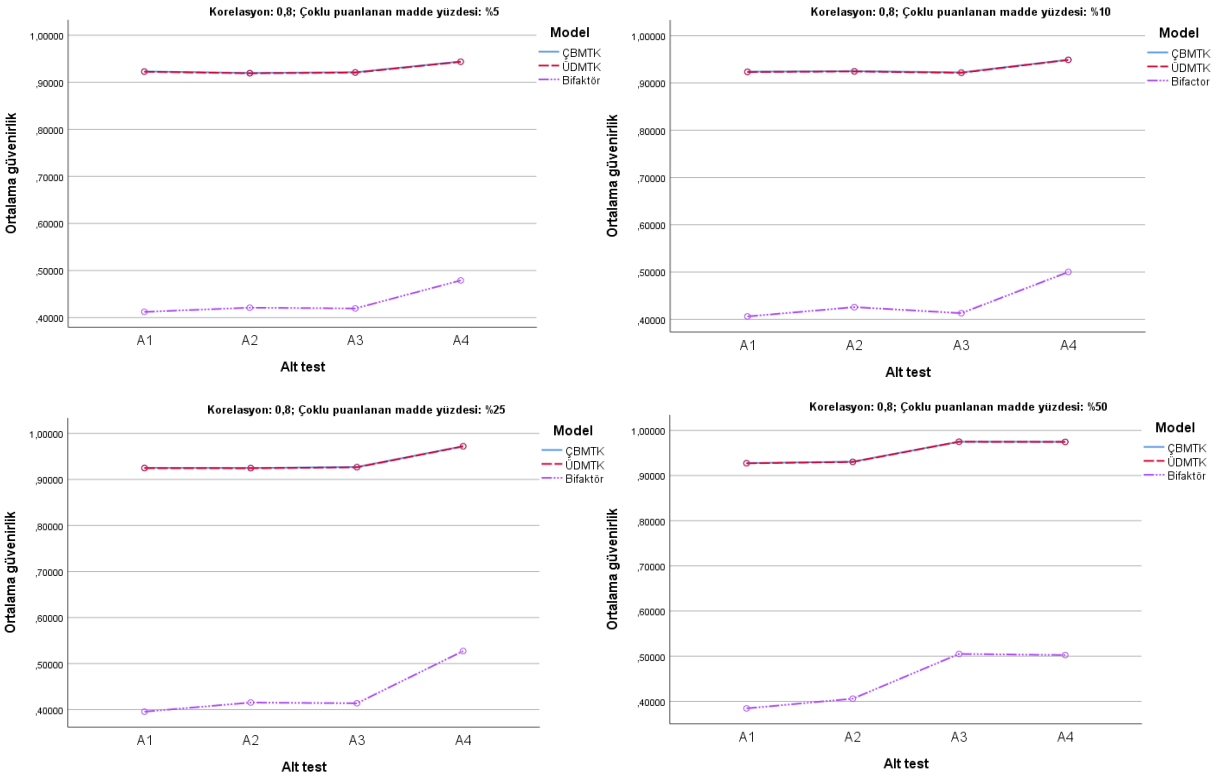
Güvenirlik değerleri açısından da bütün çoklu puanlanan madde yüzdelerinde ve test güçlük düzeylerinde modeller arasındaki neredeyse bütün ikili karşılaştırmalar anlamlıdır. Tek istisna çoklu puanlanan madde yüzdesi %10 ve orta test güçlüğü için ÇBMTK ve ÜDMTK ile elde edilen güvenirlikler arasındaki farkın anlamlı olmamasıdır. Güvenirlik değerleri en yüksekten en düşüğe doğru olan model sıralaması ÇBMTK, ÜDMTK ve Bifaktör şeklindedir. ÇBMTK ve ÜDMTK arasındaki farklar genel olarak anlamlı olsa da ortalama değerleri birbirlerine

oldukça yakındır. Hem güvenirlik hem de RMSE değerleri açısından ÇBMTK ve ÜDMTK modelleri arasındaki fark anlamlı olsa da ortalama değerleri birbirine oldukça yakındır.

Boyutlar arası korelasyon 0,8 düzeyi ve çoklu puanlanan madde yüzdesinin her bir düzeyi (%5, %10, %25 ve %50) için gerçekleştirilen analiz ile (Tablo 34) elde edilen model-test güçlüğü arasındaki anlamlı etkileşimin incelenmesinin ardından model-alt test arasındaki etkileşimin incelenmesine geçilmiştir. Tablo 34'te yer alan analiz sonuçları incelendiğinde bütün çoklu puanlanan madde yüzdesi düzeylerinde model-alt test arasındaki etkileşimin anlamlı ve büyük etkiye sahip olduğu tespit edilmiştir. Model-alt test arasındaki iki yönlü etkileşimlerin değerlendirilmesi için alt test basit ana etkisi tek yönlü ANOVA ile incelenmiştir. Bir başka ifadeyle alt testlerin birbiri arasındaki farkları görmek için yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde dört alt test için hesaplanan RMSE ve güvenirlik değerleri karşılaştırılmıştır. Alt test basit ana etkisi için elde edilen sonuçlar Tablo 37'de verilmiştir. Çoklu puanlanan madde yüzdesinin her bir düzeyi için model ve alt test düzeylerinde elde edilen ortalama RMSE ve güvenirlik değerlerini içeren grafikler sırasıyla Şekil 34 ve Şekil 35'te yer almaktadır.



Şekil 34. Alt puanlar korelasyon 0,8 - çoklu puanlanan madde yüzdesi x model x alt test (RMSE)



Şekil 35. Alt test korelasyon 0,8 - çoklu puanlanan madde yüzdesi x model x alt test (güvenirlik)

Tablo 37.

Alt Puanlar için Korelasyon 0,8 Düzeyinde ve Her Bir Çoklu Puanlanan Madde Yüzdesinde Modeller için Alt Testlerin Karşılaştırılması (RMSE, Güvenirlik)

		RMSE		Güvenirlik	
ÇPMY	Model	F	η_p^2	F	η_p^2
%5	ÇBMTK	3115,494*	,824	2205,822*	,769
	ÜDMTK	3092,547*	,823	2168,245*	,766
	Bifaktör	1645,964*	,713	1616,627*	,709
%10	ÇBMTK	4768,008*	,878	2917,733*	,815
	ÜDMTK	4681,171*	,876	2861,065*	,812
	Bifaktör	3028,718*	,820	3753,851*	,850
%25	ÇBMTK	20134,685*	,968	10328,138*	,940
	ÜDMTK	19636,895*	,967	10041,186*	,938
	Bifaktör	5517,607*	,893	8081,874*	,924
%50	ÇBMTK	27938,686*	,977	14191,214*	,955
	ÜDMTK	27082,071*	,976	13618,511*	,954
	Bifaktör	5707,339*	,896	8303,776*	,926

*p<,001; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 37'de çoklu puanlanan madde yüzdesi ve kestirim modellerinin tüm düzeylerinde RMSE ve güvenirlik değerlerinin her birinin alt testlere göre anlamlı farklılık gösterdiği ve bu farkın yüksek etki büyüklüğüne sahip olduğu görülmektedir. Anlamlı bulunan bu farkların kaynağının belirlenmesi için ikili karşılaştırmalar yapılmıştır (Ek 16). Buna göre korelasyon 0,8 iken çoklu

puanlanan madde yüzdesi %10 olduğunda ÇBMTK ve ÜDMTK modelleri ile alt testlere göre elde edilen RMSE değerleri en yüksekten en düşüğe doğru alt test 1, alt test 2, alt test 3 ve alt test 4 şeklinde sıralanmaktadır. Çoklu puanlanan madde yüzdesi %5 olduğunda alt test 2 ve alt test 3 arasındaki fark ve çoklu puanlanan madde yüzdesi %25 olduğunda ilk üç alt test arasındaki fark anlamlı değildir. Çoklu puanlanan madde yüzdesi %50 olduğunda alt test 3 ve alt test 4 arasındaki fark anlamlı değil ve diğer alt testlerden anlamlı olarak daha düşük RMSE değerlerine sahiptir. Bifaktör model için çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda en yüksek RMSE alt test 1 ile hesaplanırken alt test 2, alt test 3 ve alt test 4 sırasıyla bunu takip etmektedir. En düşük RMSE alt test 4 ile elde edilmiştir. Çoklu puanlanan madde yüzdesi %50 olduğunda ise alt test 3 ve alt test 4 ile elde edilen RMSE değerleri arasındaki fark anlamlı değil ve diğerlerinden anlamlı olarak daha düşüktür.

Güvenirlik değerleri açısından ise çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda ÇBMTK, ÜDMTK ve Bifaktör model ile alt testlere göre elde edilen güvenirlik değerleri ilk üç alt testte birbirine yakınken alt test 4 en yüksek güvenirlik değerine sahiptir. Çoklu puanlanan madde yüzdesi %50 olduğunda ise alt test 3 ve alt test 4 için elde edilen güvenirlik değerleri arasındaki fark anlamlı değildir ve diğer alt testlerden daha yüksek güvenirlik değerine sahiptirler.

Genel olarak bütün modeller ile elde edilen hem RMSE hem güvenirlik değerleri için çoklu puanlanan madde yüzdesi %5, %10 ve %25 olduğunda tamamı ikili puanlanan maddelerden oluşan ilk 3 test birbirine yakın sonuçlar üretirken alt test 4 daha düşük RMSE ve daha yüksek güvenirlik değerlerine sahiptir. Çoklu puanlanan madde yüzdesi %50 olduğunda tamamı çoklu puanlanan maddelerden oluşan alt test 3 ve alt test 4 için elde edilen değerler birbirine yakındır ve diğer alt testlerden daha düşük RMSE ve daha yüksek güvenirlik değerlerine sahiptir.

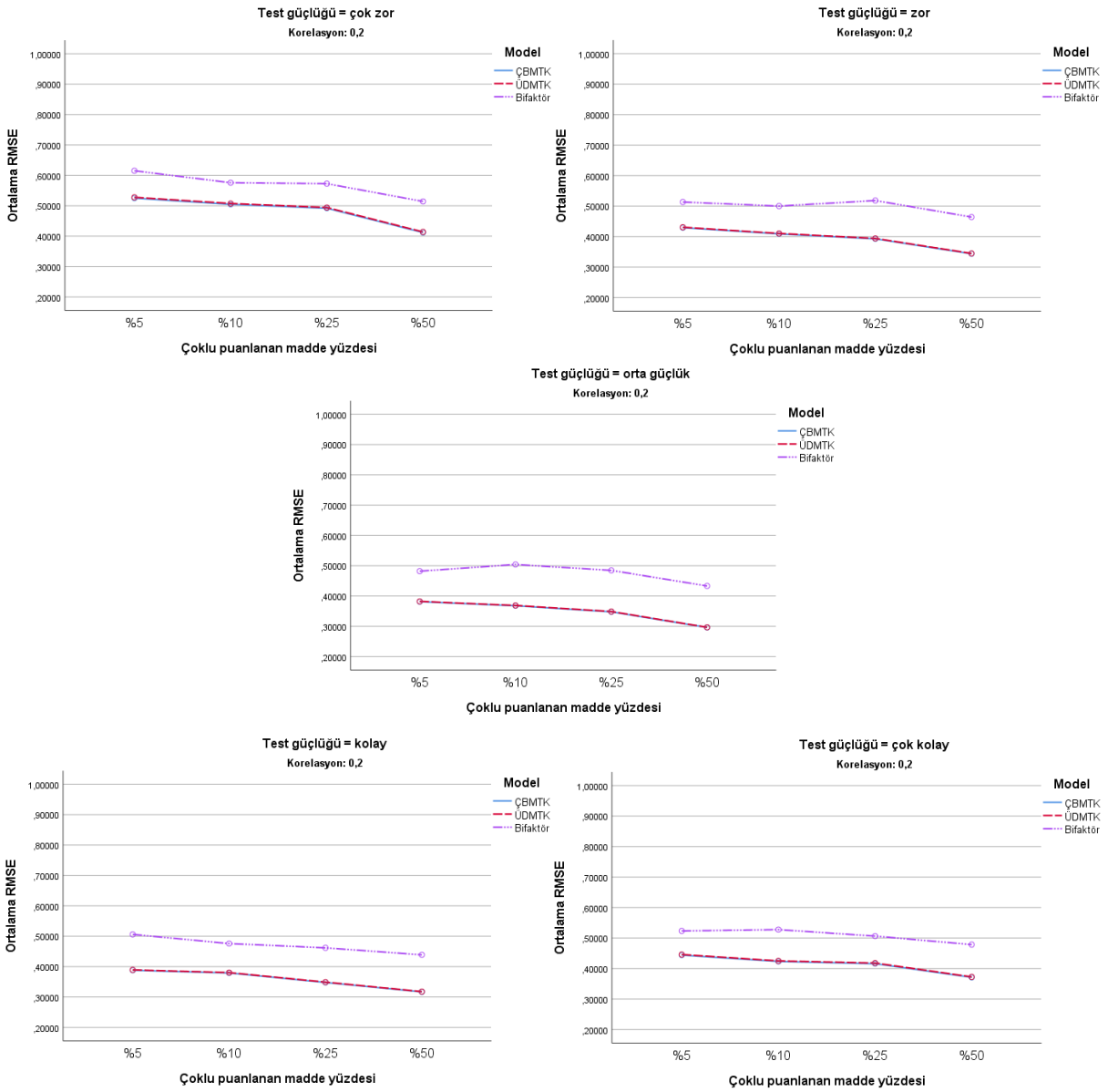
4.2.2. Çoklu Puanlanan Madde Yüzdesi Değişkeni için İncelemeler

Korelasyonun bütün düzeylerinde çoklu puanlanan madde yüzdesi, test güçlüğü ve model arasındaki etkileşim etkileri anlamlı bulunmuştur. Bunun için öncelikle çoklu puanlanan madde yüzdesinin her bir düzeyi için test güçlüğü ve model arasındaki etkileşim incelenmiştir. Test güçlüğü ve model basit ana etkileri yorumlanmıştır. Daha sonra ise yine bu üç yönlü etkileşimin devamı olarak test güçlüğünün her bir düzeyinde (çok zor, zor, orta güçlük, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki etkileşim etkisi incelenmiştir. Burada amaç çoklu puanlanan madde yüzdesi düzeyleri arasındaki farkları ortaya koyabilmektir. Bir

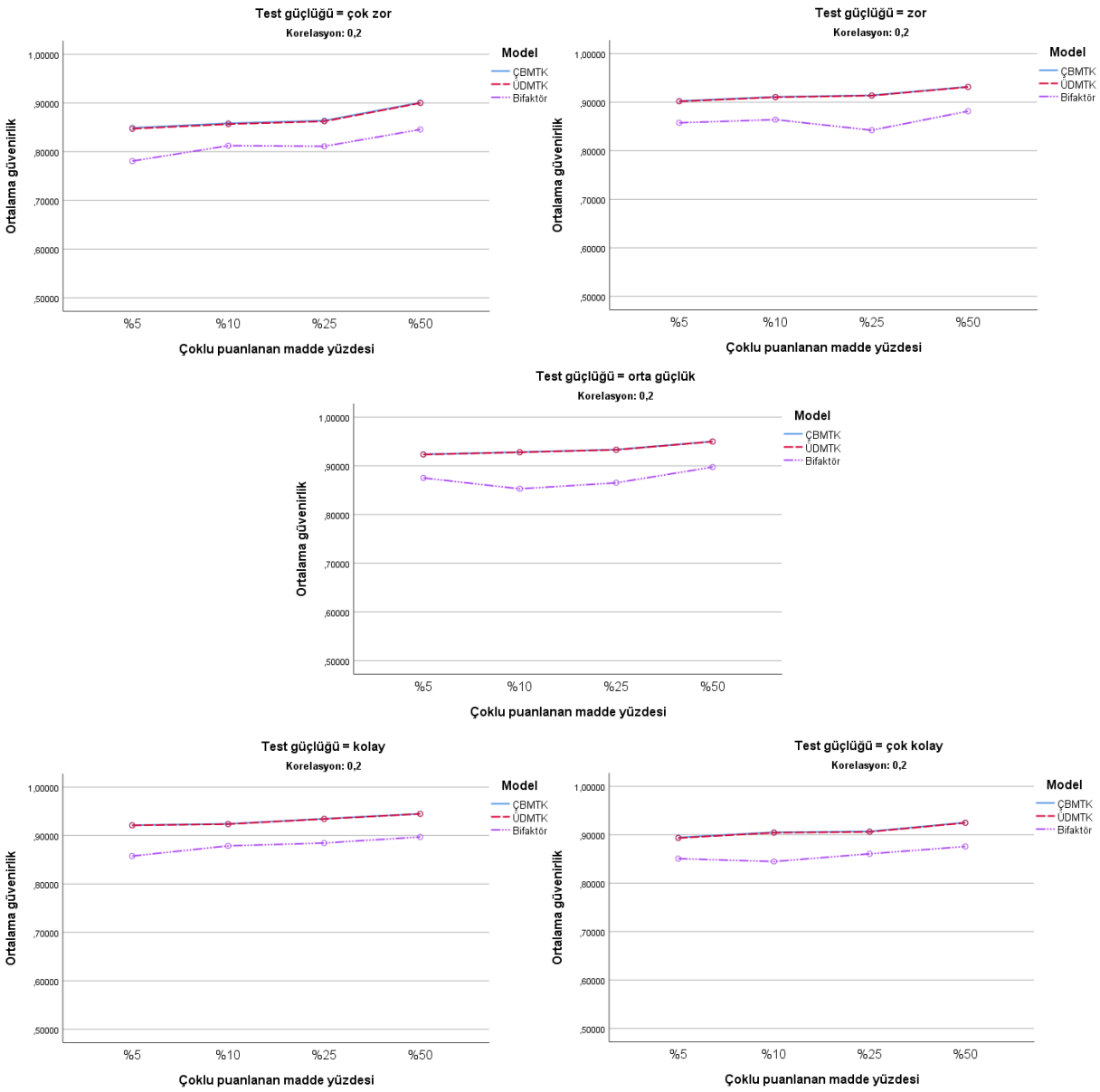
başka ifadeyle bu bölümde gerçekleştirilen analizlerde simülasyon veride toplam puan açısından çoklu puanlanan madde yüzdesinin ele alınan değerlendirme kriterleri açısından yetenek kestirimine olan etkisi ortaya konmuştur.

Korelasyon 0,2 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

Boyutlar arası korelasyon 0,2 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki etkileşim etkisi incelenmiştir. Bu amaçla model ve test güçlüğü değişkenleri ile RMSE ve güvenirlik değerlerinin karşılaştırılması için iki yönlü karma ANOVA yapılmış ve sonuçlar Tablo 38'de sunulmuştur. Test güçlüğünün her bir düzeyi için model ve çoklu puanlanan madde yüzdesi düzeylerinde elde edilen ortalama RMSE ve güvenirlik değerlerini içeren grafikler sırasıyla Şekil 36 ve Şekil 37'de yer almaktadır.



Şekil 36. Alt puanlar korelasyon 0,2 –test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE)



Şekil 37. Alt puanlar korelasyon 0,2 –test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik)

Tablo 38.

Alt Puanlar için Korelasyon 0,2 Olduğunda Test Güçlüğünün Her Bir Düzeyi için Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile İki Yönlü Karma ANOVA (RMSE, Güvenirlik)

			RMSE		Güvenirlik	
TG	Kaynak		F	η_p^2	F	η_p^2
Çok zor	Gruplar içi	M	13310,019*	,893	9206,723*	,852
		M *x* AT	943,001*	,640	1311,214*	,712
		M *x* ÇPMY	82,735*	,135	65,876*	,110
	Gruplar arası	AT	1827,065*	,775	1771,607*	,769
		ÇPMY	343,965*	,393	314,945*	,372
Zor	Gruplar içi	M	7790,669*	,830	4551,604*	,741
		M *x* AT	201,972*	,276	287,743*	,351
		M *x* ÇPMY	76,690*	,126	62,455*	,105
	Gruplar arası	AT	1131,299*	,681	1031,963*	,660
		ÇPMY	166,890*	,239	136,802*	,205
Orta güçlük	Gruplar içi	M	9756,879*	,860	6184,622*	,795
		M *x* AT	1059,426*	,666	1283,731*	,707
		M *x* ÇPMY	48,696*	,084	66,917*	,112
	Gruplar arası	AT	2425,920*	,820	2633,215*	,832
		ÇPMY	383,902*	,420	375,759*	,414
Kolay	Gruplar içi	M	8671,495*	,845	5711,360*	,782
		M *x* AT	220,467*	,293	307,770*	,367
		M *x* ÇPMY	21,291*	,039	36,303*	,064
	Gruplar arası	AT	1767,954*	,769	1429,845*	,729
		ÇPMY	399,598*	,429	364,957*	,407
Çok kolay	Gruplar içi	M	9365,987*	,855	6529,772*	,804
		M *x* AT	135,304*	,203	198,072*	,272
		M *x* ÇPMY	44,952*	,078	35,755*	,063
	Gruplar arası	AT	1803,773*	,773	1355,226*	,718
		ÇPMY	294,705*	,357	240,510*	,312

*p<,001; M: model; AT: alt test; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 38'de yer alan boyutlar arası korelasyon 0,2 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta güçlük, kolay, çok kolay) gerçekleştirilen analiz sonuçları incelendiğinde model ve çoklu puanlanan madde yüzdesi arasındaki etkileşimin RMSE ve güvenirlik değerlerinin her biri için anlamlı olduğu görülmektedir. Bu etkileşim etkilerine ait etki büyüklükleri düşük ya da orta düzeydedir. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi için kestirim modellerinin her biri için çoklu puanlanan madde yüzdesi basit ana etkisi tek yönlü ANOVA ile incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde çoklu puanlanan madde yüzdeleri (%5, %10, %25 ve %50) için hesaplanan RMSE ve güvenirlik değerleri karşılaştırılmıştır.

Çoklu puanlanan madde yüzdesi basit ana etkisi için elde edilen sonuçlar Ek 17'de verilmiştir. ANOVA çıktıları, bütün test güçlük düzeyleri altında bütün modeller için çoklu puanlanan

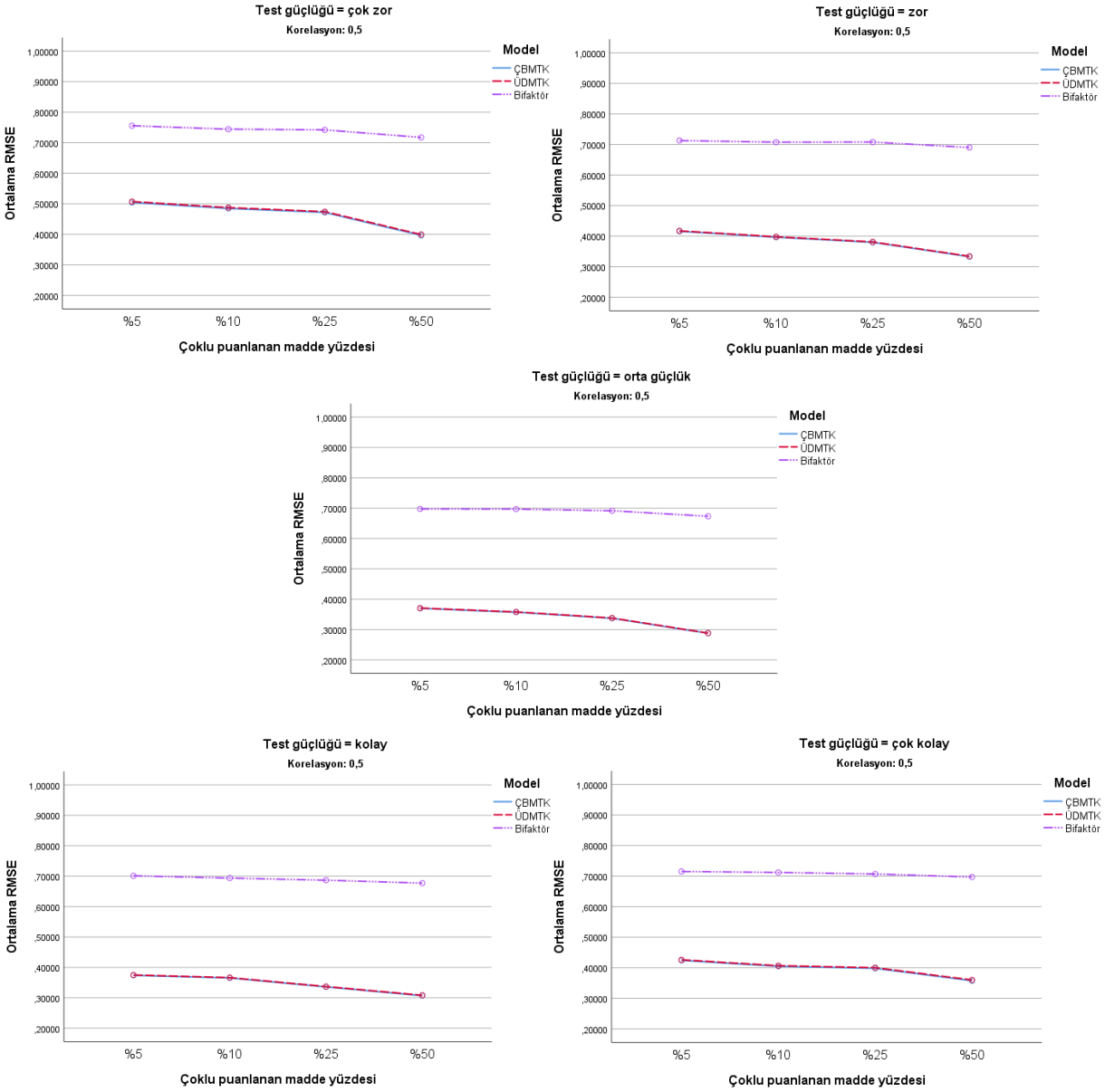
madde yüzdelerine göre elde edilen RMSE ve güvenirlik değerlerinin birbirinden anlamlı olarak farklılaştığını göstermektedir. Farkın kaynağına ilişkin yapılan ikili karşılaştırmalar ve ortalama RMSE ve güvenirlik değerleri de Ek 17'de sunulmaktadır. Buna göre test güçlük düzeyi zor ve orta güçlükte ÇBMTK ve ÜDMTK için elde edilen RMSE değerleri en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. Test güçlüğü kolay olduğunda sıralama aynı olup çoklu puanlanan madde yüzdesi %5 ve %10 arasında anlamlı farklılık yoktur. Test güçlüğü çok zor ve çok kolay olduğunda ise yine sıralama aynı fakat bu kez çoklu puanlanan madde yüzdesi %10 ve %25 arasındaki fark anlamlı değildir. Bifaktör model için test güçlüğü kolay olduğunda elde edilen RMSE değerleri en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. Test çok zor olduğunda %10 ve %25 arasındaki fark, zor olduğunda %5, %10 ve %25 arasındaki fark, orta güçlükte %5 ve %25 arasındaki fark, çok kolay olduğunda %5 ve %10 arasındaki fark anlamlı değildir. Diğer modeller ile benzer şekilde Bifaktör model için de en düşük RMSE, çoklu puanlanan madde yüzdesi %50 olduğunda elde edilmiştir.

Güvenirlik için yapılan analizlerde ise test güçlük düzeyi zor ve orta güçlükte ÇBMTK ve ÜDMTK için elde edilen güvenirlik değerleri en düşükten en yükseğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. Test güçlüğü kolay olduğunda sıralama aynı olup çoklu puanlanan madde yüzdesi %5 ve %10 arasında anlamlı farklılık yoktur. Diğer test güçlük düzeylerinde ise yine sıralama aynı fakat bu kez çoklu puanlanan madde yüzdesi %10 ve %25 arasındaki fark anlamlı değildir. Bifaktör model için test güçlüğü kolay, çok zor ve çok kolay olduğunda elde edilen güvenirlik değerleri en düşükten en yükseğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. Fakat test güçlüğü çok zor ve kolay olduğunda %25 ve %10 çoklu puanlanan madde yüzdeleri için hesaplanan güvenirlik değerleri arasındaki fark ve test güçlüğü çok kolay olduğunda %5 ve %10 çoklu puanlanan madde yüzdeleri arasındaki fark anlamlı değildir. Bifaktör model için çoklu puanlanan madde yüzdesi açısından bir örüntü yoktur. Fakat diğer modeller ile benzer şekilde Bifaktör model için de en düşük güvenirlik, çoklu puanlanan madde yüzdesi %50 olduğunda elde edilmiştir.

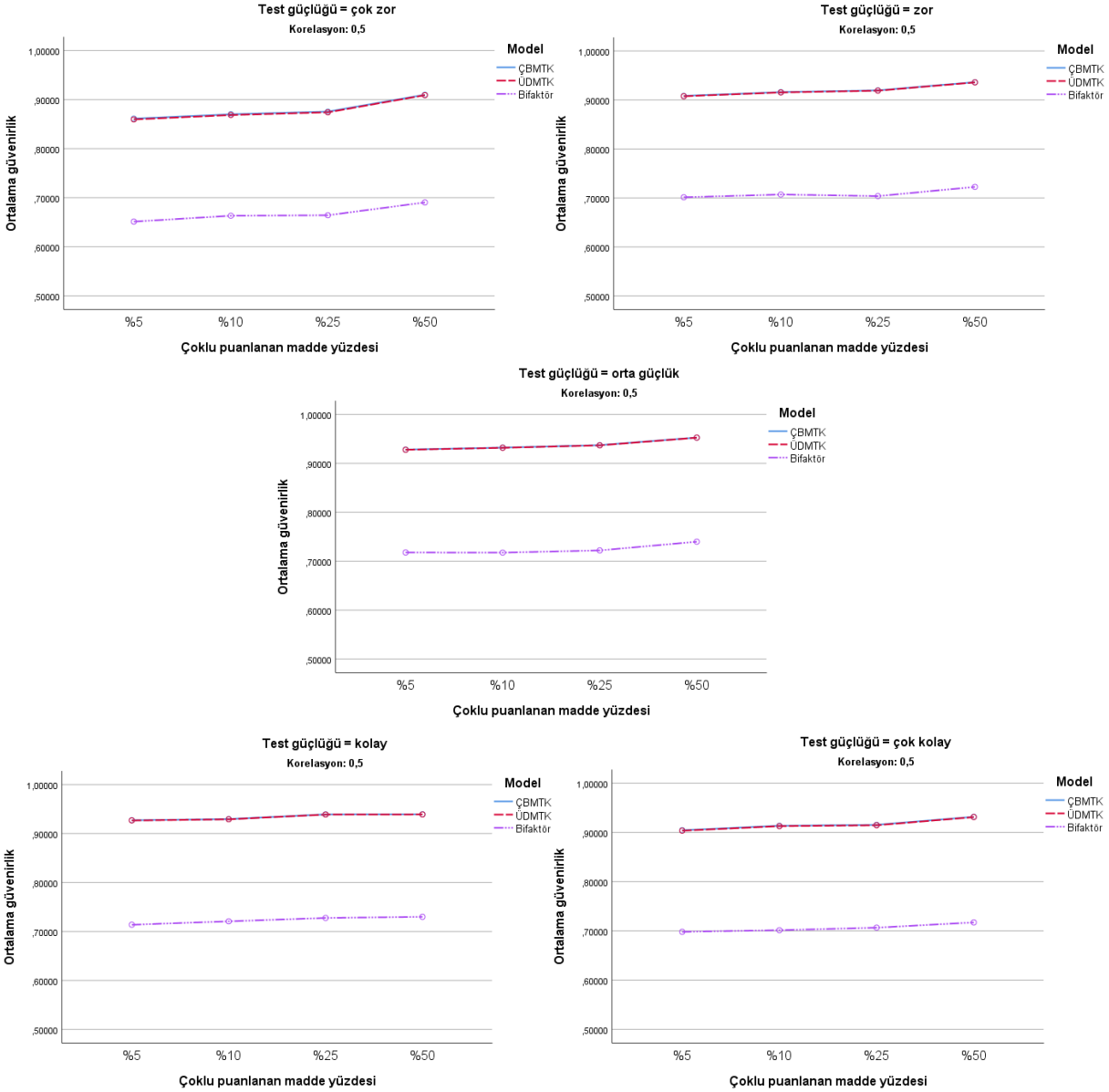
Özetle, bazı durumlarda benzerlikler olmasına rağmen genel olarak çoklu puanlanan madde yüzdesi arttıkça RMSE değerlerinin azaldığı ve güvenirlik değerlerinin arttığı söylenebilir.

Korelasyon 0,5 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

Boyutlar arası korelasyon 0,5 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki etkileşim etkisi incelenmiştir. Bu amaçla model ve test güçlüğü değişkenleri ile RMSE ve güvenirlik değerlerinin karşılaştırılması için iki yönlü karma ANOVA yapılmış ve sonuçlar Tablo 39'da sunulmuştur. Test güçlüğünün her bir düzeyi için model ve çoklu puanlanan madde yüzdesi düzeylerinde elde edilen ortalama RMSE ve güvenirlik değerlerini içeren grafikler sırasıyla Şekil 38 ve Şekil 39'da yer almaktadır.



Şekil 38. Alt puanlar korelasyon 0,5 – test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE)



Şekil 39. Alt puanlar korelasyon 0,5 –test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik)

Tablo 39.

Alt Puanlar için Korelasyon 0,5 Olduğunda Test Güçlüğünün Her Bir Düzeyi için Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Karma ANOVA (RMSE)

			RMSE		Güvenirlik	
TG	Kaynak		F	η_p^2	F	η_p^2
Çok zor	Gruplar içi	M	253102,909*	,994	535554,035*	,997
		M *x* AT	618,725*	,538	2583,027*	,829
		M *x* ÇPMY	818,697*	,607	92,291*	,148
	Gruplar arası	AT	1393,135*	,724	1594,056*	,750
		ÇPMY	312,434*	,370	254,140*	,324
Zor	Gruplar içi	M	247033,602*	,994	429196,505*	,996
		M *x* AT	544,210*	,506	802,963*	,602
		M *x* ÇPMY	401,086*	,430	41,577*	,073
	Gruplar arası	AT	886,823*	,625	1041,398*	,662
		ÇPMY	178,754*	,252	130,931*	,198
Orta güçlük	Gruplar içi	M	314094,137*	,995	725095,494*	,998
		M *x* AT	740,930*	,583	2628,889*	,832
		M *x* ÇPMY	398,460*	,429	18,891*	,034
	Gruplar arası	AT	1369,624*	,721	1883,382*	,780
		ÇPMY	316,493*	,373	281,760*	,347
Kolay	Gruplar içi	M	413531,692*	,996	373198,405*	,996
		M *x* AT	688,952*	,565	381,035*	,418
		M *x* ÇPMY	365,455*	,408	9,306*	,017
	Gruplar arası	AT	1251,227*	,702	94,249*	,151
		ÇPMY	324,076*	,379	8,501*	,016
Çok kolay	Gruplar içi	M	560944,219*	,997	790721,623*	,998
		M *x* AT	956,602*	,643	752,226*	,586
		M *x* ÇPMY	591,354*	,527	60,934*	,103
	Gruplar arası	AT	1666,027*	,758	1716,364*	,764
		ÇPMY	289,255*	,353	232,397*	,304

*p<,001; M: model; AT: alt test; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 39'da yer alan boyutlar arası korelasyon 0,5 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta güçlük, kolay, çok kolay) gerçekleştirilen analiz sonuçları incelendiğinde model ve çoklu puanlanan madde yüzdesi arasındaki etkileşimin RMSE ve güvenirlik değerlerinin her biri için anlamlı olduğu görülmektedir. RMSE için yapılan analizlerdeki etkileşim etkilerine ait etki büyüklükleri büyüktür. Ancak güvenirlik için elde edilen etkileşimler ise düşük ya da orta düzeydedir. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi amacıyla kestirim modellerinin her biri için çoklu puanlanan madde yüzdesi basit ana etkisi tek yönlü ANOVA ile incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde çoklu puanlanan madde yüzdeleri (%5, %10, %25 ve %50) için hesaplanan RMSE ve güvenirlik değerleri karşılaştırılmıştır.

Çoklu puanlanan madde yüzdesi basit ana etkisi için elde edilen sonuçlar Ek 18'de verilmiştir. ANOVA çıktıları, bütün test güçlük düzeyleri altında bütün modeller için çoklu puanlanan

madde yüzdelerine göre elde edilen RMSE değerlerinin birbirinden anlamlı olarak farklılaştığını göstermektedir. Farkın kaynağına ilişkin yapılan ikili karşılaştırmalar ve ortalama RMSE ve güvenirlik değerleri de Ek 18'de sunulmaktadır. Buna göre test güçlük düzeylerinin hepsinde ÇBMTK ve ÜDMTK için elde edilen RMSE değerleri en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. Sadece test güçlüğü çok kolay olduğunda %10 ve %25 çoklu puanlanan madde yüzdesi düzeylerindeki fark anlamlı değildir. Bifaktör model için test güçlüğü kolay olduğunda elde edilen RMSE değerleri en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. Test orta güçlükte ve çok kolay olduğunda %5 ve %10 arasındaki fark, zor ve çok zor olduğunda %5, %10 ve %25 arasındaki fark anlamlı değildir. Diğer modeller ile benzer şekilde Bifaktör model için de en düşük RMSE, çoklu puanlanan madde yüzdesi %50 olduğunda elde edilmiştir.

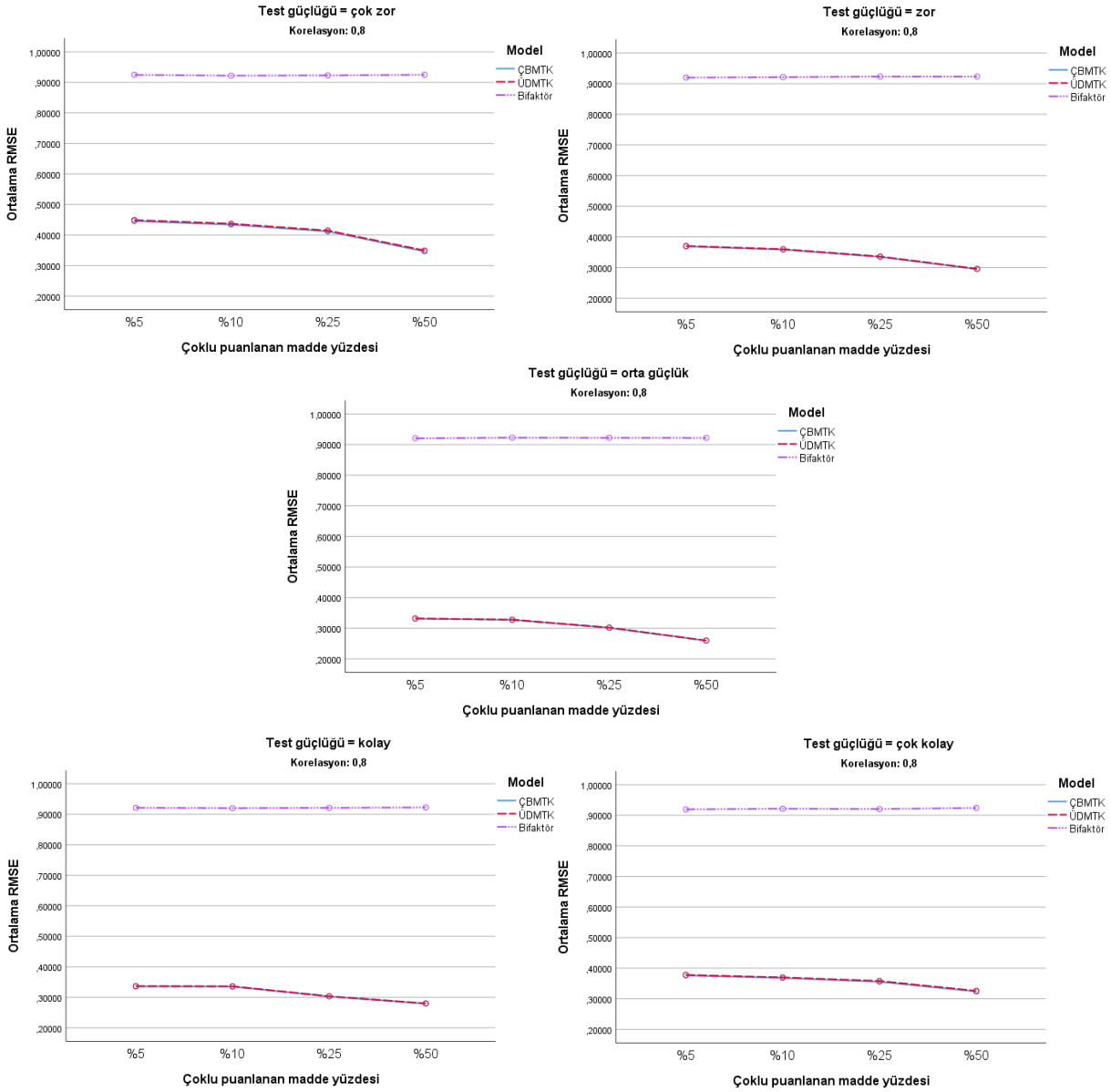
Güvenirlik değerleri için yapılan analizlerde test güçlüğü kolay olduğunda bütün modeller için farklı çoklu puanlanan madde yüzdelerinde hesaplanan güvenirlik değerleri arasında anlamlı farklılık olmadığı görülmüştür. Diğer test güçlük düzeylerinde ÇBMTK ve ÜDMTK için elde edilen güvenirlik değerleri en düşükten en yükseğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. Çoğu durumda %10 ve %25 çoklu puanlanan madde yüzdesi düzeylerindeki fark anlamlı değildir. Bifaktör model için test güçlüğü kolay olduğunda elde edilen güvenirlik değerleri en düşükten en yükseğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. Test çok kolay olduğunda %5 ve %10 arasındaki fark, zor ve orta güçlük olduğunda %5, %10 ve %25 arasındaki fark anlamlı değildir. Diğer modeller ile benzer şekilde Bifaktör model için de en düşük güvenirlik, çoklu puanlanan madde yüzdesi %50 olduğunda elde edilmiştir.

Özetle, boyutlar arası korelasyon 0,8 olduğunda bütün modeller için çoklu puanlanan madde yüzdesi arttıkça, elde edilen RMSE değerleri azalmış, güvenirlik değerleri ise artmıştır. Fakat Bifaktör model için çoklu puanlanan madde yüzdesine göre elde edilen ortalama RMSE değerlerinin birbirine yakın olduğu görülmüştür.

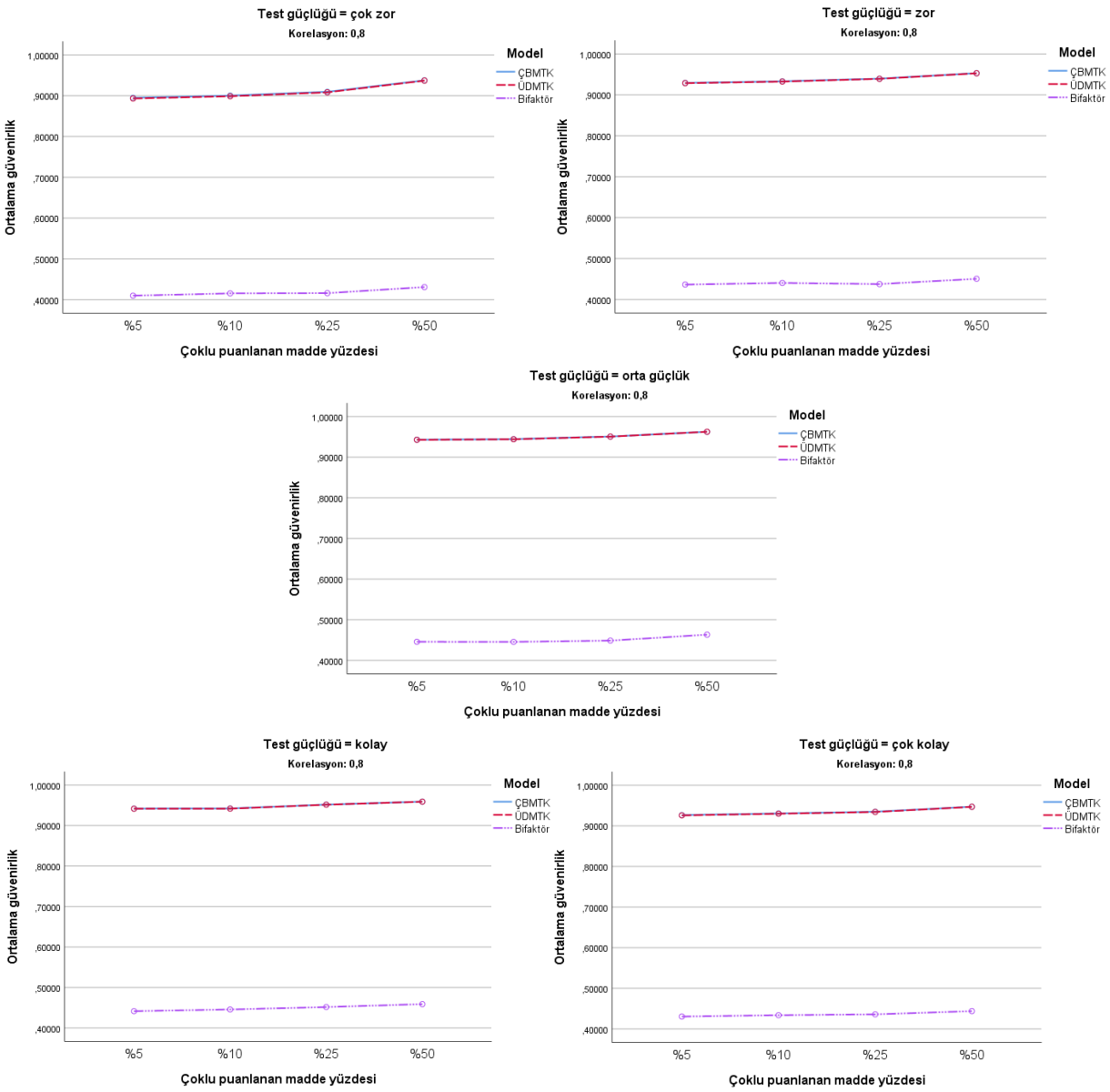
Korelasyon 0,8 Düzeyinde Üç Yönlü Etkileşimlerin İncelenmesi

Boyutlar arası korelasyon 0,8 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta, kolay ve çok kolay) model ve çoklu puanlanan madde yüzdesi arasındaki etkileşim etkisi

incelenmiştir. Bu amaçla model ve test güçlüğü değişkenleri ile RMSE ve güvenirlik değerlerinin karşılaştırılması için iki yönlü karma ANOVA yapılmış ve sonuçlar Tablo 40'da sunulmuştur. Test güçlüğünün her bir düzeyi için model ve çoklu puanlanan madde yüzdesi düzeylerinde elde edilen ortalama RMSE ve güvenirlik değerlerini içeren grafikler sırasıyla Şekil 40 ve Şekil 41'de yer almaktadır.



Şekil 40. Alt puanlar korelasyon 0,8 –test güçlüğü x model x çoklu puanlanan madde yüzdesi (RMSE)



Şekil 41. Alt puanlar korelasyon 0,8 –test güçlüğü x model x çoklu puanlanan madde yüzdesi (güvenirlik)

Tablo 40.

Alt Puanlar için Korelasyon 0,8 Olduğunda Test Güçlüğünün Her Bir Düzeyi için Model ve Çoklu Puanlanan Madde Yüzdesi Değişkenleri ile Karma ANOVA (RMSE, Güvenirlik)

			RMSE		Güvenirlik	
TG	Kaynak		F	η_p^2	F	η_p^2
Çok zor	Gruplar içi	M	1138209,895*	,999	919638,675*	,998
		M *x* AT	555,254*	,511	1930,399*	,784
		M *x* ÇPMY	2200,842*	,806	106,938*	,168
	Gruplar arası	AT	1378,978*	,722	1664,514*	,758
		ÇPMY	469,862*	,469	271,706*	,338
Zor	Gruplar içi	M	924150,358*	,998	1387204,802*	,999
		M *x* AT	581,742*	,523	1114,225*	,677
		M *x* ÇPMY	813,273*	,605	44,068*	,077
	Gruplar arası	AT	784,833*	,596	970,301*	,646
		ÇPMY	216,479*	,290	115,800*	,179
Orta güçlük	Gruplar içi	M	1078074,032*	,999	2673762,576*	,999
		M *x* AT	708,684*	,572	1870,719*	,779
		M *x* ÇPMY	790,044*	,598	11,475*	,021
	Gruplar arası	AT	1020,319*	,658	1426,487*	,729
		ÇPMY	322,475*	,378	206,604*	,280
Kolay	Gruplar içi	M	1477752,514*	,999	3459208,524*	1,000
		M *x* AT	733,682*	,580	1402,926*	,725
		M *x* ÇPMY	799,794*	,601	12,767*	,023
	Gruplar arası	AT	974,406*	,647	1246,698*	,701
		ÇPMY	317,886*	,374	226,146*	,299
Çok kolay	Gruplar içi	M	2035624,146*	,999	2829576,167*	,999
		M *x* AT	1043,730*	,663	1630,124*	,754
		M *x* ÇPMY	988,432*	,651	33,481*	,059
	Gruplar arası	AT	1426,858*	,729	1721,534*	,764
		ÇPMY	280,774*	,346	198,953*	,273

*p<,001; M: model; AT: alt test; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi

Tablo 40'ta yer alan boyutlar arası korelasyon 0,8 iken test güçlüğünün her bir düzeyinde (çok zor, zor, orta güçlük, kolay, çok kolay) gerçekleştirilen analiz sonuçları incelendiğinde model ve çoklu puanlanan madde yüzdesi arasındaki etkileşimin RMSE ve güvenirlik değerlerinin her biri için anlamlı olduğu görülmektedir. Bu etkileşim etkilerine ait etki büyüklükleri RMSE için büyüktür. Güvenirlik için elde edilen model-çoklu puanlanan madde yüzdesi arasındaki etkileşimler düşük ya da orta düzeydedir. Burada elde edilen iki yönlü etkileşimlerin değerlendirilmesi amacıyla kestirim modellerinin her biri için çoklu puanlanan madde yüzdesi basit ana etkisi tek yönlü ANOVA ile incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde çoklu puanlanan madde yüzdeleri (%5, %10, %25 ve %50) için hesaplanan RMSE ve güvenirlik değerleri karşılaştırılmıştır.

Çoklu puanlanan madde yüzdesi basit ana etkisi için elde edilen sonuçlar Ek 19'da verilmiştir. ANOVA çıktıları, bütün test güçlük düzeyleri altında ÇBMTK ve ÜDMTK için çoklu

puanlanan madde yüzdelerine göre elde edilen RMSE değerlerinin birbirinden anlamlı olarak farklılaştığını göstermektedir. Farkın kaynağına ilişkin yapılan ikili karşılaştırmalar ve ortalama RMSE değerleri de Ek 19'da sunulmaktadır. Buna göre test güçlük düzeylerinin hepsinde ÇBMTK ve ÜDMTK için elde edilen RMSE değerleri en yüksekten en düşüğe doğru çoklu puanlanan madde yüzdesi sırasıyla %5, %10, %25 ve %50 olduğunda elde edilmiştir. Sadece test güçlüğü orta ve zor olduğunda %5 ve %10 çoklu puanlanan madde yüzdesi düzeylerindeki fark anlamlı değildir. Bifaktör model için test güçlüğü çok zor, zor ve orta güçlükte olduğunda elde edilen RMSE değerleri çoklu puanlanan madde yüzdesi düzeyine göre anlamlı olarak farklılaşmamaktadır. Test kolay olduğunda sadece %5 ve %50 çoklu puanlanan madde yüzdesi düzeyleri arasındaki fark anlamlıdır. Test çok kolay olduğunda %50 çoklu puanlanan madde yüzdesi düzeyi anlamlı olarak diğerlerinden farklıdır, diğer düzeyler arasındaki fark manidar değildir. Genel olarak korelasyon 0,8 olduğunda bütün test güçlük düzeylerinde çoklu puanlanan madde yüzdesi düzeyine göre elde edilen RMSE değerleri arasındaki farklar oldukça azdır.

Güvenirlik için yapılan analizlerde ise ANOVA çıktıları, bütün test güçlük düzeyleri altında ÇBMTK ve ÜDMTK için çoklu puanlanan madde yüzdelerine göre elde edilen güvenirlik değerlerinin birbirinden anlamlı olarak farklılaştığını göstermektedir. Bifaktör model için elde edilen etki büyüklükleri küçük ya da orta düzeydedir. Farkın kaynağına ilişkin yapılan ikili karşılaştırmalar ve ortalama güvenirlik değerleri de Ek 19'da sunulmaktadır. Genel olarak korelasyon 0,8 olduğunda bütün test güçlük düzeylerinde çoklu puanlanan madde yüzdesi düzeyine göre elde edilen güvenirlik değerleri arasındaki farklar oldukça azdır.

Özetle, boyutlar arası korelasyon 0,8 olduğunda bütün modeller için çoklu puanlanan madde yüzdesi arttıkça, elde edilen RMSE değerleri azalmış, güvenirlik değerleri ise artmıştır. Fakat bu artık oldukça düşük düzeydedir. Özellikle Bifaktör model için çoklu puanlanan madde yüzdesine göre elde edilen ortalama RMSE değerleri birbirine çok yakındır.

4.2.3. Boyutlar Arası Korelasyon Değişkeni için İncelemeler

Bu bölümde gerçekleştirilen analizlerde simülasyon veride toplam puan açısından boyutlar arası korelasyonun ele alınan değerlendirme kriterleri açısından yetenek kestirimine olan etkisi ortaya konmuştur. Bu doğrultuda alt testler için elde edilen RMSE ve güvenirlik değerlerinin karşılaştırılması için gerçekleştirilen faktöriyel karma ANOVA (Tablo 23) sonucunda model,

çoklu puanlanan madde yüzdesi, test güçlüğü ve korelasyon değişkenleri arasında anlamlı çıkan dört yönlü etkileşimin yorumlanmasında bir inceleme daha yapılmıştır. Burada amaç korelasyon basit etkisinin incelenmesidir. Bu kez çoklu puanlanan madde yüzdesinin her bir düzeyinde model, test güçlüğü ve korelasyon arasındaki etkileşim etkisi incelenmiştir. Bu amaçla gerçekleştirilen faktöriyel karma ANOVA sonuçları Tablo 41'de sunulmuştur.

Tablo 41.

Alt Puanlar için Çoklu Puanlanan Madde Yüzdesinin Her Bir Düzeyinde Model, Test Güçlüğü ve Korelasyon Değişkenleri ile Faktöriyel Karma ANOVA (RMSE, Güvenirlik)

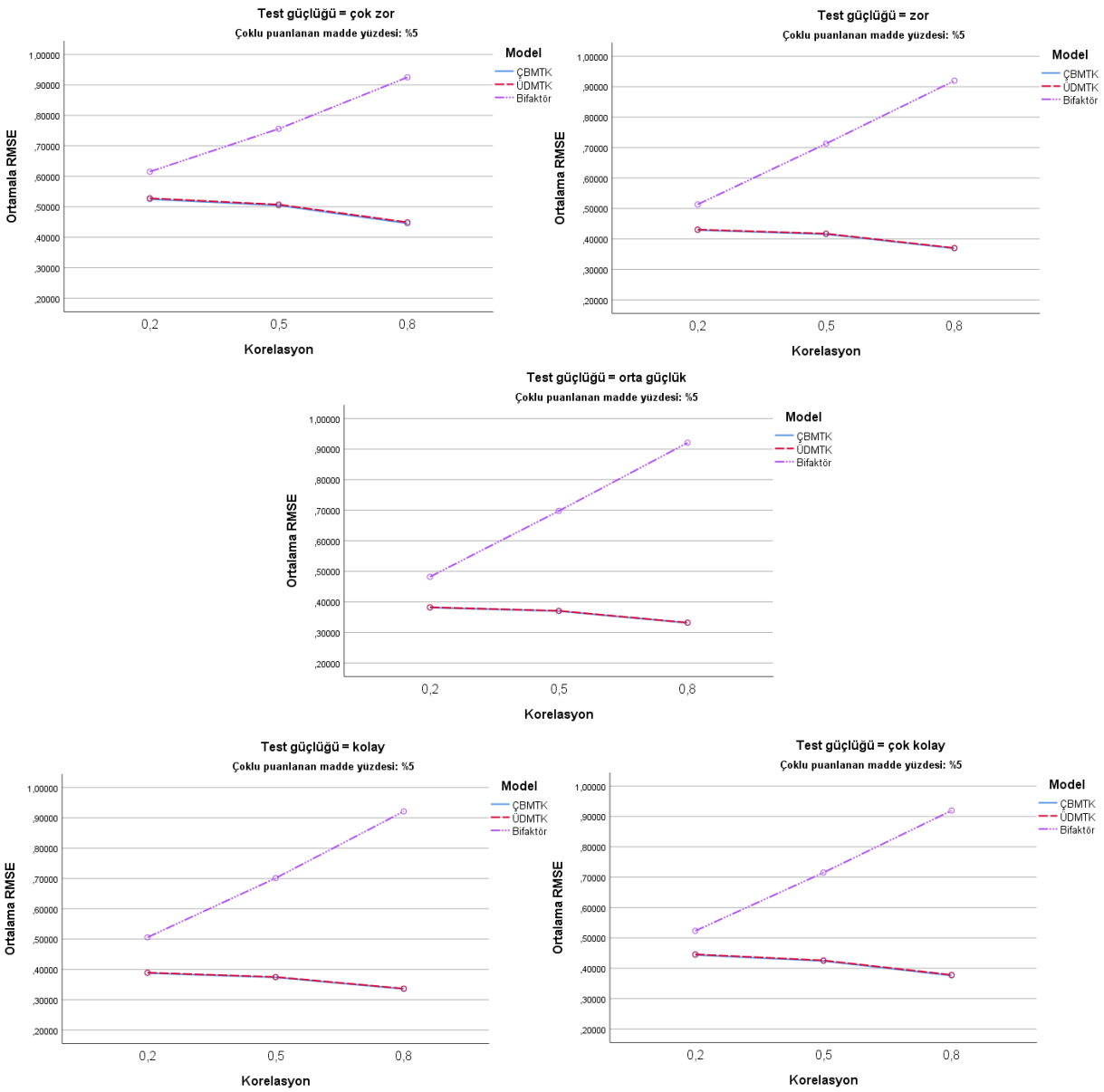
			RMSE		Güvenirlik	
ÇPMY	Kaynak		F	η_p^2	F	η_p^2
%5	Gruplar içi	M	540971,378*	,989	681566,314*	,991
		M *x* AT	760,079*	,276	1383,810*	,410
		M *x* K	95342,431*	,970	178403,584*	,984
		M *x* TG	923,424*	,382	44,958*	,029
		M *x* K *x* TG	140,075*	,158	39,418*	,050
%10	Gruplar içi	M	418786,818*	,986	619490,134*	,990
		M *x* AT	640,116*	,243	1778,021*	,471
		M *x* K	70792,802*	,959	160457,560*	,982
		M *x* TG	802,580*	,349	92,549*	,058
		M *x* K *x* TG	51,055*	,064	24,250*	,031
%25	Gruplar içi	M	432098,864*	,986	669585,438*	,991
		M *x* AT	1410,365*	,414	2759,314*	,581
		M *x* K	70711,203*	,959	170743,133*	,983
		M *x* TG	880,181*	,370	66,004*	,042
		M *x* K *x* TG	51,170*	,064	20,116*	,026
%50	Gruplar içi	M	2621385,824*	,998	1510813,824*	,996
		M *x* AT	9008,690*	,819	2631,844*	,569
		M *x* K	419861,650*	,993	402815,579*	,993
		M *x* TG	2370,417*	,613	36,752*	,024
		M *x* K *x* TG	143,870*	,161	3,585*	,005

*p<,001; M: model; AT: alt test; TG: test güçlüğü; ÇPMY: çoklu puanlanan madde yüzdesi; K: korelasyon

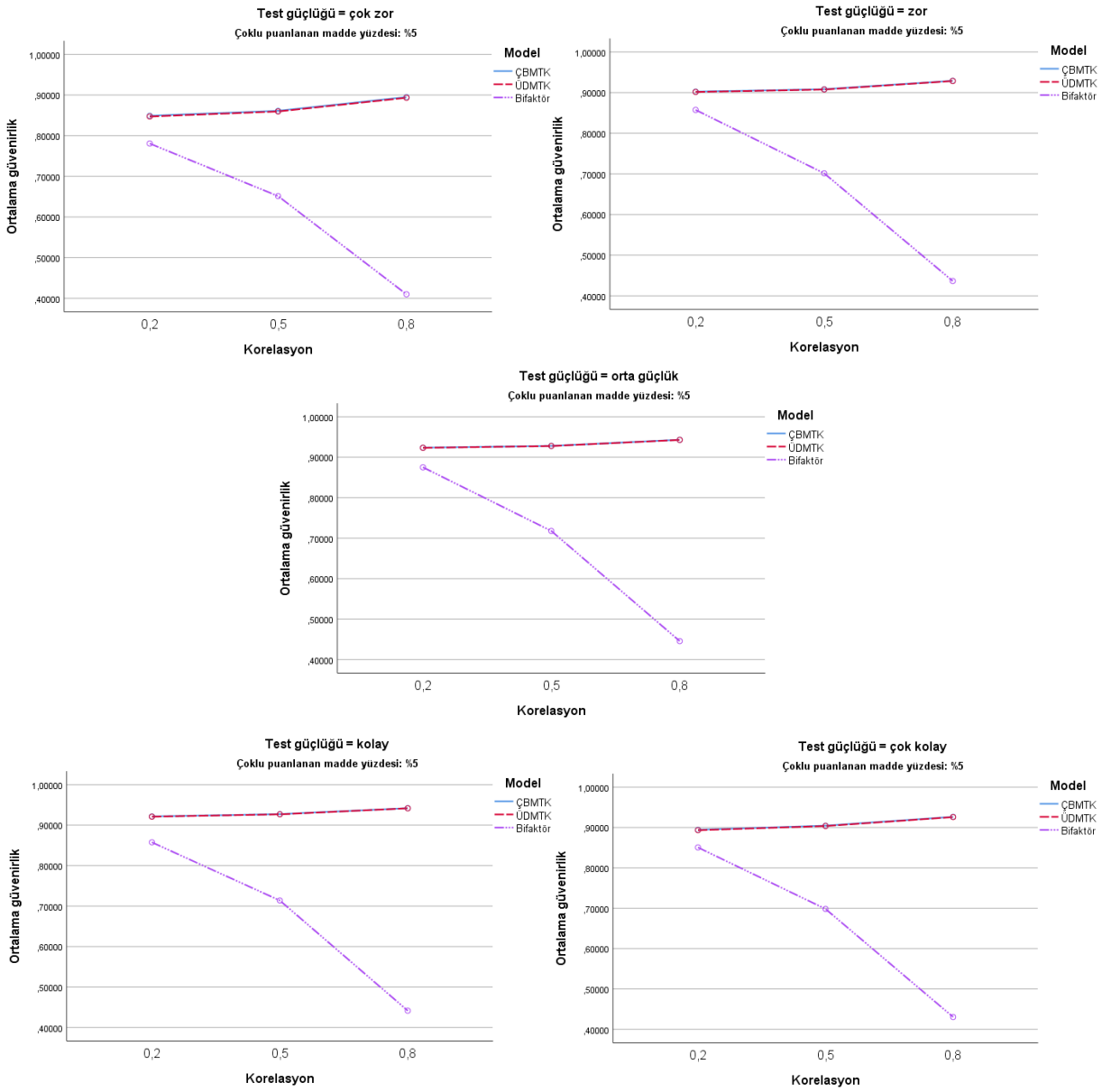
Tablo 41 incelendiğinde RMSE için çoklu puanlanan madde yüzdesinin bütün düzeylerinde model, korelasyon ve test güçlüğü arasındaki üç yönlü etkileşimlerin anlamlı olduğu görülmektedir. Güvenirlik için ise çoklu puanlanan madde yüzdesi %5, %10 ve %25 düzeylerinde model, korelasyon ve test güçlüğü arasındaki üç yönlü etkileşimler anlamlıdır. Çoklu puanlanan madde yüzdesi %50 için ise üç yönlü etkileşim yaklaşımı istatistiksel olarak anlamlı olsa dahi etki büyüklüğü anlamlı değildir. Dolayısıyla bu düzey için güvenirlik değerleri ile yapılan analizlerde model ve korelasyon arasındaki iki yönlü etkileşim incelenmiştir.

Anlamlı bulunan üç yönlü etkileşimleri incelemek amacıyla basit iki yönlü etkileşim yaklaşımı kullanılmıştır. RMSE için çoklu puanlanan madde yüzdesinin her bir düzeyi (%5, %10, %25

ve %50) ve güvenirlik için ilk üç düzey içerisinde test güçlüğünün bütün düzeyleri için model ve korelasyon arasındaki etkileşim etkisi incelenmiştir. Bu amaçla karma ANOVA yapılmıştır. Analiz sonuçları Ek 20'de yer almaktadır. Buna göre çoklu puanlanan madde yüzdesinin ve test güçlüğünün her bir düzeyinde model ve korelasyon arasındaki iki yönlü etkileşim anlamlı bulunmuştur. Burada elde edilen iki yönlü etkileşim etkisinin değerlendirilmesi için korelasyon basit ana etkisi incelenmiştir. Bir başka ifadeyle yetenek kestirim modellerinin (ÇBMTK, ÜDMTK ve Bifaktör) her birinde boyutlar arası korelasyon düzeyleri (0,2, 0,5 ve 0,8) için hesaplanan RMSE değerleri karşılaştırılmıştır. Korelasyon basit ana etkisi için elde edilen sonuçlar Ek 20'de verilmiştir. Çoklu puanlanan madde yüzdesi %5 için her bir test güçlüğü düzeyinde model ve korelasyon değişkenlerine ait ortalama RMSE ve güvenirlik değerlerini gösteren grafikler sırasıyla Şekil 42 ve Şekil 43'te yer almaktadır. Çoklu puanlanan madde yüzdesinin diğer düzeyleri için grafikler Ek 20'de sunulmaktadır.



Şekil 42. Alt puanlar çoklu puanlanan madde yüzdesi %5 –test güçlüğü x model x korelasyon (RMSE)



Şekil 43. Alt puanlar çoklu puanlanan madde yüzdesi %5 –test güçlüğü x model x korelasyon (güvenirlik)

Çoklu puanlanan madde yüzdesi ve test güçlüğü değişkenlerinin düzeylerinde anlamlı bulunan model ve korelasyon arasındaki iki yönlü etkileşimin incelendiği korelasyon basit etki testi sonuçları bütün modeller için korelasyon düzeyleri arasındaki farkın anlamlı olduğunu göstermektedir. Anlamlı bulunan bu farkların kaynağına ilişkin yapılan ikili karşılaştırma sonuçları Ek 20'de yer almaktadır. Buna göre genel olarak sonuçlar paralellik göstermektedir ve bütün karşılaştırmalar anlamlıdır. Bütün çoklu puanlanan madde yüzdelerinde ve bütün test güçlük düzeylerinde ÇBMTK ve ÜDMTK ile yapılan analizlerde korelasyon arttıkça RMSE değerleri düşüş göstermiştir. Bifaktör model için ise alt testler ile yapılan analizlerde toplam

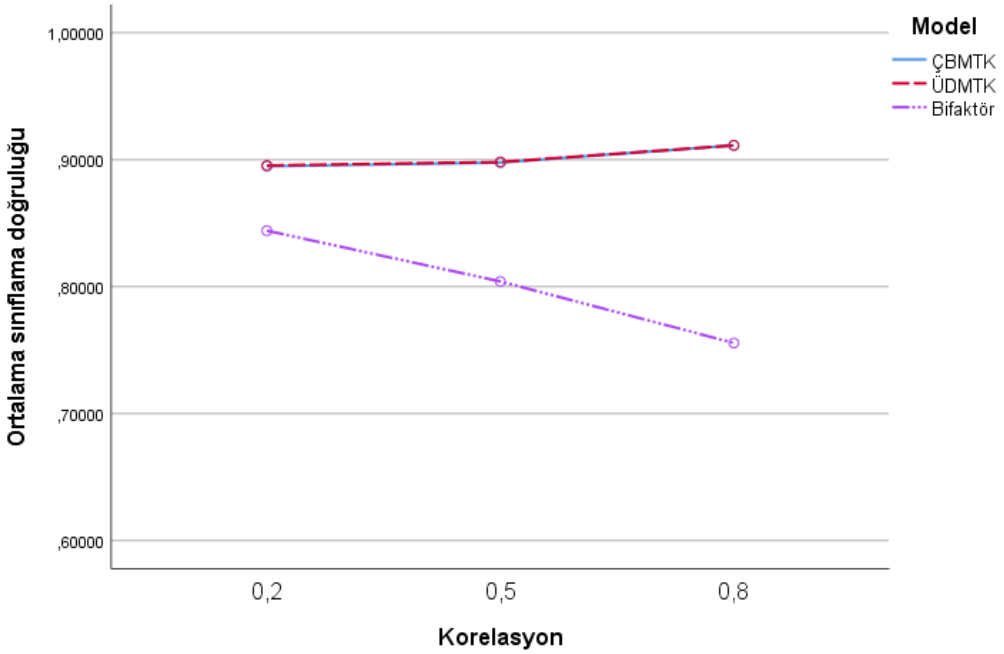
testten farklı olarak korelasyon arttıkça RMSE değerleri artmıştır. Korelasyon düzeylerinde elde edilen RMSE değerleri arasındaki ortalama farklar test çok kolay ya da çok zor olduğunda, test orta güçlükte olduğu duruma göre daha fazladır.

Güvenirlik değerleri ile yapılan analizlerde ise %5, %10 ve %25 çoklu puanlanan madde yüzdelerinde ve bütün test güçlük düzeylerinde ÇBMTK ve ÜDMTK ile yapılan analizlerde korelasyon arttıkça güvenirlik değerleri artış göstermiştir. Bifaktör model için ise alt testler ile yapılan analizlerde toplam testten farklı olarak korelasyon arttıkça güvenirlik değerleri düşmüştür. Çoklu puanlanan madde yüzdesi %50 için ise model ve korelasyon iki yönlü etkileşiminin incelenmesi için test güçlüğünden bağımsız olarak korelasyon basit etkisi incelenmiştir. Sonuçlar diğer çoklu puanlanan madde yüzdesi düzeyleri ile aynıdır.

4.2.4. Alt Test Sınıflama Doğruluğu Kriteri için İncelemeler

Tablo 23'te yer alan ve sınıflama doğruluğu değerleri ile gerçekleştirilen analiz sonucunda üç ve dört yönlü etkileşimlere dair istatistik değeri anlamlı olsa da kısmi eta kare değerlerinin 0,01'in altında olduğu, bu etkileşime dair etkinin oldukça düşük olduğu görülmüştür. Bu sebeple iki yönlü etkileşim etkileri incelenmiştir. Anlamlı bulunan iki yönlü etkileşim etkileri için değişkenler sırayla sabit tutularak diğer değişkenin düzeyleri arasındaki farklar incelenmiştir.

Sınıflama doğruluğu için model ve korelasyon arasındaki iki yönlü etkileşim etkisi anlamlı olup büyük etki büyüklüğü ile diğer iki yönlü etkileşimlerden daha büyüktür. Bu etkileşim etkisi için grafiksel gösterim Şekil 44'te yer almaktadır. Analizlerde kullanılan model ve boyutlar arası korelasyon değişkenleri için basit ana etkiler incelenmiştir. Bunun için gerçekleştirilen analiz sonuçları Tablo 42'de sunulmuştur.



Şekil 44. Alt puanlar –model x korelasyon (sınıflama doğruluğu)

Tablo 42.

Alt Puanlar Model x Korelasyon Etkileşim Etkisi için Basit Ana Etkiler (Sınıflama Doğruluğu)

		F	η_p^2
Korelasyon basit ana etkisi	Model		
	ÇBMTK	2544,773*	,175
	ÜDMTK	2406,894*	,167
	Bifaktör	31925,215*	,727
Model basit ana etkisi	Korelasyon		
	0,2	29673,996*	,713
	0,5	84090,329*	,875
	0,8	224273,431*	,949

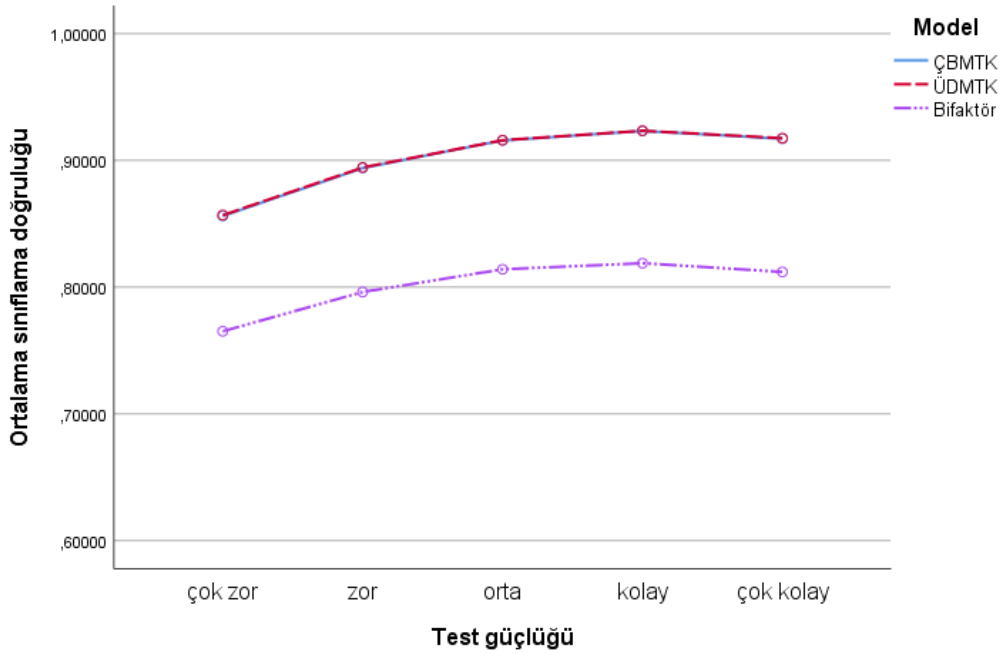
*p<,001

Tablo 42'de görüldüğü gibi boyutlar arası korelasyon basit ana etkisi için her bir model özelinde (ÇBMTK, ÜDMTK ve Bifaktör) korelasyon düzeylerinde hesaplanan sınıflama doğruluğu değerlerinin karşılaştırılması sonucunda farkın istatistiksel olarak anlamlı olduğu tespit edilmiştir. Her bir model için yapılan analizlere ait etki büyüklükleri incelendiğinde Bifaktör model için hesaplanan etki büyüklüğünün diğerlerinden daha yüksek olduğu, bu model için boyutlar arasındaki korelasyon düzeyinin farklılaşmasının sınıflama doğruluğu değerleri üzerindeki etkisinin daha önemli olduğu görülmektedir. Bütün modeller için farkın hangi korelasyon düzeyleri arasında olduğunun belirlenmesi amacıyla yapılan ikili karşılaştırma sonuçları Ek 21'de sunulmuştur. Buna göre bütün ikili karşılaştırmalar anlamlıdır. ÇBMTK ve ÜDMTK yöntemleri için korelasyon düzeyi arttıkça alt puanlar için hesaplanan sınıflama

doğruluğu değerleri artarken Bifaktör model için korelasyon arttıkça alt puanlara ait sınıflama doğruluğu değerleri daha düşük hesaplanmıştır.

Model *x* korelasyon etkileşimi model basit ana etkisi için her bir korelasyon düzeyinde (0,2, 0,5 ve 0,8) ÇBMTK, ÜDMTK ve Bifaktör modellerine ait sınıflama doğruluğu değerleri karşılaştırılmış ve modellere göre bu farklar anlamlı bulunmuştur. Korelasyon düzeylerinin her biri için anlamlı bulunan bu farkların hangi modeller arasında olduğunun incelenmesi için ikili karşılaştırmalar yapılmıştır (Ek 21). ÇBMTK ve ÜDMTK arasındaki fark anlamlı olsa da ortalama sınıflama doğruluğu değerleri arasındaki fark oldukça azdır. Bifaktör model bütün korelasyon düzeylerinde en düşük sınıflama doğruluğu değerlerine sahiptir. Özellikle 0,8 korelasyon düzeyinde Bifaktör model ile diğer modeller arasındaki fark fazladır.

Model ve test güçlüğü değişkenleri arasındaki iki yönlü etkileşim etkisinin de anlamlı olduğu görülmektedir. Bu etkileşim etkisi için grafiksel gösterim Şekil 45'te yer almaktadır. Model ve test güçlüğü değişkenleri için basit ana etkiler incelenmiştir. Bunun için gerçekleştirilen analiz sonuçları Tablo 43'de sunulmuştur.



*Şekil 45.*Alt puanlar - model x test güçlüğü (sınıflama doğruluğu)

Tablo 43.

Alt Puanlar Model x Test Güçlüğü Etkileşim Etkisi için Basit Ana Etkiler (Sınıflama Doğruluğu)

		F	η_p^2
Test güçlüğü ana etkisi	Model		
	ÇBMTK	14981,704*	,715
	ÜDMTK	15013,342*	,715
	Bifaktör	4681,275*	,439
Model ana etkisi	Test güçlüğü		
	Çok zor	50431,886*	,808
	Zor	54725,629*	,821
	Orta güçlük	57681,954*	,828
	Kolay	60830,392*	,836
	Çok kolay	62650,937*	,840

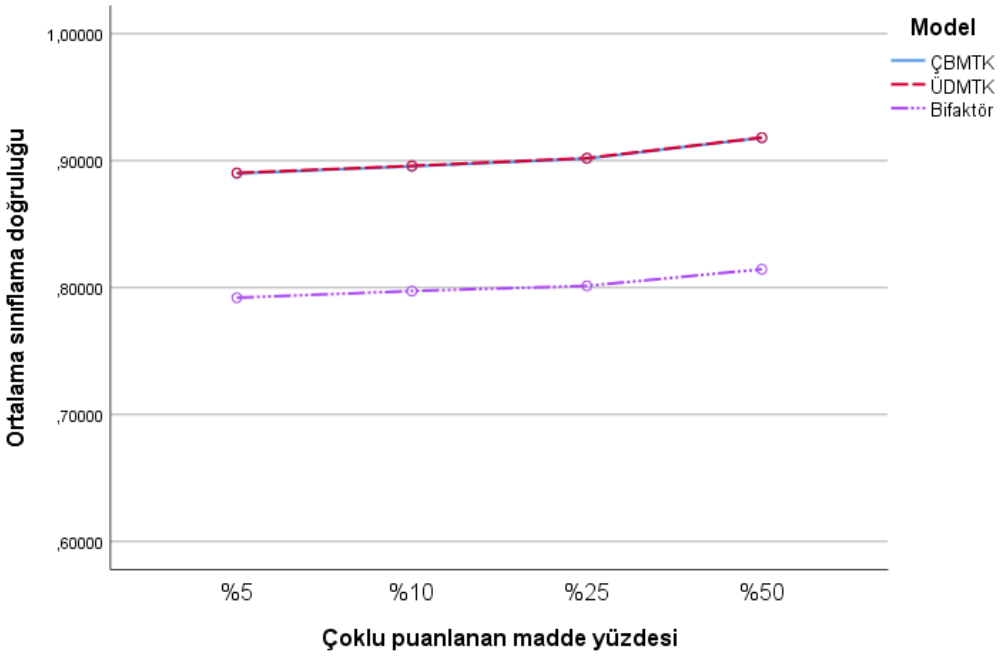
*p<,001

Tablo 43 incelendiğinde, test güçlüğü ana etkisi için her bir model özelinde (ÇBMTK, ÜDMTK ve Bifaktör) test güçlük düzeylerine ait sınıflama doğruluğu değerleri arasındaki farkın anlamlı olduğu görülmektedir. Her bir model için yapılan analizlere ait etki büyüklükleri incelendiğinde Bifaktör model için hesaplanan etki büyüklüğünün diğerlerinden daha düşük olduğu, bu model için test güçlük düzeyinin farklılaşmasının sınıflama doğruluğu değerleri üzerindeki etkisinin diğer modellere kıyasla daha düşük olduğu görülmektedir. Bütün modeller için farkın hangi test güçlükleri arasında olduğunun belirlenmesi amacıyla yapılan ikili karşılaştırma sonuçları Ek 10'da sunulmuştur. Buna göre kestirim modellerinin hepsi için test güçlükleri arasındaki bütün ikili karşılaştırmalar anlamlıdır. ÇBMTK ve ÜDMTK modellerinde test güçlük düzeyleri için hesaplanan sınıflama doğruluğu değerleri en düşükten en yükseğe doğru sırasıyla test kolay, çok kolay, orta, zor ve çok zor olduğunda elde edilmiştir. Bifaktör modeli için bu sıralama kolay, orta, çok kolay, zor ve çok zor şeklindedir.

Model *x* test güçlüğü etkileşimi model ana etkisi için her bir test güçlüğü düzeyinde (çok zor, zor, orta güçlük, kolay ve çok kolay) ÇBMTK, ÜDMTK ve Bifaktör modelleri karşılaştırılmış ve modeller arasındaki fark anlamlı bulunmuştur. Test güçlüğü düzeylerinin her biri için anlamlı bulunan bu farkların hangi modeller arasında olduğunun incelenmesi için ikili karşılaştırmalar yapılmıştır (Ek 22). ÇBMTK ve ÜDMTK arasındaki fark anlamlı olsa da ortalama sınıflama doğruluğu değerleri arasındaki fark oldukça azdır. Bifaktör model test güçlüğünün bütün düzeylerinde en yüksek sınıflama doğruluğu değerlerine sahiptir.

Alt puanlar için yapılan faktöriyel karma ANOVA sonuçlarına (Tablo 23) göre anlamlı çıkan bir diğer iki yönlü etkileşim etkisi model ve çoklu puanlanan madde yüzdesi değişkenleri arasındadır. Bu etkileşim etkisi için grafiksel gösterim Şekil 46'da yer almaktadır. Model ve

çoklu puanlanan madde yüzdesi değişkenleri için basit ana etkiler incelenmiştir. Bunun için gerçekleştirilen analiz sonuçları Tablo 45'te sunulmuştur.



Şekil 46. Alt puanlar –model x çoklu puanlanan madde yüzdesi (sınıflama doğruluğu)

Tablo 44.

Alt Puanlar Model x Çoklu Puanlanan Madde Yüzdesi Etkileşim Etkisi için Basit Ana Etkiler (Sınıflama Doğruluğu)

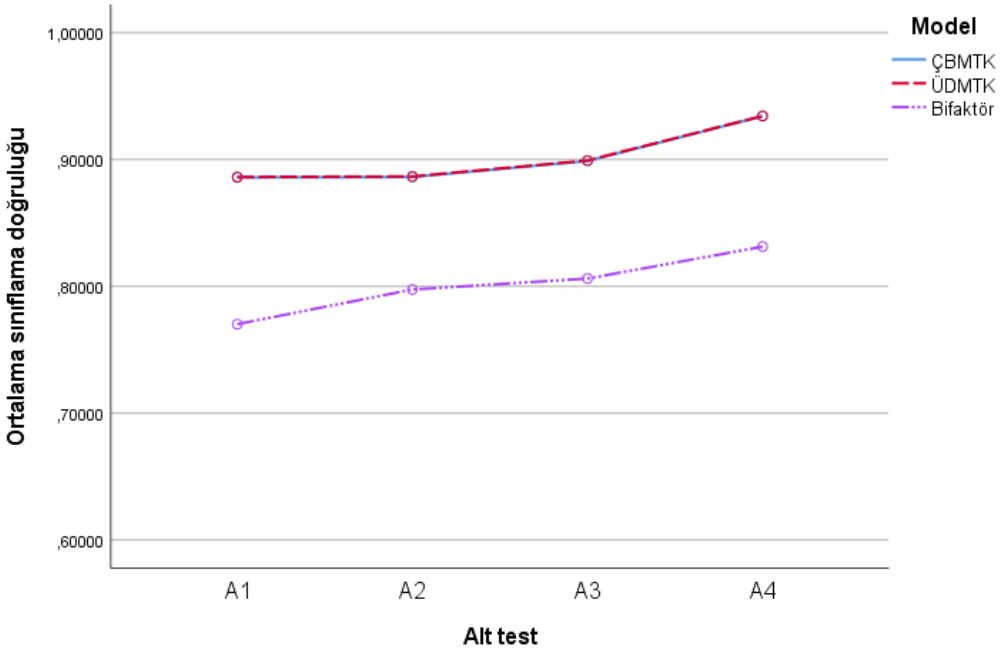
		F	η_p^2
Çoklu puanlanan madde yüzdesi	Model		
	ÇBMTK	3644,165*	,314
	ÜDMTK	3652,474*	,314
	Bifaktör	1113,352*	,122
Model ana etkisi	Çoklu puanlanan madde yüzdesi		
	%5	68776,844*	,852
	%10	68953,182*	,852
	%25	71348,312*	,856
	%50	75231,712*	,863

*p<,001

Tablo 44 incelendiğinde, çoklu puanlanan madde yüzdesi basit ana etkisi için her bir model için (ÇBMTK, ÜDMTK ve Bifaktör) çoklu puanlanan madde yüzdesi düzeylerine ait sınıflama doğruluğu değerleri arasındaki farkın anlamlı olduğu görülmektedir. Her bir model için yapılan analizlere ait etki büyüklükleri incelendiğinde Bifaktör model için hesaplanan etki büyüklüğünün orta ve diğer modeller için yüksek düzeyde olduğu görülmektedir. Farkın hangi çoklu puanlanan madde yüzdeleri arasında olduğunun belirlenmesi amacıyla yapılan ikili

karşılaştırma sonuçları Ek 23'te sunulmuştur. Buna göre bütün ikili karşılaştırmalar anlamlıdır. Bütün modeller altında hesaplanan sınıflama doğruluğu değerleri en düşükten en yükseğe doğru çoklu puanlanan madde yüzdesi sırasıyla %50, %25, %10 ve %5 olduğunda elde edilmiştir. Bir diğer ifadeyle, çoklu puanlanan madde sayısı arttıkça sınıflama doğruluğu değerleri anlamlı olarak artış göstermiştir. Modeller arasındaki fark için ÇBMTK ve ÜDMTK ile elde edilen değerlerin birbirine benzer ve Bifaktör model ile hesaplanan değerlerden daha yüksek olduğu söylenebilir.

Alt puanlar için yapılan faktöriyel karma ANOVA sonuçlarına (Tablo 23) göre anlamlı çıkan bir diğer iki yönlü etkileşim etkisi model ve blok değişken olan alt test arasındadır. Bu etkileşim etkisi için grafiksel gösterim Şekil 47'de yer almaktadır. Model ve alt test için basit ana etkiler incelenmiştir. Bunun için gerçekleştirilen analiz sonuçları Tablo 45'te sunulmuştur.



Şekil 47. Alt puanlar –model x alt test (sınıflama doğruluğu)

Tablo 45.

Alt Puanlar Model x Alt Test Etkileşim Etkisi için Basit Ana Etkiler (Sınıflama Doğruluğu)

		F	η_p^2
Alt test ana etkisi	Model		
	ÇBMTK	12798,814*	,616
	ÜDMTK	12890,216*	,618
	Bifaktör	7769,926*	,493
Model ana etkisi	Alt test		
	Alt test 1	95100,705*	,888
	Alt test 2	56778,972*	,826
	Alt test 3	61541,468*	,837
	Alt test 4	73939,848*	,861

*p<,001

Tablo 45 incelendiğinde, alt test ana etkisi için her bir model özelinde (ÇBMTK, ÜDMTK ve Bifaktör) alt testlere ait sınıflama doğruluğu değerleri arasındaki farkın anlamlı olduğu görülmektedir. ÇBMTK, ÜDMTK ve Bifaktör model için farkın alt testler arasında olduğunun belirlenmesi amacıyla yapılan ikili karşılaştırma sonuçları Ek 24'te sunulmuştur. Buna göre ÇBMTK ve ÜDMTK model için alt test 1 ve alt test 2 arasındaki fark anlamlı değildir ve sırasıyla alt test 3 ve alt test 4 ile elde edilen sınıflama doğruluğu değerleri anlamlı olarak daha yüksektir. Bifaktör model için alt testlere göre elde edilen sınıflama doğruluğu değerleri en yüksekten en düşüğe doğru sırasıyla alt test 4, alt test 3, alt test 2 ve alt test 1 ile elde edilmiştir. Modeller arası fark yine diğer değişkenlerde olduğu gibi ÇBMTK ve ÜDMTK arasındaki fark anlamlı olsa da ortalama sınıflama doğruluğu değerleri arasındaki fark oldukça azdır ve Bifaktör model bütün alt testlerde en düşük sınıflama doğruluğu değerlerine sahiptir.

4.3. Gerçek Veri Analizi için Bulgular

Bu bölümde gerçek veri ile gerçekleştirilen analiz bulgularına yer verilmiştir. ÇBMTK, ÜDMTK, Bifaktör kestirim modellerinin yetenek kestirimlerine ait hata, güvenirlik ve sınıflama doğruluğu değerlerine olan etkisi hem toplam puan hem de alt puan açısından incelenmiştir.

4.3.1. Ön Analizler

Bu çalışma kapsamında kullanılan gerçek veri Türkiye'de bir devlet üniversitesinin Yabancı Diller Meslek Yüksek Okulu tarafından hazırlanıp uygulanan İngilizce yeterlilik sınavına aittir.

Veri karma formatta olup okuma, kelime bilgisi, dinleme ve yazma olmak üzere dört alt alandan oluşmaktadır. Ham veriye ait boyutlar arası korelasyon değerleri Tablo 46'da sunulmaktadır.

Tablo 46.

Gerçek Veriye Ait Alt Puanların Birbiri ve Toplam Puan ile Olan Korelasyon Değerleri

	Dinleme	Okuma	Kelime	Yazma	Toplam
Dinleme	1				
Okuma	0,780*	1			
Kelime	0,651*	0,691*	1		
Yazma	0,501*	0,525*	0,402*	1	
Toplam	0,900*	0,932*	0,829*	0,643*	1

*p<,01

Tablo 46'da yer alan korelasyonlar incelendiğinde alt test puanları arasındaki korelasyonun orta ve yüksek düzeyde olduğu görülmektedir. Yazma alt testi puanları ile diğer alt test puanları arasındaki korelasyon diğerlerine göre daha düşüktür. Toplam test puanı ile dinleme, okuma ve kelime alt test puanları arasındaki korelasyon değerleri yüksektir. Fakat toplam test puanı ve yazma alt testi arasındaki korelasyon diğer alt test puanlarına göre daha düşük ve orta düzeydedir.

İçerik olarak farklı alt alanlara sahip olan İngilizce yeterlilik sınavı verisinin çok boyutluluğunun incelenmesi için Poly-DETECT (doğrulayıcı) ve Doğrulayıcı Faktör Analizi (DFA) uygulanmıştır. Tablo 47'de içerik temelli Poly-DETECT analizi sonuçları yer almaktadır.

Tablo 47.

Poly-DETECT Analizi Sonuçları

İndeks	Hesaplanan Değer	Karşılık gelen sınıflama	
DETECT	0,461	Orta düzey çok boyutluluk	.40 < DETECT < 1.00
ASSI	0,327	Tek boyutluluktan sapma	ASSI > .25
RATIO	0,747		RATIO > .36

Tablo 47'de yer alan çalışma verisi için hesaplanan ASSI (0,327) ve RATIO (0,747) indekslerinin tek boyutluluktan sapmaya işaret etmektedir. Ayrıca DETECT indeksi de orta düzeyde çok boyutluluğa işaret etmektedir. Poly-DETECT analizinden elde edilen bu bulgular, verinin çok boyutlu olduğuna işaret etmektedir. Ek olarak, gerçek veride yer alan dört alt alanı içeren dört faktörlü yapı DFA ile test edilmiştir. Model-veri uyumunun incelenmesi için elde edilen uyum indeksleri ve bunlara ait kriterler Tablo 48'de sunulmuştur.

Tablo 48.

DFA Model-Veri Uyum İndeksleri ve Kriter Değerler

İndeks	Değer	İyi uyum
TLI	0,980	TLI ≥ 0,95
CFI	0,981	CFI ≥ 0,95
RMSEA	0,014	RMSEA ≤ 0,05

Gerçek veri ile gerçekleştirilen DFA ile elde edilen CFI ve TLI indeksleri modelin veriye iyi uyum gösterdiğini ortaya koymaktadır (≥0,95). Benzer şekilde, RMSEA değeri de iyi uyuma işaret etmektedir (≤0,05). DFA sonuçlarına göre, dört faktörlü yapı veriye iyi uyum göstermektedir. Bu bulgu içerik temelli çok boyutluluğu destekler niteliktedir. DFA çıktılarından elde edilen Mcdonald ω güvenirlik katsayıları dinleme, okuma, kelime ve yazma alt testleri için sırasıyla 0,951, 0,962, 0,963 ve 0,975'tir.

DFA ve Poly-DETECT analizleri ile verinin çok boyutlu olduğu ortaya konduktan sonra madde parametreleri ile alt testler ve toplam teste ilişkin yetenek kestirimleri ve bunlara ilişkin standart hata değerleri elde edilmiştir.

4.3.2. Madde Parametreleri

İngilizce yeterlilik sınavına ait veriler BMIRT programı kullanılarak ÇBMTK, ÜDMTK ve Bifaktör modeller ile analiz edilmiştir. Üç tanesi çoklu ve 82 tanesi ikili puanlanan maddelere ait parametreler her üç model için de elde edilmiş ve elde edilen madde ayırt edicilik, güçlük ve şans parametrelerine ait betimsel istatistikler Tablo 49'da sunulmuştur.

Tablo 49.

Gerçek Veri Madde Parametrelerine Ait Betimsel İstatistikler

			Ayırt edicilik				Güçlük				Şans			
		Madde sayısı	$\bar{X}$	ss	min	maks	$\bar{X}$	ss	min	maks	$\bar{X}$	ss	min	maks
ÇBMTK	Dinleme	30	1,128	0,410	0,536	1,977	-0,741	1,219	-3,292	1,154	0,193	0,021	0,143	0,212
	Okuma	33	1,110	0,465	0,050	2,023	-0,954	0,982	-3,406	0,868	0,207	0,013	0,178	0,244
	Kelime	20	1,713	0,567	0,872	2,933	1,155	2,009	-1,726	5,882	0,164	0,037	0,062	0,204
	Yazma	2	1,120	1,037	0,387	1,853	-0,916	1,963	-4,900	2,433	-	-	-	-
	Toplam	85	1,259	0,539	0,050	2,933	-0,463	1,633	-4,900	5,882	0,191	0,029	0,062	0,244
ÜDMTK	Dinleme	30	1,273	0,446	0,618	2,145	-0,747	1,206	-3,305	1,107	0,193	0,022	0,143	0,212
	Okuma	33	1,283	0,516	0,203	2,305	-0,943	0,999	-3,400	0,972	0,209	0,015	0,175	0,248
	Kelime	20	1,927	0,659	0,977	3,362	1,163	2,036	-1,670	5,921	0,165	0,038	0,060	0,217
	Yazma	2	1,357	1,316	0,426	2,287	-0,994	2,154	-5,485	2,638	-	-	-	-
	Toplam	85	1,433	0,605	0,203	3,362	-0,469	1,671	-5,485	5,921	0,193	0,030	0,060	0,248
Bifaktör	Dinleme	30	0,407	0,273	0,000	1,052	-0,741	1,226	-3,289	1,105	0,194	0,022	0,145	0,213
	Okuma	33	0,328	0,376	0,052	1,326	-0,974	1,009	-3,399	0,974	0,208	0,015	0,172	0,251
	Kelime	20	0,909	0,324	0,361	1,465	1,148	1,962	-1,678	5,689	0,165	0,038	0,060	0,215
	Yazma	2	0,608	0,013	0,599	0,617	-0,681	1,328	-2,937	1,440	-	-	-	-
	Toplam	85	0,499	0,398	0,000	1,465	-0,445	1,552	-3,399	5,689	0,192	0,030	0,060	0,251

Tablo 49 incelendiğinde testte yer alan bütün maddelerin ortalama ayırt edicilikleri ÜDMTK için 1,433 olup diğer modeller ile elde edilen değerlerden daha yüksektir. Bifaktör model ile elde edilen ortalama ayırt edicilik düşük düzeydedir. Alt testler bazında da benzer sonuçlar elde edilmiştir. Yazma alt testi en yüksek ortalama ayırt edicilik değerine sahiptir. Bütün maddelere ait ortalama güçlük değerleri her üç modelde de birbirine yakın ve yaklaşık olarak -0,5 ortalama güçlük değeri elde edilmiştir. Ele alınan İngilizce yeterlilik testinin kolay yönde orta güçlük düzeyine yakın olduğu söylenebilir. Alt testler bazında genel olarak dinleme, okuma ve yazma alt testlerine ait güçlük düzeylerinin kolay, fakat kelime alt testinin zor olduğu söylenebilir. Testin geneline ait ortalama şans parametreleri de güçlük değerlerinde olduğu gibi her üç modelde de birbirine yakındır. Maddelere ait şans parametreleri yaklaşık olarak 0,19'dur. Kelime alt testine ait ortalama şans parametresi diğer alt testlerden bir miktar daha düşüktür. Yazma alt testi çoklu puanlanan maddelerden oluştuğu için şans parametresi kestirilmemiştir.

4.3.3. Yetenek Parametreleri, Kestirim Hatası, Güvenirlik ve Sınıflama Doğruluğu

Çalışma kapsamında ele alınan modeller ÇBMTK, ÜDMTK ve Bifaktör ile toplam test ve alt testlere ilişkin yetenek kestirimleri BMIRT programı kullanılarak elde edilmiştir. İlgili modeller ile İngilizce yeterlilik testine giren 5085 kişi için testin geneli ve alt testler için elde edilen yetenek parametrelerine ait betimsel istatistikler Tablo 50'de sunulmuştur.

ÇBMTK, ÜDMTK ve Bifaktör modeller ile elde edilen yetenek parametreleri alt testler ve toplam test bazında yaklaşık olarak 0 ortalamaya sahiptir. Standart sapma değerleri ÇBMTK ve ÜDMTK modelleri için yaklaşık 1 iken Bifaktör model için daha düşüktür. Çarpıklık ve basıklık değerleri incelendiğinde bütün modeller için elde edilen bütün yetenek kestirimlerinin normal dağıldığı görülmektedir.

Tablo 50.

Gerçek Veri Yetenek Parametrelerine Ait Betimsel İstatistikler

		Kişi sayısı	$\bar{X}$	Standart sapma	Minimum	Maksimum	Çarpıklık	Basıklık
ÇBMTK	Dinleme	5085	0,001	0,952	-3,009	2,754	-0,057	-0,354
	Okuma	5085	0,005	1,062	-3,238	2,739	-0,030	-0,538
	Kelime	5085	0,005	1,081	-3,156	2,668	-0,042	-0,647
	Yazma	5085	0,000	1,043	-2,694	2,935	0,154	-0,471
	Toplam	5085	0,008	0,982	-3,063	2,562	-0,268	-0,198
ÜDMTK	Dinleme	5085	0,004	0,954	-3,146	2,938	-0,014	-0,408
	Okuma	5085	0,004	0,945	-3,132	2,866	-0,005	-0,434
	Kelime	5085	0,004	0,955	-3,121	2,916	-0,016	-0,438
	Yazma	5085	0,002	0,934	-2,791	2,896	0,108	-0,377
	Toplam	5085	0,006	0,891	-2,994	2,362	-0,266	-0,136
Bifaktör	Dinleme	5085	0,007	0,958	-3,141	2,820	-0,008	-0,403
	Okuma	5085	0,001	0,556	-1,998	2,013	-0,126	-0,076
	Kelime	5085	0,002	0,573	-2,677	1,479	-0,740	0,774
	Yazma	5085	-0,003	0,669	-1,973	2,666	0,379	-0,039
	Toplam	5085	0,001	0,657	-2,269	1,890	-0,353	-0,040

Kestirilen yetenek parametrelerine ait standart hatalar da elde edilmiştir. Alt testler ve testin tamamı için kestirilen yeteneklere ait standart hata ortalamaları ve standart sapma değerleri Tablo 51'de yer almaktadır.

Tablo 51.

Gerçek Veri Analizi için Toplam Test ve Alt Testlerde Elde Edilen Yetenek Parametrelerine Ait Standart Hata Ortalama ve Standart Sapma Değerleri

	Alt test				
Model	Dinleme (30 madde)	Okuma (33 madde)	Kelime bilgisi (20 madde)	Yazma (2 madde)	Toplam test (85 madde)
ÇBMTK	0,432 (0,050)	0,404 (0,073)	0,445 (0,085)	0,522 (0,047)	0,252 (0,029)
ÜDMTK	0,320 (0,038)	0,290 (0,042)	0,352 (0,060)	0,452 (0,050)	0,295 (0,035)
Bifaktör	0,825 (0,042)	0,811 (0,066)	0,739 (0,075)	0,752 (0,037)	0,288 (0,035)

Tablo 51'de her bir yetenek kestirimi için elde edilen standart hataların ortalama ve standart sapma değerleri yer almaktadır. Alt test yetenek kestirimi için ÇBMTK ve ÜDMTK modelleri ile elde edilen değerler birbirine yakın olmakla birlikte ÜDMTK ile elde edilen hataların daha düşük olduğu görülmektedir. Bifaktör model ise bütün alt testlerde en yüksek ortalama hata değerine sahiptir. Toplam test için ise elde edilen ortalama hata değerleri birbirine yakındır ve alt testler için elde edilen hatalardan daha düşüktür. En düşük ortalama hata ÇBMTK model ile en yüksek ortalama hata ise ÜDMTK model ile kestirilmiştir. Toplam test ve alt testler için yetenek kestirimlerine ait standart hatalar arasındaki farkların istatistiksel olarak anlamlı olup olmadığının incelenmesi için tekrarlı ölçümler ANOVA uygulanmıştır. Sonuçlar Tablo 52'de sunulmuştur.

Tablo 52.

Gerçek Veri Analizi için Standart Hataların Tekrarlı Ölçümler ANOVA ile Karşılaştırılması

Toplam/alt test	Kaynak	Kareler toplamı	*sd*	Kareler ortalaması	*F*	η_p^2	Anlamlı fark
Dinleme	Model	716,594	2	716,594	628674,513*	,992	Bütün ikili karşılaştırmalar
	Hata	5,795	10168	,001			
Okuma	Model	762,691	2	762,691	412478,881*	,988	Bütün ikili karşılaştırmalar
	Hata	9,401	10168	,001			
Kelime	Model	415,005	2	415,005	240242,340*	,979	Bütün ikili karşılaştırmalar
	Hata	8,782	10168	,001			
Yazma	Model	251,047	2	251,047	223127,815*	,978	Bütün ikili karşılaştırmalar
	Hata	5,720	10168	,001			
Toplam	Model	5,362	2	5,362	16046,520*	,759	Bütün ikili karşılaştırmalar
	Hata	1,699	10168	,000			

*p<,001

Yetenek kestirimlerine ilişkin standart hataların modellere göre değişip değişmediğinin incelendiği tekrarlı ölçümler ANOVA sonuçlarına göre bütün alt testler bazında hatalar arasındaki farklar yüksek etki büyüklüğü ile anlamlıdır. Farkın kaynağının incelenmesi için gerçekleştirilen ikili karşılaştırma testleri, bütün ikili karşılaştırmaların anlamlı olduğunu göstermektedir. Buna göre standart hata değerleri düşükten yükseğe doğru sırasıyla ÜDMTK, ÇBMTK ve Bifaktör modelleri ile elde edilmiştir. Alt testler ile yapılan analizlerde, ÇBMTK ve ÜDMTK modellerinin Bifaktör modele göre daha doğru yetenek kestirimi yaptığı belirlenmiştir. Toplam test için de yetenek kestirimlerine ait standart hataların kestirim modeline göre birbirinden anlamlı olarak farklılaştığı tespit edilmiştir. Anlamlı bulunan bu farkın incelenmesi için yapılan ikili karşılaştırmalar, bütün farkların anlamlı olduğunu göstermektedir. Buna göre alt testlerden farklı olarak, toplam test için elde edilen yetenek kestirimlerine ait en yüksek hata ÜDMTK modeli ile kestirilmiştir. En düşük hataya sahip olan model ise ÇBMTK kestirim modelidir. Genel olarak toplam teste ilişkin hataların her üç model için de düşük olduğu ve kestirim doğruluğunun yüksek olduğu söylenebilir.

Toplam test ve alt testler için ÇBMTK, ÜDMTK ve Bifaktör modeller ile elde edilen yetenek kestirimleri, güvenirlik açısından da incelenmiştir. Çalışma kapsamındaki bütün modeller altında kestirilen yeteneklere ilişkin marjinal güvenirlik katsayıları Tablo 53'te sunulmuştur.

Tablo 53.

Gerçek Veri Analizi için Marjinal Güvenirlik Katsayıları

Model	Alt test				Toplam test (85 madde)
	Dinleme (30 madde)	Okuma (33 madde)	Kelime bilgisi (20 madde)	Yazma (2 madde)	
ÇBMTK	0,856	0,874	0,841	0,778	0,934
ÜDMTK	0,896	0,914	0,873	0,794	0,912
Bifaktör	0,312	0,332	0,448	0,432	0,916

Hesaplanan marjinal güvenirlik katsayıları incelendiğinde bütün alt testlerde ÇBMTK ve ÜDMTK modelleri için elde edilen güvenirlik değerlerinin Bifaktör model ile elde edilen değerlerden daha yüksek olduğu görülmüştür. Alt testler için Bifaktör modelde elde edilen güvenirlik değerleri oldukça düşüktür. Toplam test bazında ise her üç kestirim modeli ile elde edilen yetenek kestirimlerine ait güvenirlik değerleri oldukça yüksektir ve birbirlerine yakındır. Toplam test için en yüksek güvenirliğin hesaplandığı model ÇBMTK modelidir.

Son olarak, toplam test ve alt testler için ÇBMTK, ÜDMTK ve Bifaktör modeller ile elde edilen yetenek kestirimlerine ait sınıflama doğruluğu değerleri incelenmiştir. Çalışma kapsamındaki bütün modeller altında kestirilen yeteneklere ilişkin sınıflama doğruluğu değerleri Tablo 54'te sunulmuştur.

Tablo 54.

Gerçek Veri Analizi için Sınıflama Doğruluğu Değerleri

Model	Alt testler				Toplam test (85 madde)
	Dinleme (30 madde)	Okuma (33 madde)	Kelime bilgisi (20 madde)	Yazma (2 madde)	
ÇBMTK	0,898	0,913	0,882	0,865	0,931
ÜDMTK	0,913	0,924	0,895	0,870	0,920
Bifaktör	0,860	0,703	0,726	0,745	0,808

Hesaplanan sınıflama doğruluğu değerleri incelendiğinde bütün alt puanlarda ve toplam puanda en yüksek sınıflama doğruluğu değerleri ÜDMTK ile elde edilmiştir. ÇBMTK ve ÜDMTK modelleri ile yapılan yetenek kestirimi sonunda elde edilen sınıflama doğruluğu değerleri birbirine yakın olup Bifaktör model ile ulaşılan değerlerden daha yüksektir. Toplam teste ait yetenek kestirimleri için sınıflama doğruluğu değerleri her üç model için de alt test ile hesaplanan değerlerden daha yüksektir.

BÖLÜM 5

TARTIŞMA, SONUÇ VE ÖNERİLER

Bu bölümde araştırma sonucunda elde edilen bulgular alan yazındaki bulgular ile tartışılarak sonuca bağlanmış ve elde edilen sonuçlar doğrultusunda araştırmacı ve uygulayıcılara önerilerde bulunulmuştur.

6.1. Sonuç ve Tartışma

Bu çalışmada, toplam puan ve alt puanların birlikte kestirimini sağlayan yöntemlerden Madde Tepki Kuramı'na dayalı çok boyutlu MTK, üst düzey MTK ve Bifaktör modeller incelenmiştir. Karma-formatta hem simülasyon hem de gerçek veri ile elde edilen yetenek kestirimlerine ilişkin kestirim doğruluğu, güvenirlik ve sınıflama doğrulukları açısından ele alınan modeller karşılaştırılmıştır. Simülasyon çalışmasında incelenen diğer değişkenler, testte yer alan çoklu puanlanan madde yüzdesi, alt boyutlar arasındaki korelasyon ve test güçlüğüdür. Ayrıca 5085 kişiye uygulanan dört alt testten oluşan bir İngilizce yeterlilik sınavına ait veriler, ele alınan her üç model ile analiz edilmiş ve elde edilen yetenek kestirimlerinin hataları, güvenirlikleri ve alınan geçti-kaldı kararları ile ilgili sınıflama doğrulukları incelenmiştir.

Bu çalışmanın ana odağını alt testlerden oluşan çok boyutlu bir testte toplam puan ve alt puanların birlikte raporlanması, bu doğrultuda geçerli ve güvenilir yorumlar yapılabilmesi oluşturmaktadır. İdeal olarak uygulanan çok boyutlu bir test sonrasında toplam puan ve alt puanların raporlanması ana hedeftir (Yao, 2012a). Alt puanlardan elde edilen tanılayıcı bilgi, en çok gelişim gösterilen içerik alanının hangisi olduğuna karar verilmesi, öğrencilerin farklı içerik alanlarındaki gelişiminin izlenmesi ve potansiyel öğrenme güçlükleri ve

eksikliklerinin belirlenmesi gibi durumlarda kullanılabilmektedir (Haladyna & Kramer, 2004). Yapılan ölçme uygulaması her bir yanıtlayıcının başarı düzeyini değerlendirmenin yanı sıra eğitim kalitesini yükseltmek adına yanıtlayıcıların zayıf ve güçlü yönlerini de ortaya koyabilmektedir. Bu noktada, birey, kurum ya da ülke bazında alt puanların raporlanıp raporlanmayacağı kararı alınmadan önce bazı önemli noktaların ele alınması gerekmektedir. Raporlanan puanların geçerli, güvenilir ve karşılaştırılabilir olması ve bu standartların alt puanlara da uygulanabilir olması önemli bir husustur. Doğru olmayan toplam puanlar, olası yanlış geçti-kaldı kararlarına sebep olabilir. Benzer şekilde doğru olmayan alt test puanları öğretimin iyileştirilmesi ve eksiklerin telafi edilmesinde yanlış kararlar alınmasına sebep olabilir (Yao, 2012a).

Bu çalışma kapsamında yetenek kestirimlerinin elde edilmesinde ele alınan ÇBMTK, ÜDMTK ve Bifaktör modelleri tek boyutlu modellerden farklı olarak çok boyutlu yaklaşımı içermektedir. Birçok durumda testlerden elde edilen puanların çok boyutlu bir yaklaşımla analiz edilip değerlendirilmesi fayda sağlamaktadır. Johnson ve Carlson (1994) tek bir oturumda birden çok alt testten oluşan bir test uygulamasında bu alt testlerden elde edilen yeteneklerin birbiriyle yüksek düzeyde ilişkili olduğunu belirtmektedir. Çok boyutlu bir yapının kullanılması ile bu ilişki puanlamada dikkate alınmaktadır (de la Torre & Patz, 2005). Çok boyutlu modellere ilişkin yapılan çalışmalar faktörler arası ilişkiyi ve bu ilişkiye dayalı elde edilen bilgiyi dikkate alan yöntemler ile yetenek kestiriminde daha yüksek performans elde edildiğini göstermektedir. (de la Torre vd., 2011; DeMars, 2005; Shin, 2007; Wang vd., 2004).

Araştırmada kullanılan gerçek veri genel İngilizce yeterliliği altında ilişkili alt alanları ölçen alt testlerden oluşan bir testten elde edilmiştir. Bilindiği gibi özellikle dil becerisi farklı alt becerileri (dinleme, okuma, yazma vb.) içerisinde barındırmaktadır ve dil becerisinin ölçülmesindeki ilk hedeflerden biri genel dil yeterliliğinin ortaya konulmasıdır. Bu genel yeterlilik ilişkili bütün alt becerileri kapsayıp bir bütün oluşturmaktadır. Bu yönüyle bu testlerden elde edilen veriler tek boyutlu yaklaşımla analiz edilebilmektedir. Fakat Reckase ve Xu (2015) İngilizce yeterliliğini ölçen teste ait veri setinin tek boyutlu Rasch modeline iyi uyum göstermesine rağmen böyle testlerde maddelere verilen yanıtlar arasındaki ilişkilerin daha iyi açıklanması için birden fazla boyuta ihtiyaç olduğunu belirtmektedir. Çalışmada, ÇBMTK modeli ile alt puanların raporlanmasına ilişkin geçerlilik ve güvenirlik kanıtları elde edilmiş ve özellikle geliştirilmesi gereken alt becerilerin ortaya konmasında ÇBMTK'nin önemi vurgulanmıştır. Özellikle anadiline göre farklı grupların farklı ihtiyaç

ve eksikliklerinin ortaya konmasının öğretmenler, eğitim yöneticileri ve eğitim araştırmacılarına bilgi sağlayacağı belirtilmektedir. Öte yandan Longabah ve Peyton (2018) dil becerisine dair alt alanların incelenmesinde ilişkili değillermiş gibi ayrı ayrı analiz edilmesi ve bu ilişkinin dikkate alınmamasının dil edinim teorisine aykırı olduğunu ifade etmektedir. Çalışmada, dört alt alandan oluşan bir İngilizce yeterlilik testi için alt puanlar, klasik test kuramı, tek boyutlu MTK, takviyeli MTK ve çok boyutlu MTK modelleri ile kestirilmiştir. Bu çalışma kapsamında alt alanlar arasındaki ilişkiyi dikkate alan son iki modelin diğerlerinden daha iyi performans gösterdiği ve çok boyutlu MTK modeli ile en güvenilir kestirimler elde edildiği raporlanmıştır.

Simülasyon çalışmasında hem alt test hem de testin geneli için modellerin performansı karşılaştırılmıştır. Alt puanların kestirimi açısından bütün koşullarda ÇBMTK ve ÜDMTK modelleri ile düşük RMSE ve yüksek güvenirlik ve sınıflama doğruluğu performansları elde edilirken bu iki model arasındaki farkın oldukça az olduğu görülmüştür. de la Torre ve Song (2009) ÜDMTK modeli kullanıldığında alt testlere ait yetenek kestirimlerinin tek boyutlu modele göre daha az hata ile kestirim yaptığını belirtmiştir. Yao ve Boughton (2007) ele alınan bütün koşullarda Bayes ÇBMTK modelinin tek boyutlu modele göre daha güvenilir ve doğru alt puan kestirimleri yaptığını bulmuştur. Bu çalışmada da ÜDMTK ve ÇBMTK modelleri ile doğru ve güvenilir alt puan kestirimleri elde edilmiştir. Ayrıca diğer bazı çalışmalarla alt puanların kestiriminde ÇBMTK ve ÜDMTK yöntemlerinin birbirine benzer sonuçlar ürettiği ortaya konmuştur (de la Torre vd., 2011; Soysal, 2017; Yao, 2010). Bifaktör model ile elde edilen RMSE, güvenirlik ve sınıflama doğruluğu değerleri bu modelin en kötü performansa sahip olduğunu göstermektedir. Liu vd. (2019) de çalışmasında Bifaktör genel modelin alt puanların raporlanmasında kullanılmasını tavsiye etmemiştir. Bifaktör model içerisinde genel ve alt boyutların hesaplanmasında farklı ağırlıklandırma yöntemlerinin kullanılmasını önermişlerdir. Korelasyon 0,2 olduğunda modeller arasındaki fark daha azken korelasyon arttıkça fark açılmış ve Bifaktör model ile daha yüksek RMSE ve daha düşük güvenirlik elde edilmiştir. Benzer şekilde Soysal (2017) da çalışmasında alt testlerin puanlamasında Bifaktör model kullanıldığında boyutlar arasındaki korelasyon arttıkça Bifaktör model ile elde edilen kestirimlerin hatasının arttığını ve güvenirlik değerlerinin düştüğünü ifade etmektedir. Bu çalışmada ele alınan İngilizce yeterlilik sınavı gerçek verisi için elde edilen sonuçlar da simülasyon sonuçlarını doğrulamaktadır. ÇBMTK ve ÜDMTK modelleri ile elde edilen RMSE değerlerinin Bifaktör model ile elde edilen değerlerden daha düşük olduğu gözlenmiş olup güvenirlik ve sınıflama doğruluğu değerleri daha yüksektir.

ÇBMTK, ÜDMTK ve Bifaktör modellerinin performansı testin geneli için de incelenmiş ve alt puanlar ile benzer şekilde toplam puanlar açısından modellere göre farklılıklar olduğu tespit edilmiştir. Yao'nun (2010) çalışması ile benzer şekilde ÇBMTK maksimum bilgi yöntemi ile elde edilen toplam teste ait yetenek kestirimleri en düşük RMSE ve en yüksek güvenirlik değerlerine sahiptir. Ayrıca sınıflama doğruluğu açısından da en iyi performans gösteren model ÇBMTK'dir. de la Torre ve Song (2009) toplam puan ve alt puanların birlikte kestirilmesinde ÜDMTK modelini kullandıkları çalışmada modelin üst düzey olarak tanımlanan genel yeteneğin kestirilmesinde iyi performansa sahip olduğunu belirtmiştir. Ayrıca bu araştırma kapsamında yer almasa da ÜDMTK modelinin birden fazla genel faktör ve bir üst düzey faktör daha içeren yapılar için de kullanılabileceğini vurgulamıştır (de la Torre & Song, 2009). Yao (2010) çalışmasında verilerin üretildiği gerçek modeli de bir değişken olarak ele almış ve hem ÇBMTK hem de ÜDMTK modellerine göre veri üretmiştir. Verilerin üretildiği gerçek model ÇBMTK olduğunda ÇBMTK maksimum bilgi yöntemi bütün kriterlerde en iyi sonucu vermiştir. ÜDMTK gerçek model olduğu durumda ÜDMTK en iyi sonucu vermiştir ve ÇBMTK modeli de korelasyon yüksek olduğunda HOIRT kadar iyi performans sergilemiştir. Bu açıdan bu çalışmada, alt puanlar açısından benzerken toplam puanlar açısından ÇBMTK ve ÜDMTK modelleri arasındaki fark gerçek model ÇBMTK ve gerçek toplam puanlar ÇBMTK maksimum bilgi yöntemine göre elde edilmesinden kaynaklanmış olabilir.

Alt puanlardaki durumun tam tersi olarak toplam puanlar için boyutlar arası korelasyon düzeyi düşük (0,2) olduğunda modeller arasındaki fark daha yüksek iken korelasyon artışıyla modeller arası farklılık azalmıştır. Toplam puan analizlerinde boyutlar arası korelasyon arttıkça ÜDMTK ve Bifaktör modelleri birbirine yakın değerler üretirken korelasyon 0,8 olduğunda başta güvenirlik değerleri olmak üzere RMSE ve sınıflama doğruluğu değerleri açısından modeller arasındaki fark oldukça azalmıştır. Bu çalışma sonuçları ile benzer şekilde Soysal (2017) da ÇBMTK, ÜDMTK ve Bifaktör modeller ile toplam puan ve alt puanların birlikte kestirilmesi üzerine çalışmış ve toplam puan için en düşük RMSE ve en yüksek güvenirlik değerlerinin ÇBMTK modeli ile elde edildiğini bulmuştur. Ayrıca korelasyon 0,8 olduğunda incelediği tüm modeller açısından elde edilen yetenek kestirim performanslarının benzer olduğunu belirtmektedir. Bir diğer çalışmada da boyutlar arasındaki korelasyon 0,7'den yüksek olduğunda ÇBMTK, ÜDMTK ve Bifaktör modelleri arasındaki farkın anlamlı olmadığı bulunmuştur (Yao, 2010). Yine aynı çalışmada, boyutlar arası korelasyon düşük olduğunda özellikle ÜDMTK ve Bifaktör modeller için toplam puana

ait güvenirlikler daha düşük olduğu ifade edilmektedir. Ayrıca simülasyon çalışmasına ek olarak gerçek veri analizi yapılmış ve gerçek model bilinmediği için modellerin göreli performansı ortaya konmuştur. Gerçek veri ile gerçekleştirilen analizlerde ÇBMTK, ÜDMTK ve Bifaktör modellerin toplam teste ilişkin yetenek kestirimleri açısından birbirine benzediği, alt testlere ait yeteneklerin kestiriminde ise ÇBMTK ve ÜDMTK modelleri birbirine yakın sonuçlar verirken bu modellerin Bifaktör modelden daha iyi performans sergiledikleri görülmektedir. Benzer şekilde bu araştırmada ele alınan gerçek veride de boyutlar arası korelasyon yüksek olduğu için bütün değerlendirme kriterlerinde toplam puan için modeller arası farklar azken alt puanlar açısından en iyi performans ÜDMTK ile elde edilmiş, bunu ÇBMTK takip etmiş ve bu iki modelden oldukça farklı olarak en kötü performans Bifaktör model ile elde edilmiştir. Gerçek verinin kullanıldığı bir başka çalışmada da alt boyutlar arasındaki korelasyonun yüksek (0,75) olduğu ve böylece analizde kullanılan ÇBMTK ve ÜDMTK ile birbirine benzer alt puan kestirimleri elde edildiği, modellerin birbirine yakın doğruluk ve güvenirlik ile kestirim yaptığı ve iyi performans gösterdikleri belirlenmiştir (de la Torre vd., 2011).

Çalışma sonuçları, genel olarak Bifaktör modelin diğer modellerden daha yüksek hata, daha düşük güvenirlik ve sınıflama doğruluğu ile kestirim yaptığını göstermektedir. Bu durum özellikle toplam puan analizlerinde boyutlar arası korelasyon düşük olduğu koşulda ve alt puan analizlerinde ise boyutlar arası korelasyon yüksek olduğunda daha belirgindir. Yao'nun (2010) çalışmasında da Bifaktör model için elde edilen toplam puan ve alt puanlar için yetenek kestirim doğruluğu ve güvenirliğin düşük olduğu bulunmuştur. Liu vd. (2019) genel faktör ve alt faktörlerin ayrı ayrı ele alınmasının yani bunlar arasındaki diklik varsayımının yanıltıcı olabileceğini ve olduğundan daha düşük yetenek kestirimlerine sebep olabileceğini belirtmektedir. Örneğin matematiği ölçen genel bir faktörde yüksek puan alan öğrencinin çok düşük geometri puanı olması yanıltıcı olabilir; geometri alt alanı, genel matematik yeterliliğinin bir parçasıdır. Benzer şekilde, alt alanların dâhil edilmediği toplam puan da olduğundan daha düşük kestirilecektir. DeMars (2013) Bifaktör model altında genel ve alt faktörlerin ağırlıklandırılmış birleşimi kullanılarak elde edilen toplam ve alt puanların daha güvenilir olacağını ve daha az hataya sahip olacağını önermiştir. Liu vd. (2019) Bifaktör modeldeki genel ve alt faktörlerin istatistiksel özelliklerine dayalı olarak toplam ve alt puanların hesaplanması için birçok yöntem önermişlerdir. Çalışmada genel faktör ve alt faktörler için farklı ağırlıklandırma yöntemleri kullanılmış ve buna göre toplam puan ve alt puanlar kestirilmiştir. Çalışma sonucunda, bu araştırmada da ele alınan Bifaktör genel

modelin en kötü performansı gösterdiği, ağırlıklandırmalarda ilgili faktörlerdeki maddelere ait ayırt edicilikleri dikkate alan Bifaktör-M4 ve Bifaktör-M6 yöntemleri ile daha doğru ve güvenilir toplam ve alt puanlar elde edildiğini belirtilmiştir.

Brown ve Croudace (2015, s. 331) belirli bir ölçme uygulaması için hangi aracın daha uygun olduğunun sadece model uyumlarına göre değil kavramsal alt yapıya göre de belirlenmesi gerektiğini ifade etmektedir. Ölçmenin odağının ne olduğu (genel faktör ya da alt faktörler) puanlama modelinin seçilmesinde bazı gereklilikler ortaya koymaktadır. Model seçimine teorik açıdan yaklaşıldığında bazı testler içeriğinde birbirinden içerik olarak farklı fakat birbiri ile orta ya da yüksek düzeyde ilişkiye sahip birden çok yapıyı ölçmek için geliştirilmiş olabilir. Bu durumda ilişkili özellik modeli (ÇBMTK) uygun olacaktır. Brown ve Croudace (2015, s. 331) bu modelin, alt faktörler arasında paylaşılan ortak varyansı ortaya koymada ve genel faktör hakkında bilgi verme konusunda başarısız olduğunu söylese de Yao (2010, 2012a) tarafından geliştirilen maksimum bilgi yöntemi ile genel yeteneğe dair kestirimler elde etmek mümkündür. İkinci düzey model (ÜDMTK) ise aynı anda hem genel hem de alt alanlar hakkında bilgi sunabilmektedir. Bu yönüyle hem testin ölçtüğü genel yetenek hem de alt alanlara ait yeteneklerin ortaya konması önemli olduğunda bu modelin kullanımı faydalı olacaktır. Bifaktör modelin ise genel faktöre ilişkin bilgi ana odak fakat alt faktörler de dâhil edilmek istendiğinde kullanışlıdır. Bifaktör modelde alt faktörler, genel faktör tarafından açıklanan varyanstan arda kalan varyansı da ortaya koyarak daha iyi genel yetenek kestirimi hedeflemektedir.

Çalışmada ÇBMTK, ÜDMTK ve Bifaktör modelleri ile elde edilen yetenek kestirimlerine olan etkisi incelenen değişkenlerden biri boyutlar arası korelasyondur. Gerçek veri dört alt boyuttan oluşmaktadır ve simülasyon çalışmasında veriler dört alt boyut olacak şekilde aralarındaki korelasyonlar düşük, orta ve yüksek şeklinde tanımlanarak 0,2, 0,5 ve 0,8 olarak belirlenmiştir. Toplam puan için boyutlar arasındaki korelasyon arttıkça ÇBMTK modeli için hesaplanan RMSE değerleri bir miktar artış gösterirken ÜDMTK ve Bifaktör modelleri için önemli ölçüde düşüş göstermiştir. Her üç model için de elde edilen güvenirlik ve sınıflama doğruluğu değerleri boyutlar arası ilişki arttıkça artış göstermiştir. Bu artış miktarının ÜDMTK ve Bifaktör modeller için fazla, ÇBMTK model için ise düşük düzeyde olduğu tespit edilmiştir. Alt puanlar için ise boyutlar arasındaki korelasyon arttıkça ÇBMTK ve ÜDMTK modelleri için hesaplanan RMSE değerleri az miktarda düşüş gösterirken Bifaktör model için önemli ölçüde düşüş göstermiştir. Güvenirlik ve sınıflama doğruluğu değerleri için boyutlar arası korelasyon arttıkça ÇBMTK ve ÜDMTK modelleri ile elde

edilen değerlerin artış gösterdiği fakat Bifaktör model için korelasyon arttıkça daha düşük değerler elde edildiği görülmüştür. Bu araştırmada kullanılan İngilizce yeterlilik sınavına ait gerçek veride boyutlar arasındaki korelasyonun yüksek olduğu bulunmuştur. Yapılan çalışmalar dil yeterliliği ile ilgili bir alt alanın diğer alt alanlarla ilişkili olduğunu belirtmektedir (Alderson, 2007; Chapelle, 2011; Solano-Flores & Trumbull, 2003). Dolayısıyla simülasyon çalışması sonuçları ile benzer şekilde hem toplam puan hem de alt puan bazında ÇBMTK ve ÜDMTK modelleri Bifaktör modele göre daha iyi performans göstermiştir. Boyutlar arasındaki ilişki yüksek olduğu için alt testlerin kestirilmesinde hesaplanan RMSE değerleri Bifaktör model için oldukça yüksek ve güvenirlik değerleri düşüktür. Sınıflama performansı için fark çok olmasa da diğer iki yöntemden daha düşük sınıflama doğruluğu değerleri elde edilmiştir.

Alan yazındaki birçok çalışmada ÇBMTK'nin yüksek korelasyona sahip birden çok alt testin analizinde daha etkili olduğu bulunmuştur (de la Torre, 2008, 2009; de la Torre & Patz, 2005; Hong, Lam & de la Torre, 2010; Sheng & Wikle, 2008; Yao & Boughton, 2007). de la Torre ve Hong (2010) tarafından yapılan çalışmada ÜDMTK model ile toplam puan ve alt puan kestirimleri elde edilmiş ve boyutlar arası korelasyon artışına göre yapılan incelemelerde korelasyon arttıkça gerçek ve kestirilen yetenekler arasındaki korelasyonun arttığı ve bu artışın toplam puan bazında daha çok olduğu, alt puanlarda farkın daha az olduğu görülmüştür. Bu çalışmada da ÜDMTK için benzer sonuçlar elde edilmiştir.

Bifaktör model için toplam puandaki durumdan farklı olarak korelasyon arttıkça alt puanların hesaplanmasındaki performans daha da kötüleşmiştir. Toplam puan ve alt puanların birlikte kestirilmesinde Bifaktör modelin incelendiği çalışmalarda da genel yetenek kestirimi için hesaplanan RMSE değerleri boyutlar arası korelasyon arttıkça azalırken alt testlere ait yetenek kestiriminde korelasyon arttıkça artmıştır (Liu vd., 2019; Soysal, 2017; Yao, 2010). İki alt faktöre sahip Bifaktör modelin kullanıldığı bir diğer çalışmada da alt faktörlere ait kestirilen parametrelerle gerçek parametreler arasındaki korelasyon, boyutlar arasındaki korelasyon (0,8) yüksek olduğunda, boyutlar arası korelasyon düşük (0,2) olduğu duruma göre daha düşüktür (Jang, 2014). Bu çalışmada Bifaktör model ile elde edilen toplam puanlara ait RMSE, güvenirlik ve sınıflama doğruluğu değerlerinin özellikle 0,2 korelasyon düzeyi için oldukça düşük ve diğer değişkenlere göre de birbirinden farklı olduğu yani bir örüntüye sahip olmadığı görülmüştür. Ayrıca Bifaktör model için alt testler arası farklar incelendiğinde alt test 1'e ait RMSE değerlerinin diğer alt testlerden yüksek ve güvenirlik ve sınıflama doğruluğu değerlerinin düşük olduğu

görülmüştür. Bu çalışmanın kapsamında olmasa da bu durumun olası sebepleri için madde parametreleri incelenmiş ve Bifaktör model için genel faktöre ve birinci alt teste ait madde ayırt ediciliklerinin diğer modeller ile elde edilen değerlerden daha düşük olduğu, ayırt edicilik parametrelerine ilişkin hatanın daha yüksek olduğu tespit edilmiştir. Hambleton, Jones ve Rogers (1993) madde parametreleri kestirimlerine ait hatanın, maddenin doğru tanımlanmasında negatif bir etkiye sahip olduğunu ve bunun da doğru olmayan ya da tutarsız yetenek kestirimine yol açtığını belirtmektedir. Bir başka ifadeyle, doğru madde parametrelerinin elde edilmesi yetenek kestirimlerinin daha doğru kestirilmesi için gereklidir. Ayrıca ayırt edicilik parametresi, maddelerin sağladığı bilgi hususunda önemli bir role sahiptir. Yüksek ayırt ediciliğe sahip maddeler daha çok bilgi sağlarken ayırt ediciliği düşük olan maddeler daha az bilgi sağlamaktadır; MTK bağlamında bilgi, ölçme hatası ile ilişkilidir ve bilgi yüksek olduğunda hata daha düşüktür (Embretson & Reise, 2000; Wang, 2012). Simülasyon çalışması ile benzer şekilde gerçek veri analizlerinde de diğer modellerden daha kötü performans sergileyen Bifaktör model ile elde edilen madde ayırt ediciliklerinin düşük olduğu görülmüştür. Ayrıca Barış-Pekmezci (2018) de çalışmasında, boyutlar arası korelasyon arttıkça Bifaktör model için elde edilen madde ayırt edicilik parametrelerinin yanlılığının arttığını ifade etmektedir.

Çalışma kapsamında, yetenek kestiriminde etkisi incelenen bir diğer değişken testte yer alan maddeler içerisindeki çoklu puanlanan madde yüzdesidir. Bunun için belirlenen düzeyler %5, %10, %25 ve %50 olup toplamda 120 maddeden oluşan testin içinde bu yüzdelere göre sırasıyla 6, 12, 25 ve 60 madde çoklu puanlanmaktadır. Çoklu puanlanan maddeler %5, %10 ve %25 için alt test 4'te yer alırken %50 çoklu puanlanan madde yüzdesi düzeyinde alt test 3 ve alt test 4'ün tamamı çoklu puanlanan maddelerden oluşmaktadır. Çalışma sonuçları toplam test için gerçekleştirilen analizlerde ÇBMTK için çoklu puanlanan madde yüzdesi arttıkça RMSE değerlerinin azaldığını, güvenirlik ve sınıflama doğruluklarının arttığını göstermektedir. ÜDMTK ve Bifaktör modeller için ise korelasyon ve test güçlüğü değişkenlerinin düzeylerinde farklı sonuçlar elde edilmiş olup genel bir örüntüden söz edilememektedir. Alt testler bağlamında ise genel anlamda bütün modeller için çoklu puanlanan madde yüzdesi arttıkça RMSE değerlerinin azaldığı, güvenirlik ve sınıflama doğruluğu değerlerinin arttığı görülmüştür.

ÇBMTK ve ÜDMTK modelleri için tamamı ikili puanlanan maddelerden oluşan alt testler arasındaki güvenirlik, hata ve sınıflama doğruluğu değerlerinin birbirine yakın olduğu görülmüştür. Benzer şekilde çoklu puanlanan madde yüzdesi %50 olduğunda hem alt test 3

hem de alt test 4, tamamı çoklu puanlanan maddelerden oluşmaktadır ve elde edilen değerlendirme kriterlerinin her üçü için de değerler birbirine yakındır. İçerisinde çoklu puanlanan maddeler olan alt testlerde çoklu puanlanan madde yüzdesi arttıkça güvenirlik artmıştır. Karma format test yapısına sahip bir test verisinin kullanıldığı bir çalışmada da ÇBMTK ile yapılan analizlerde içerisinde çoklu puanlanan maddelerin olduğu alt testlerden elde edilen puanlara ilişkin güvenirliklerin daha yüksek standart hataların ise daha düşük olduğu bulunmuştur (Longabach & Peyton, 2018). Bifaktör model için elde edilen sonuçlar da benzer olup tek fark alt test 1 için yetenek kestirim performansının alt test 2 ve alt test 3'ten farklı ve daha kötü olmasıdır.

Çoklu puanlanan madde yüzdesi değişkeni ve buna bağlı olarak alt testlere ilişkin elde edilen bu sonuçlar, çoklu puanlanan maddelerin yer aldığı durumlarda yetenek kestirim hatasının daha az, güvenirliklerin daha çok ve sınıflama performansının daha iyi olduğunu göstermektedir. Kavramsal olarak da çoklu puanlanan maddelerin, ikili puanlanan maddelere göre daha derin ve detaylı ölçümler yapabileceği varsayılmaktadır (Ercikan, Schwarz, Julian, Burket & Link, 1998). Shin (2007) karma format testlerde alt testler arası ilişkiyi dikkate alan Bock, OPI, Wainer ve Shin MCMC modelleri ile alt puanların kestirilmesini inceledikleri çalışmada çoklu puanlanan madde sayısının çoktan seçmeli madde sayısına oranını bir değişken olarak ele almışlardır. Bu oran için düzeyler %0, %20 ve %50 olup çalışma bulguları çoklu puanlanan madde sayısı arttıkça bütün yöntemlerde güvenirliğin arttığını göstermektedir. Buradaki sonuçlar da çalışmadan elde edilen sonuçları desteklemektedir.

Karma format testler için bir diğer hipotez ise ikili puanlanan maddeler ile çoklu puanlanan maddelerin farklı özellikleri ya da boyutları ölçtüğü ile ilgilidir (Traub, 1993). Alt faktörler testin içerik yapısına göre oluşabildiği gibi madde formatına göre de oluşabilmektedir Yao & Schwarz, 2007). Lee vd. (2019) karma format testlerde tek boyutlu MTK'nin kullanımında boyutluluk varsayımı açısından dikkatli olunması gerektiğini vurgulayarak madde formatı etkisi olduğu durumlarda test puanlarının yorumlanması ve kullanımının doğru olmayacağını ifade etmiştir. Çalışmalarında madde formatı etkisinin dikkate alındığı çok boyutlu yöntemlerin kullanımını önermişlerdir. Basit yapılı ve Bifaktör MTK modelleri karma format bir testin ölçülmesinde ele alınmış ve format etkisi yani çok boyutluluk arttıkça bu modellerin tek boyutlu MTK modeline göre daha iyi performans gösterdiği görülmüştür. Alan yazındaki bu bilgiler ışığında bu çalışmadaki tamamı çoklu puanlanan maddelerden ya da ikili puanlanan maddelerden oluşan alt testlerin birbiri ile benzer sonuçlara sahipken

aralarında fark oluşması, karma format test yapısına sahip araştırma verisindeki madde formatı etkisinin boyutluluğu arttırdığı şeklinde yorumlanabilir. Dolayısıyla bu çalışmada ele alınan bütün çok boyutlu modeller için çoklu puanlanan madde yüzdesi arttıkça model performanslarının iyileşmesi çok boyutluluğun artması ile de ilişkilendirilebilir. Gerçek veri analizinde ise çoklu puanlanan madde sayısı üçtür ve bunlardan ikisi birlikte yazma alt testini oluşturmaktadır. Simülasyon çalışmasından farklı olarak çoklu puanlanan maddelerin yer aldığı bu alt test için hata yüksek, güvenirlik ve sınıflama doğruluğu değerleri düşüktür. Burada oluşan fark madde sayısından kaynaklanmış olabilir. Çünkü madde sayısı azaldıkça sağlanan bilgi miktarı bu doğrultuda ise güvenirlik azalmaktadır (Bulut, 2013; Chin, 2011; Yao, 2010; Yavuz, 2014). Reckase ve Xu (2015) İngilizce dil becerisini ölçen bir teste ait veriyi ÇBMTK modeli ile analiz etmiş ve madde sayısı daha az olan alt testlere ait güvenirliklerin daha düşük olduğunu raporlamışlardır.

Test güçlüğünün hem toplam test hem de alt testler bazında yetenek kestirimine etkisi olup olmadığı da incelenmiştir. Test güçlüğü için orta, kolay, çok kolay, zor ve çok zor şeklinde düzeyler belirlenmiştir. Toplam puan açısından genel olarak test güçlüğü orta düzeyde iken en düşük RMSE ve en yüksek güvenirlik değerleri elde edilmiştir. Test zorlaştıkça ve kolaylaştıkça kestirim modellerinin performansının kötüleştiği görülmüştür. Özellikle test çok zor olduğunda bütün modeller için en yüksek RMSE ve en düşük güvenirlik değerleri elde edilmiştir. Modellerin sınıflama performansı test güçlüğü açısından incelendiğinde ise test çok kolay ve kolay olduğunda daha yüksek sınıflama doğrulukları elde edilmiştir. Toplam puan incelemelerinde korelasyon düşük (0,2) olduğunda ÜDMTK ve Bifaktör modeller için farklı sonuçlar elde edilmiş ve tam olarak yukarıdaki gibi bir örüntü gözlemlenmemiştir. Korelasyon yükseldikçe test güçlüğü için yapılan yorumlar ÇBMTK ile benzerdir. Test güçlüğünün alt puanlardaki yetenek kestirimine etkisi de toplam puan ile benzer olup RMSE değerleri en düşükten en yükseğe doğru test güçlüğü sırasıyla orta, kolay, zor, çok kolay ve çok zor olduğunda elde edilmiştir. Güvenirlik için durum tam tersi olup en yüksek güvenirlik orta güçlük ile elde edilirken en düşük güvenirlik test çok zor olduğunda elde edilmiştir. Boyutlar arasındaki korelasyon arttıkça test güçlüklerine göre oluşan bu farklarda azalma görülmüştür.

Madde güçlüklerinin güvenirliğe etkisi diğer bazı çalışmalarda da belirtilmiştir (Allen & Yen, 1979; De Gruijter & van der Kamp, 2008). Paek (2015) çalışmasında test güçlüğünü kolay, orta ve zor diye tanımlayarak ortalama güçlük değerleri sırasıyla -1,5, 0 ve 1,5 olarak belirlemiş ve güvenirliğe olan etkisini incelemiştir. Çalışma sonuçları test güçlüğü orta ve

kolay olduğunda güvenirliklerin birbirine yakın ve yüksek, test zor olduğunda ise güvenirliğin düşük olduğunu göstermektedir. Ayrıca alan yazındaki çalışmalarda test güçlüğü ve yanıtlayıcı grubunun yetenek düzeyi denk olduğunda daha iyi kestirim yapıldığını göstermektedir (Embretson, 2007, s.63; Morse, Johanson & Griffeth, 2012; Paek, 2015; Yao, 2012a). Örneğin, Yao (2012a) dört alt testten oluşan gerçek bir veri ile gerçekleştirdiği çalışmasında madde güçlük ortalaması yetenek ortalamasına en yakın olan alt boyutun diğerlerinden daha yüksek güvenirliğe sahip olduğunu ifade etmektedir. Güvenirliği en yüksek olan bu alt boyuta ait madde güçlüklerinin ortalaması 0,45 ve yetenek parametrelerinin ortalaması 0'dır. Test güçlüğünün orta güçlüğe çok yakın ve yeteneklerin de ortalama düzeyde olduğu söylenebilir. Bu çalışmada da yetenek kestirimlerinin ortalaması sıfır ve test orta güçlükte olduğunda güçlük parametrelerinin ortalaması 0'dır. Gerçek veri analizinde de yetenek kestirimlerinin ortalaması 0 ve test güçlüğü orta düzeye çok yakındır.

6.2. Sonuçlar

Bu araştırma ile elde edilen sonuçlar aşağıda maddeler halinde özetlenmektedir.

1. Alt puanların kestirimi açısından bütün koşullarda ÇBMTK ve ÜDMTK modelleri ile düşük RMSE ve yüksek güvenirlik ve sınıflama doğruluğu performansları elde edilirken bu iki model arasındaki farkın oldukça az olduğu görülmüştür.
2. Alt puan kestirimlerinde, boyutlar arası korelasyon 0,2 olduğunda modeller arasındaki fark daha azken korelasyon arttıkça fark açılmış ve Bifaktör model ile daha yüksek RMSE ve daha düşük güvenirlik elde edilmiştir.
3. ÇBMTK maksimum bilgi yöntemi ile elde edilen toplam teste ait yetenek kestirimleri en düşük RMSE ve en yüksek güvenirlik ve sınıflama doğruluğu değerlerine sahiptir. ÜDMTK modeline ait performans bunu takip etmektedir ve en kötü performans özellikle boyutlar arası korelasyon 0,2 olduğunda Bifaktör model ile elde edilmiştir.
4. Toplam puan analizlerinde boyutlar arası korelasyon arttıkça ÜDMTK ve Bifaktör modelleri birbirine yakın değerler üretirken korelasyon 0,8 olduğunda başta güvenirlik değerleri olmak üzere RMSE ve sınıflama doğruluğu değerleri açısından modeller arasındaki fark oldukça azalmıştır.
5. Genel olarak Bifaktör modelin diğer modellerden daha yüksek hata, daha düşük güvenirlik ve sınıflama doğruluğu ile kestirim yaptığını göstermektedir. Bu durum

özellikle toplam puan analizlerinde boyutlar arası korelasyon düşük olduğu koşulda ve alt puan analizlerinde ise boyutlar arası korelasyon yüksek olduğunda daha belirgindir.

6. Toplam puan için boyutlar arasındaki korelasyon arttıkça ÇBMTK modeli için hesaplanan RMSE değerleri bir miktar artış gösterirken ÜDMTK ve Bifaktör modelleri için önemli ölçüde düşüş göstermiştir. Her üç model için de elde edilen güvenirlik ve sınıflama doğruluğu değerleri boyutlar arası ilişki arttıkça artış göstermiştir.
7. Alt puanlar için ise boyutlar arasındaki korelasyon arttıkça ÇBMTK ve ÜDMTK modelleri için hesaplanan RMSE değerleri az miktarda düşüş gösterirken Bifaktör model için önemli ölçüde düşüş göstermiştir. Güvenirlik ve sınıflama doğruluğu değerleri için boyutlar arası korelasyon arttıkça ÇBMTK ve ÜDMTK modelleri ile elde edilen değerlerin artış gösterdiği fakat Bifaktör model için korelasyon arttıkça daha düşük değerler elde edildiği görülmüştür.
8. Çoklu puanlanan madde yüzdesi arttıkça toplam test için gerçekleştirilen analizlerde ÇBMTK için RMSE değerlerinin azaldığını, güvenirlik ve sınıflama doğruluklarının artmıştır. ÜDMTK ve Bifaktör modeller için ise korelasyon ve test güçlüğü değişkenlerinin düzeylerinde farklı sonuçlar elde edilmiş olup genel bir örüntüden söz edilememektedir. Alt testler bağlamında ise genel anlamda bütün modeller için çoklu puanlanan madde yüzdesi arttıkça RMSE değerlerinin azaldığı, güvenirlik ve sınıflama doğruluğu değerlerinin arttığı görülmüştür.
9. Çoklu puanlanan madde yüzdesi değişkeni ve buna bağlı olarak alt testlere ilişkin elde edilen sonuçlar, çoklu puanlanan maddelerin yer aldığı durumlarda yetenek kestirim hatasının daha az, güvenirliklerin daha çok ve sınıflama performansının daha iyi olduğunu göstermektedir.
10. Toplam puan açısından genel olarak test güçlüğü orta düzeyde iken en düşük RMSE ve en yüksek güvenirlik değerleri elde edilmiştir. Test zorlaştıkça ve kolaylaştıkça kestirim modellerinin performansının kötüleştiği görülmüştür. Özellikle test çok zor olduğunda bütün modeller için en yüksek RMSE ve en düşük güvenirlik değerleri elde edilmiştir.
11. Modellerin sınıflama performansı test güçlüğü açısından incelendiğinde ise test çok kolay ve kolay olduğunda daha yüksek sınıflama doğrulukları elde edilmiştir.

12. Toplam puan incelemelerinde korelasyon düşük (0,2) olduğunda ÜDMTK ve Bifaktör modeller için farklı sonuçlar elde edilmiş ve tam olarak yukarıdaki gibi bir örüntü gözlemlenmemiştir. Korelasyon yükseldikçe test güçlüğü için yapılan yorumlar ÇBMTK ile benzerdir.
13. Test güçlüğünün alt puanlardaki yetenek kestirimine etkisi de toplam puan ile benzer olup RMSE değerleri en düşükten en yükseğe doğru test güçlüğü sırasıyla orta, kolay, zor, çok kolay ve çok zor olduğunda elde edilmiştir. Güvenirlik için durum tam tersi olup en yüksek güvenirlik orta güçlük ile elde edilirken en düşük güvenirlik test çok zor olduğunda elde edilmiştir. Boyutlar arasındaki korelasyon arttıkça test güçlüklerine göre oluşan bu farklarda azalma görülmüştür.

6.3. Öneriler

Bu araştırma kapsamında elde edilen sonuçlar doğrultusunda uygulayıcılara ve araştırmacılara yönelik önerilerde bulunulmuştur.

Uygulayıcılara yönelik öneriler

- Çalışma sonuçlarına dayalı olarak çok boyutlu testlerde hem alt puan hem de toplam puan raporlamasında ÇBMTK ve ÜDMTK modellerinin kullanılması önerilmektedir. Özellikle alt puanların hesaplanmasında bu iki model arasındaki fark oldukça azdır. Bifaktör model ise boyutlar arası korelasyon yüksek olduğunda toplam puan hesaplamasında diğer modeller ile yakın performans gösterdiği için kullanılabilir olsa da alt boyutların hesaplanmasında önerilmemektedir.
- Yetenek kestiriminde etkisi tespit edilen test güçlüğü açısından özellikle testin çok zor olduğu durumlara dikkat edilmelidir. Dolayısıyla model seçimi yapılırken test güçlüklerinin göz önünde bulundurulması özellikle test çok zor yada çok kolay olduğunda ÇBMTK modelinin kullanılması önerilmektedir.
- Karma format bir testte yer alan çoklu puanlanan maddelerin sayısının arttırılması modellerin performansını iyileştirilmesi açısından önerilmektedir.
- Gerçek veri analizinde gerçek model bilinmediği için hangi modelin kullanılması gerektiği hususunda dikkatli olunmalıdır. Simülasyon çalışması sonuçları verinin özelliklerine göre model seçiminde faydalı olabilir. Ayrıca testin yapısı ve raporlamanın amacı da alan uzmanları tarafından ortaya konmalıdır. İmkân

bulunduğunda modellerin hepsi ile analiz gerçekleştirilip ilgili değerlendirme kriterlerine göre kestirimlerin göreli performansları incelenebilir.

Araştırmacılara yönelik öneriler

- Bu çalışmada basit yapılı çok boyutluluk incelenmiştir. Bir maddenin birden fazla boyut ya da içerik alan ile ilişkili olduğu kompleks yapılı çok boyutluluk ele alınarak çalışma tekrarlanabilir.
- Çok boyutlu testlerde toplam puan ve alt puanın birlikte kestirilmesine olanak sağlayan ÇBMTK, ÜDMTK ve Bifaktör modeller ile elde edilen yetenek kestirimlerinin performansı başka değişkenler ile incelenebilir.
- Bifaktör model diklik varsayımından ötürü özellikle ilişkili durumlarda alt puanların kestirilmesinde iyi performans göstermemektedir. Bifaktör model için toplam puan ve alt puanların hesaplanmasında farklı ağırlıklandırmalar kullanılarak boyutlar arası ilişkinin dikkate alındığı yöntemler ile model performansı incelenebilir.

KAYNAKLAR

Ackerman, T. A. (1989). Unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. *Applied Psychological Measurement, 13*, 113-127. https://doi.org/10.1177/014662168901300201

Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education, 7*(4), 255-278. https://doi.org/10.1207/s15324818ame0704_1

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*, 37-51. https://doi.org/10.1111/j.1745-3992.2003.tb00136.x

Adams, R., Wilson, M., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*, 1–23. https://doi.org/10.1177/0146621697211001

Alderson, J. (2007). The challenge of (diagnostic) testing: Do we know what we are measuring? J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Ed.), *Language testing reconsidered* içinde (s. 21–40). Ottawa, Ontario: University of Ottawa Press.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. New York, NY: Brooks/Cole.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014*). Standards for educational and psychological testing*. Washington, DC: AERA.

Ames, A. J. (2015). *Bayesian model criticism: prior sensitivity of the posterior predictive checks method.* Doktora tezi, The Graduate School at The University of North Carolina, Greensboro.

Ansley, T. N., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement, 9*, 37-48. doi:10.1177/014662168500900104

Baker, F. B. (2001). *The basics of item response theory*. United States of America: ERIC Clearinghouse on Assessment and Evaluation.

Barış-Pekmezci, F. (2018). *İki faktör modelde (bifactor) diklik varsayımının farklı koşullar altında sınanması.* Doktora tezi. Hacettepe Üniversitesi.

Bock, R.D., & Schilling, S.G. (2003). IRT Based Item Factor Analysis. M. du Toit (Ed.), *IRT from SSI: BILOGMG, MULTILOG, PARSCALE, TESTFACT* içinde (s. 584-591). Lincolnwood, Illinois: Scientific Software International.

Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64*(2), 153-168. DOI: 10.1007/BF02294533

Brennan, R. L. (2012). *Utility indexes for decisions about subscores*. Center for Advanced Studies in Measurement and Assessment (CASMA). Research Report, 33. https://education.uiowa.edu/sites/education.uiowa.edu/files/documents/centers/casma/publications/casma-research-report-33.pdf sayfasından erişilmiştir.

Brown, A., & Croudace, T. (2015). *Handbook of item response theory modeling: Applications to typical performance assessment*. New York: Routledge.

Bulut, O. (2013). *Between-person and within-person subscore reliability: Comparison of unidimensional and multidimensional IRT models.* Doktora tezi. University Of Minnesota. USA

Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods*, *16*(3), 221–248. http://doi.org/10.1037/a0023350

Carrol, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

Chalmers, R. P. (2012). A Multidimensional item response theory package for the R Environment. *Journal of Statistical Software, 48*(6), 1-29. DOI: 10.18637/jss.v048.i06

Chang, Y.-W., & Davison, M. L. (1992, Nisan). *A comparison of unidimensional and multidimensional IRT approaches to test information in a test battery*. Paper presented

at the annual meeting of the American Educational Research Association, San Franscisco. https://files.eric.ed.gov/fulltext/ED344940.pdf sayfasından erişilmiştir.

Chapelle, C. (2011). Conceptions of validity. G. Fulcher & F. Davidson (Ed.), *The Routledge handbook of language testing* içinde (s. 21–33). London: Routledge.

Chin, T. Y. (2011). *Accuracy and robustness of diagnostic methods: Comparing performance across domain score, multidimensional item response, and diagnostic categorization models*. Doktora tezi. University Of Nebraska. USA

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cronbach, L.J. , & Snow, R.E. (1977). *Aptitude and instructional methods*. New York: Irvington.

De Gruijter, D. N. M., & van der Kamp, L. J. Th. (2008). *Statistical test theory for the behavioral sciences*. Boca Raton, FL: Chapman & Hall/CRC.

de la Torre, J. (2008). Multidimensional scoring of abilities: The ordered polytomous response case. *Applied Psychological Measurement, 32*, 355-370. https://doi.org/10.1177/0146621607303784

de la Torre, J. (2009). Improving the quality of ability estimates through multidimensional scoring and incorporation of ancillary variables. *Applied Psychological Measurement, 33*, 465-485. https://doi.org/10.1177/0146621608329890

de la Torre, J., & Hong, Y. (2010). Parameter estimation with small sample size a higher-order IRT model approach. *Applied Psychological Measurement, 34*, 267-285. https://doi.org/10.1177/0146621608329501

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. https://doi.org/10.3102/10769986030003295

de la Torre, J., & Song, H. (2009). Simultaneous estimation of overall and domain abilities: A higher-order IRT model approach. *Applied Psychological Measurement, 33*(8), 620–639. http://doi.org/10.1177/0146621608326423

de la Torre, J., Song, H., & Hong, Y. (2011). A comparison of four methods of IRT subscoring. *Applied Psychological Measurement, 35*(4), 296–316. http://doi.org/10.1177/0146621610378653

DeMars, C. E. (2005, Ağustos). *Scoring subscales using multidimensional item response theory models.* The Annual Meeting of the American Psychological Association'da

sunulmuş poster, Washington, DC. http://commons.lib.jmu.edu/cgi/viewcontent.cgi?article=1030&context=gradpsych sayfasından erişilmiştir.

DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing, 13*, 354-378. https://doi.org/10.1080/15305058.2013.799067

Edwards, M. C. (2010). A markov chain monte carlo approach to confirmatory item factor analysis. *Psychometrika, 75*, 474-497. https://doi.org/10.1007/s11336-010-9161-9

Embretson, S. E. (2007). Impact of measurement scale in modeling development processes and ecological factors. T. D. Little, J. A. Bovaird, & N. A. Card (Ed.), *Modeling contextual effects in longitudinal studies* içinde (s. 63–87). Lawrence Erlbaum Associates Publishers.

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

Ercikan, K., Schwarz, R. D., Julian, M. W., Burket, G., & Link, V. (1998). Calibration and scoring of tests with multiple-choice and constructed-response item types. *Journal of Educational Measurement, 3,* 137-154. https://doi.org/10.1111/j.1745-3984.1998.tb00531.x

Fan, F. (2016). *Subscore reliability and classification consistency: a comparison of five methods*. Doktora tezi, University of Massachusetts Amherst.

Finch, H. (2010). Item parameter estimation for the MIRT model: Bias and precision of confirmatory factor analysis-based models. *Applied Psychological Measurement, 34*(1), 10-26. https://doi.org/10.1177/0146621609336112

Fu, J., & Qu, Y. (2010, Mayıs). *A comparison of subscore estimation methods on simulated data*. Paper presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*(3), 423–436. http://doi.org/10.1007/BF02295430

Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., Kupfer, D. J., Frank, E., Grochocinski, V. J., & Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement, 31*(1), 4–19. https://doi.org/10.1177/0146621606289485

Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and*

Behavioral Statistics, 33(2), 204-229. https://doi.org/10.3102/1076998607302636

Haberman, S. J., & Sinharay, S. (2010). Reporting of subscores using multidimensional item response theory. *Psychometrika, 75*(2), 209-227. https://doi.org/10.1007/S11336-010-9158-4

Haladyna, T. M., & Kramer, G. A. (2004). The validity of subscores for a credentialing test. *Evaluation & the Health Professions, 27*(4), 349–368. doi: 10.1177/0163278704270010

Hambleton, R. K, & Novick, M. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement, 10*(3), 159-170. https://doi.org/10.1111/j.1745-3984.1973.tb00793.x

Hambleton, R. K., & Han, N. (2005). Assessing the fit of IRT models to educational and psychological test data: A five step plan and several graphical displays. W. R. Lenderking & D. Revicki (Ed.), *Advances in health outcomes research methods, measurement, statistical analysis, and clinical applications* içinde (s. 57-78). Washington, DC: Degnon Associates.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement: Issues and Practice, 12*(3), 38–47. https://doi.org/10.1111/j.1745-3992.1993.tb00543.x

Hambleton, R. K., Jones, R. W., & Rogers, H. J. (1993). Influence of item parameter estimation errors in test development. *Journal ofEducational Measurement, 30*(2), 143-155. https://doi.org/10.1111/j.1745-3984.1993.tb01071.x

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer Nijhoff.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage.

Hambleton, R. K., Swaminathan, H., Cook, L. L., Eignor, D. R., & Gifford, J. A. (1977). Developments in latent trait theory: A review of models, technical issues, and applications. *Review of Educational Research, 48,* 467–510. https://eric.ed.gov/?id=ED137364 sayfasından erişilmiştir.

Han, K. T., & Hambleton, R. K. (2007). *User's manual: WinGen* (Center for Educational Assessment Report No. 642). Amherst, MA: University of Massachusetts, School of

Education.

Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score models. *Journal of Educational Measurement, 27,* 345-359. https://doi.org/10.1111/j.1745-3984.1990.tb00753.x

Hartig, J., & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within- and between-item multidimensionality. *Zeitschrift fu¨r Psychologie/ Journal of Psychology, 216*(2), 89–101. https://doi.org/10.1027/0044-3409.216.2.89

Hartig, J., & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation, 35*(2-3), 57–63. https://doi.org/10.1016/j.stueduc.2009.10.002

Harwell, M., Stone, C. A., Hsu, T. C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied psychological measurement, 20*(2), 101-125. https://doi.org/10.1177/014662169602000201

Hong, Y., Lam, T. C., & de la Torre, J. (2010). *Ancillary variables and multidimensional scoring of polytomous responses.* Paper presented at the annual meeting of the National Council on Measurement in Education, Denver.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huynh, H. (1976). On the reliability of decisions in domain-referenced testing. *Journal of Educational Measurement, 13*(4), 253-264. https://doi.org/10.1111/j.1745-3984.1976.tb00016.x

Huynh, H. (1990). Computation and statistical inference for decision consistency indexes based on the Rasch model. *Journal of Educational Statistics, 15*(4), 353-368. https://doi.org/10.3102/10769986015004353

Jang, E. E., & Roussos, L. (2007). An investigation into the dimensionality of ToEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement, 44*(1), 1-21. https://doi.org/10.1111/j.1745-3984.2007.00024.x

Jang, H. (2014). *Exploring the Estimation of Examinee Locations Using Multidimensional Latent Trait Models Under Different Distributional Assumptions*. Doktora tezi. Michigan State University. Measurement and Quantitative Methods.

Johnson, E. G., & Carlson, J. (1994). *The NAEP 1992 Technical Report* (Report No. 23-TR-20). Washington, DC: National Center for Education Statistics.

Kahraman, N., & Kamata, A. (2004). Increasing the precision of subscale scores by using out-of-scale information. Applied *Psychological Measurement, 28*(6), 407-426. https://doi.org/10.1177/0146621604268736

Kelderman, H., & Rijkes, C. P. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika, 59*(2), 149-176. https://doi.org/10.1007/BF02295181

Kirk, R. (1996). *Statistics: an introduction*. Nelson Education.

Kirk, R. (2013). *Experimental design: Procedures for the behavioral sciences*. Sage.

Kuo, T. C. (2015). *Bayesian Estimation Of Multi-Unidimensional Graded Response IRT Models.* Doktora tezi, Southern Illinois University.

Kuo, T. C., & Sheng, Y. (2016). A Comparison of Estimation Methods for a Multiunidimensional Graded Response IRT Model. *Frontiers in Psychology, 7,* 880. https://doi.org/10.3389/fpsyg.2016.00880

Lane, S. (2005, Nisan). *Status and future directions for performance assessments in education*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Lathrop, Q. (2015). Practical Issues in Estimating Classification Accuracy and Consistency with R Package cacIRT. *Practical Assessment, Research, and Evaluation, 20*(1), 18. https://doi.org/10.7275/43vm-p442

Lazarsfeld, P.F. (1950). The logical and mathematical foundation of latent structure analysis & The interpretation and mathematical foundation of latent structure analysis. S.A. Stouffer et al. (eds.), *Measurement and Prediction*, 362-472. Princeton, NJ: Princeton University Press.

Lee, J. (2012). *Multidimensional item response theory: an investigation of interaction effects between factors on item parameter recovery using Markov Chain Monte Carlo.* Doktora tezi. Michigan State University, Measurement and Quantitative Methods.

Lee, W. C. (2010). Classification consistency and accuracy for complex assessments using item response theory. *Journal of Educational Measurement, 47*(1), 1–17.

http://doi.org/10.1111/j.1745-3984.2009.00096

Lee, W. C., Hanson, B. A., & Brennan, R. L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement, 26*(4), 412-432. https://doi.org/10.1177/014662102237797

Lee, W. C., Kim, S. Y., Choi, J., & Kang, Y. (2020). IRT Approaches to Modeling Scores on Mixed-Format Tests. *Journal of Educational Measurement, 57*(2), 230-254. https://doi.org/10.1111/jedm.12248

Levy, R., Mislevy, R. J., & Sinharay, S. (2009). Posterior predictive model checking for multidimensionality in item response theory. *Applied Psychological Measurement, 33* (7), 519–537. https://doi.org/10.1177/0146621608329504

Li, C. (2018). *Subscore reporting for double coded innovative items embedded in multiple contexts.* Doktora tezi. Graduate School of the University of Maryland, College Park.

Li, C. (2018). *Subscore Reporting for Double-Coded Innovative Items Embedded in Multiple Contexts.* Doktora tezi. University of Maryland.

Li, Y., & Lissitz, R. W. (2012). Exploring the full-information bifactor model in vertical scaling with construct shift. *Applied Psychological Measurement, 36*(1), 3-20. https://doi.org/10.1177/0146621611432864

Lin, P. (2008). *IRT vs. factor analysis approaches in analyzing multigroup multidimensional binary data: The effect of structural orthogonality, and the equivalence in test structure, item difficulty, & examinee groups*. Doktora tezi. University of Maryland.

Liu, Y., Li, Z., & Liu, H. (2019). Reporting valid and reliable overall scores and domain scores using bi-factor model. *Applied psychological measurement, 43*(7), 562-576. https://doi.org/10.1177/0146621618813093

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32*, 179-197. https://doi.org/10.1111/j.1745-3984.1995.tb00462.x

Longabach, T. (2015). *A comparison of subscore reporting methods for a state assessment of English language proficiency*. Doktora tezi, University of Kansas.

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)*. http://doi.org/10.1177/0265532217689949

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*.

Massachusetts: Addison-Wesley.

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

Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement Issues and Practice, 26*(4), 29-37. DOI: 10.1111/j.1745-3992.2007.00106.x

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Md Desa, Z. N. D. (2012). *Bi-factor multidimensional item response theory modeling for subscores estimation, reliability, and classification*. Doktora tezi, University of Kansas. Lawrence, Kansas.

Monaghan, W. (2006). *The facts about subscores* (ETS R&D Connections No. 4). Princeton, NJ: Educational Testing Service. https://www.ets.org/Media/Research/pdf/RD_Connections4.pdf sayfasından erişilmiştir.

Morse, B. J., Johanson, G. A., & Griffeth, R. W. (2012). Using the graded response model to control spurious interactions in moderated multiple regression. *Applied Psychological Measurement, 36*(2), 122-146. https://doi.org/10.1177/0146621612438725

Muraki, E., & Carlson, J. E. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement, 36*(1), 73-90. https://doi.org/10.1177/014662169501900109

Nandakumar, R. (1994). Assessing dimensionality of a set of item responses-comparison of different approaches. *Journal of Educational Measurement, 31*(1), 17-35. http://www.jstor.org/stable/1435095 sayfasından erişilmiştir.

Paek, I. (2015). An investigation of the impact of guessing on coefficient α and reliability. *Applied Psychological Measurement, 39*(4), 264-277. https://doi.org/10.1177/0146621614559516

Patz, R. J., & Junker, B. W. (1999a). Application and extensions of MCMC in IRT: Multiple item types missing data, and related responses. *Journal of Educational and Behavioral Statistics, 24* , 342-366. https://doi.org/10.3102/10769986024004342

Patz, R. J., & Junker, B. W. (1999b). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics, 24*, 146-178. https://doi.org/10.3102/10769986024002146

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. https://www.R-project.org/ sayfasından erişilmiştir

Reckase, M. D. & Hirsh, T. (1991). *Interpretation of number correct scores when the true number of dimensions assessed by a test is greater than two*. Paper presented at the annual meeting of the National Council on Measurement in Education. Chicago, IL. https://files.eric.ed.gov/fulltext/ED331845.pdf sayfasından erişilmiştir.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9*(4), 401-412. https://doi.org/10.1177/014662168500900409

Reckase, M. D. (1995). A linear logistic multidimensional model for dichotomous items response data. W. J. van der Linden and Hambleton (Ed.), *Handbook of Modern Item Response Theory* içinde (s. 271-286). New York: Springer.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21*(1), 25-36. https://doi.org/10.1177/0146621697211002

Reckase, M. D. (2009). *Multidimensional item response theory (statistics for social and behavioral sciences)*. 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*(5), 805-825. https://doi.org/10.1177/0013164414554416

Reckase, M. D., Ackerman, T.A., Carlson, J. E. (1988). Building a unidimensional test using multidimensional items. *Journal of Educational Measurement. 25*(3), 193–204. DOI: 10.1111/j.1745-3984.1988.tb00302.x

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544-559. https://doi.org/10.1080/00223891.2010.496477

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*(1), 19-31. https://doi.org/10.1007/s11136-007-9183-7

Rudner, L. M. (2001). Computing the expected proportions of misclassified examinees. *Practical Assessment Research & Evaluation, 7(14).* https://doi.org/10.7275/an9m-2035

Rudner, L. M. (2005). Expected classification accuracy. *Practical Assessment Research & Evaluation, 10*(13). https://doi.org/10.7275/56a5-6b14

Samejima, F. (1974). Normal Ogive Model on Continuous Response Level in Multidimensional Latent Space. *Psychometrika, 39*(1), 111-121. https://doi.org/10.1007/BF02291580

Sass, D. A., Schmitt, D. A., & Walker, C. M. (2008) Estimating non-normal latent trait distributions within item response theory using true and estimated item parameters. *Applied Measurement in Education, 21*(1), 65-88. https://doi.org/10.1080/08957340701796415

Sheng Y. (2005). *Bayesian analysis of hierarchical IRT models: Comparing and combining the unidimensional & multi-unidimensional IRT models.* Doktora tezi, Faculty of the Graduate School University of Missouri, Columbia.

Sheng Y. and Wikle C. K. (2007). Comparing multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement, 68*(6), 413-430. https://doi.org/10.1177/0013164406296977

Shin, D. (2007). *A comparison of method of estimating subscale scores for mixed-format tests* (Pearson Educational Measurement Research Report). http://assets.pearsonglobalschools.com/asset_mgr/legacy/200746/Estimating%20Subscale%20Scores%20for%20Mixed-Format%20Tests_3961_1.pdf sayfasından erişilmiştir.

Sinharay, S. (2006). Bayesian item fit analysis for unidimensional item response theory models. *British journal of mathematical and statistical psychology, 59*(2), 429-449. https://doi.org/10.1348/000711005X66888

Sinharay, S. (2010). How often do subscores have added value? Results from operational and simulated data. *Journal of Educational Measurement, 47*(2), 150-174. https://doi.org/10.1111/j.1745-3984.2010.00106.x

Sinharay, S., Haberman, S. J., & Wainer, H. (2011). Do adjusted subscores lack validity? Don't blame the messenger. *Educational and Psychological Measurement, 71*(5), 789–797. http://doi.org/10.1177/0013164410391782

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher, 32*(2), 3–13. https://doi.org/10.3102/0013189X032002003

Soysal, S. (2017). *Toplam Test Puanı ve Alt Test Puanlarının Kestiriminin Hiyerarşik Madde Tepki Kuramı Modelleri ile Karşılaştırılması*. Doktora tezi. Hacettepe Üniversitesi.

Soysal, S., & Kelecioğlu, H. (2018). Toplam test ve alt test puanlarının kestiriminin hiyerarşik madde tepki kuramı modelleri ile karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 9*(2), 178-201. https://doi.org/10.21031/epod.404089

Stone, C. A., Ye, F., Zhu, X., & Lane, S. (2010). Providing subscale scores for diagnostic information: A case study when the test is essentially unidimensional. *Applied Measurement in Education, 23*(1), 63–86. http://doi.org/10.1080/08957340903423651

Subkoviak, M. J. (1976). Estimating reliability from a single administration of a criterion-referenced test. *Journal of Educational Measurement, 13*, 265-276. https://doi.org/10.1111/j.1745-3984.1976.tb00017.x

Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion-referenced tests: a decision-theoretic formulation. *Journal of Educational Measurement, 11*(4), 263–267. http://doi.org/10.1111/j.1745-3984.1974.tb00998.x

Şahin, M. G., & Yıldırım, Y. (2018). The examination of item difficulty distribution, test length and sample size in different ability distribution. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 9*(3), 277-294. https://doi.org/10.21031/epod.385000

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics: International edition. Pearson.

te Marvelde, J. M., Glas, C. A., Van Landeghem, G., & Van Damme, J. (2006). Application of multidimensional item response theory models to longitudinal data. *Educational and Psychological Measurement, 66*(1), 5-34. https://doi.org/10.1177/0013164405282490

Toribio, S. G., & Albert, J. H. (2011). Discrepancy measures for item fit analysis in item response theory. *Journal of Statistical Computation and Simulation, 81*(10), 1345-1360. https://doi.org/10.1080/00949655.2010.485131

Traub, R. E. (1993). On the equivalence of traits assessed by multiple-choice and constructed- response tests. R. E. Bennett & W. C. Ward (Ed.), *Construction versus choice in cognitive measurement* içinde (s. 29-44). Hillsdale, NJ: Lawrence Erlbaum.

Wainer, H., & Thissen, D. (1996). How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice, 15*(1), 22-29. https://doi.org/10.1111/j.1745-3992.1996.tb00803.x

Wainer, H., Sheehan, K. M., & Wang, X. (2000). Some paths toward making Praxis scores

more useful. *Journal of Educational Measurement, 37*(2), 113-140. https://doi.org/10.1111/j.1745-3984.2000.tb01079.x

Wainer, H., Vevea, J. L., Camacho, F., Reeve, B. B., Rosa, K., Nelson, L., … Thissen, D. (2001). Augmented sores—"borrowing strength" to compute scores based on small numbers of items. D. Thissen & H. Wainer (Ed.), *Test scoring* içinde (s. 343– 387). Mahwah, NJ: Lawrence Erlbaum Associates.

Walker, C. M., & Beretvas, S. N. (2003). Comparing multidimensional and unidimensional proficiency classifications: multidimensional IRT as a diagnostic aid. *Journal of Educational Measurement, 40*(3), 255-275. https://doi.org/10.1111/j.1745-3984.2003.tb01107.x

Wang, H. (2012). *A Monte Carlo comparison of polytomous item estimation based on higher-order item response theory models versus higher-order confirmatory factor analysis models.* Doktora tezi. University of Pittsburgh.

Wang, T., Kolen, M. J., & Harris, D. J. (2000). Psychometric properties of scale scores and performance levels for performance assessments using polytomous IRT. *Journal of Educational Measurement, 37*(2), 141-162. https://doi.org/10.1111/j.1745-3984.2000.tb01080.x

Wang, W., Song, L., Ding, S., and Meng, Y. (2016). Estimating classification accuracy and consistency indices for multidimensional latent ability. L. A. van der Ark, D. M. Bolt, W. Wang, J. A. Douglas, and M. Wiberg (Ed.) *Quantitative Psychology Research* içinde (Cham: Springer International Publishing AG), 89–103. https://doi.org/10.1007/978-3-319-38759-8_8

Wang, W.-C., Chen, P.-H., & Cheng, Y.-Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods, 9*(1), 116–136. http://doi.org/10.1037/1082-989X.9.1.116

Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). The comparative effects of compensatory and non-compensatory two-dimensional data on unidimensional IRT estimation. *Applied Psychological Measurement, 12*(3), 239-252. https://doi.org/10.1177/014662168801200303

Wedman, J., & Lyrén, P. E. (2015). Methods for examining the psychometric quality of subscores: A review and application. *Practical Assessment, Research, and Evaluation, 20*(1), 21. https://doi.org/10.7275/ng3q-0d19.

Woods, C. M., & Thissen, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika, 71*(2), 281-301. https://doi.org/10.1007/s11336-004-1175-8

Yao, L. (2003). *BMIRT: Bayesian multivariate item response theory [computer software]*. Monterey, CA: DefenseManpower Data Center.

Yao, L. (2010). Reporting valid and reliable overall scores and domain scores. *Journal of Educational Measurement, 47*(3), 339–360. http://doi.org/10.1111/j.1745-3984.2010.00117

Yao, L. (2011). Multidimensional linking for domain scores and overall scores for nonequivalent groups. *Applied Psychological Measurement, 35*(1), 48-66. https://doi.org/10.1177/0146621610373095

Yao, L. (2012a). Multidimensional item response theory for score reporting. Y. Cheng, & H.- H. Chang (Ed.) *Advances in modern international testing: Transition from summative to formative assessment*. Charlotte, NC: Information Age.

Yao, L. (2012b). Multidimensional CAT item selection methods for domain scores and composite scores: Theory and Applications. *Psychometrika, 77*(3), 495-523. https://doi.org/10.1007/s11336-012-9265-5

Yao, L. (2016). *The BMIRT toolkit*. Monterey.

Yao, L., & Boughton, K. A. (2007). A Multidimensional item response modeling approach for improving subscale proficiency estimation and classification. *Applied Psychological Measurement*, *31*(2), 83–105. http://doi.org/10.1177/0146621606291559

Yao, L., & Schwarz, R. D. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. *Applied Psychological Measurement, 30*(6), 469-492. https://doi.org/10.1177/0146621605284537

Yavuz, G. (2014). *Çok boyutlu madde tepki kuramı modelleri ve paket programları için karşılaştırmalı analizler*. Doktora tezi. Hacettepe üniversitesi.

Yavuz, G., & Hambleton, R. K. (2017). Comparative analyses of MIRT models and software (BMIRT and flexMIRT). *Educational and psychological measurement, 77*(2), 263-274. https://doi.org/10.1177/0013164416661220

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of educational measurement, 30*(3), 187-213. https://doi.org/10.1111/j.1745-3984.1993.tb00423.x

Zhang, B. (2010). Assessing the accuracy and consistency of language proficiency classification under competing measurement models. *Language Testing*, *27*(1), 119–140. https://doi.org/10.1177/0265532209347363

Zhang, J. (2007). Conditional covariance theory and detect for polytomous items. *Psychometrika, 72*(1), 69-91. http://doi.org/10.1007/s11336-004-1257-7

Zhang, J., & Stout, W. (1999a). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika, 64*(2), 129-152. https://doi.org/10.1007/BF02294532

Zhang, J., & Stout, W. (1999b). The theoretical DETECT index of dimensionality and its applicationto approximate simple structure. *Psychometrika, 64*(2), 213-249. https://doi.org/10.1007/BF02294536

Zhang, M. (2016). *Exploring dimensionality of scores for mixed-format tests.* Doktora tezi, University of Iowa.

EKLER

Ek 1. SimuMIRT Örnek Betik Dosyası

```
120 5000 4 0 0 0 0 1 0.2 0.2 0.2 0.2 1 0.2 0.2 0.2 0.2 1 0.2 0.2 0.2 0.2 1 4 9231 867989
1        1        1.2400   0.0000   0.0000   0.0000   1.6000   0.2970
2        1        1.9510   0.0000   0.0000   0.0000   -0.5280  0.2570
3        1        0.5930   0.0000   0.0000   0.0000   3.1390   0.3630
4        1        0.6430   0.0000   0.0000   0.0000   0.4200   0.2280
5        1        1.3400   0.0000   0.0000   0.0000   1.3240   0.2800
6        1        2.1970   0.0000   0.0000   0.0000   1.4650   0.2120
7        1        2.2190   0.0000   0.0000   0.0000   1.9350   0.3970
8        1        2.2250   0.0000   0.0000   0.0000   0.8690   0.3660
9        1        2.3000   0.0000   0.0000   0.0000   1.9190   0.2000
10       1        2.1340   0.0000   0.0000   0.0000   1.1120   0.3930
11       1        1.0370   0.0000   0.0000   0.0000   1.2470   0.2610
12       1        2.4880   0.0000   0.0000   0.0000   1.7420   0.2390
13       1        0.5850   0.0000   0.0000   0.0000   2.9030   0.4170
14       1        1.8520   0.0000   0.0000   0.0000   1.4210   0.1770
15       1        0.9320   0.0000   0.0000   0.0000   1.7130   0.3710
16       1        1.4210   0.0000   0.0000   0.0000   3.5150   0.2920
17       1        0.9140   0.0000   0.0000   0.0000   3.1550   0.1290
18       1        2.2500   0.0000   0.0000   0.0000   2.1910   0.2460
19       1        0.6800   0.0000   0.0000   0.0000   1.3930   0.2430
20       1        2.4010   0.0000   0.0000   0.0000   2.6300   0.3710
21       1        1.6610   0.0000   0.0000   0.0000   4.1820   0.2060
22       1        1.8140   0.0000   0.0000   0.0000   -0.2350  0.2330
23       1        0.6080   0.0000   0.0000   0.0000   0.6140   0.3090
24       1        1.2200   0.0000   0.0000   0.0000   2.7660   0.2920
25       1        0.7670   0.0000   0.0000   0.0000   2.9070   0.2560
26       1        1.9010   0.0000   0.0000   0.0000   2.2240   0.2660
27       1        1.7460   0.0000   0.0000   0.0000   2.1140   0.1750
28       1        1.6180   0.0000   0.0000   0.0000   2.5710   0.2200
29       1        2.1410   0.0000   0.0000   0.0000   3.0890   0.1410
30       1        0.5160   0.0000   0.0000   0.0000   0.3110   0.2030
31       1        0.0000   1.4860   0.0000   0.0000   2.8810   0.2000
32       1        0.0000   1.3120   0.0000   0.0000   2.3060   0.2980
33       1        0.0000   0.5730   0.0000   0.0000   2.3050   0.2680
34       1        0.0000   0.5960   0.0000   0.0000   0.6670   0.2000
35       1        0.0000   1.7380   0.0000   0.0000   0.8810   0.2640
36       1        0.0000   0.5310   0.0000   0.0000   2.2060   0.1680
37       1        0.0000   1.1330   0.0000   0.0000   1.9290   0.3420
38       1        0.0000   0.7930   0.0000   0.0000   2.6990   0.2420
39       1        0.0000   1.5860   0.0000   0.0000   1.2970   0.1660
40       1        0.0000   1.0420   0.0000   0.0000   2.3090   0.3890
41       1        0.0000   1.0820   0.0000   0.0000   1.4770   0.2610
42       1        0.0000   0.7500   0.0000   0.0000   1.8250   0.2260
43       1        0.0000   2.4440   0.0000   0.0000   1.6840   0.3220
44       1        0.0000   1.6880   0.0000   0.0000   3.8330   0.1560
45       1        0.0000   2.2750   0.0000   0.0000   1.5350   0.2000
46       1        0.0000   0.8890   0.0000   0.0000   3.3560   0.2020
47       1        0.0000   1.7110   0.0000   0.0000   0.2590   0.1850
48       1        0.0000   2.4110   0.0000   0.0000   1.7660   0.2780
49       1        0.0000   2.0700   0.0000   0.0000   -0.1100  0.2290
50       1        0.0000   0.6850   0.0000   0.0000   -0.3310  0.2080
51       1        0.0000   0.9310   0.0000   0.0000   1.9600   0.3890
52       1        0.0000   0.7920   0.0000   0.0000   2.2960   0.1280
53       1        0.0000   0.7340   0.0000   0.0000   3.8600   0.2580
54       1        0.0000   1.4540   0.0000   0.0000   3.8500   0.2140
55       1        0.0000   2.4440   0.0000   0.0000   1.1710   0.2960
56       1        0.0000   1.0750   0.0000   0.0000   0.7420   0.3380
57       1        0.0000   1.3480   0.0000   0.0000   3.4600   0.1970
58       1        0.0000   1.6280   0.0000   0.0000   1.2630   0.2670
59       1        0.0000   0.5470   0.0000   0.0000   3.0390   0.2970
60       1        0.0000   1.7040   0.0000   0.0000   2.2780   0.1830
61       1        0.0000   0.0000   2.2120   0.0000   1.9880   0.1680
62       1        0.0000   0.0000   0.5420   0.0000   2.5140   0.0530
63       1        0.0000   0.0000   1.1480   0.0000   1.0500   0.2490
```

64	1	0.0000	0.0000	1.3490	0.0000	1.8030	0.1000	
65	1	0.0000	0.0000	1.7550	0.0000	2.3100	0.3050	
66	1	0.0000	0.0000	0.7130	0.0000	1.8470	0.3240	
67	1	0.0000	0.0000	1.9230	0.0000	2.5030	0.1690	
68	1	0.0000	0.0000	1.4410	0.0000	2.2020	0.1620	
69	1	0.0000	0.0000	1.9390	0.0000	2.5720	0.2600	
70	1	0.0000	0.0000	0.6170	0.0000	0.7770	0.2300	
71	1	0.0000	0.0000	0.9700	0.0000	1.8220	0.3080	
72	1	0.0000	0.0000	1.2020	0.0000	1.1930	0.1020	
73	1	0.0000	0.0000	2.1650	0.0000	1.6770	0.1060	
74	1	0.0000	0.0000	1.9790	0.0000	2.3120	0.3610	
75	1	0.0000	0.0000	1.9690	0.0000	3.0400	0.3220	
76	1	0.0000	0.0000	1.1750	0.0000	1.5310	0.2470	
77	1	0.0000	0.0000	1.6790	0.0000	2.0090	0.2540	
78	1	0.0000	0.0000	1.5460	0.0000	3.3390	0.0750	
79	1	0.0000	0.0000	1.8350	0.0000	1.6700	0.2580	
80	1	0.0000	0.0000	1.2590	0.0000	2.4440	0.2600	
81	1	0.0000	0.0000	2.4810	0.0000	1.3680	0.3520	
82	1	0.0000	0.0000	1.3730	0.0000	1.7840	0.3370	
83	1	0.0000	0.0000	0.8770	0.0000	1.8390	0.1620	
84	1	0.0000	0.0000	1.2360	0.0000	2.1090	0.2200	
85	1	0.0000	0.0000	0.7260	0.0000	1.6990	0.1610	
86	1	0.0000	0.0000	1.6040	0.0000	2.1170	0.1700	
87	1	0.0000	0.0000	1.4530	0.0000	1.2980	0.2320	
88	1	0.0000	0.0000	1.8040	0.0000	1.3940	0.1640	
89	1	0.0000	0.0000	0.6360	0.0000	0.8060	0.3810	
90	1	0.0000	0.0000	1.7480	0.0000	3.3300	0.2610	
91	1	0.0000	0.0000	0.0000	1.1230	2.4530	0.2950	
92	1	0.0000	0.0000	0.0000	2.1090	1.7810	0.1520	
93	1	0.0000	0.0000	0.0000	0.9040	-0.6280	0.2510	
94	1	0.0000	0.0000	0.0000	2.4700	3.2230	0.1810	
95	1	0.0000	0.0000	0.0000	2.3110	3.3760	0.0880	
96	1	0.0000	0.0000	0.0000	1.7750	3.5990	0.3240	
97	1	0.0000	0.0000	0.0000	1.5980	1.6960	0.2490	
98	1	0.0000	0.0000	0.0000	1.8630	4.3380	0.1360	
99	1	0.0000	0.0000	0.0000	2.4590	2.1920	0.1890	
100	1	0.0000	0.0000	0.0000	1.4350	0.9560	0.0500	
101	1	0.0000	0.0000	0.0000	1.5700	2.6380	0.2730	
102	1	0.0000	0.0000	0.0000	2.2330	1.6760	0.1490	
103	1	0.0000	0.0000	0.0000	0.6820	3.9980	0.3130	
104	1	0.0000	0.0000	0.0000	1.7650	2.3920	0.3430	
105	1	0.0000	0.0000	0.0000	2.3620	2.3780	0.2430	
106	1	0.0000	0.0000	0.0000	1.3820	2.8580	0.2980	
107	1	0.0000	0.0000	0.0000	0.9230	1.2430	0.2440	
108	1	0.0000	0.0000	0.0000	1.0130	2.1410	0.2600	
109	4	0.0000	0.0000	0.0000	2.4120	0.7580	0.8880	1.6990
110	4	0.0000	0.0000	0.0000	1.4780	0.1790	0.5210	3.8030
111	4	0.0000	0.0000	0.0000	0.7430	0.4920	1.0140	2.3900
112	4	0.0000	0.0000	0.0000	2.3630	0.1600	0.3690	2.1760
113	4	0.0000	0.0000	0.0000	1.6490	0.8750	0.9520	1.9940
114	4	0.0000	0.0000	0.0000	0.6280	1.3310	1.3490	2.0380
115	4	0.0000	0.0000	0.0000	1.3920	2.7730	2.8320	2.9550
116	4	0.0000	0.0000	0.0000	1.1730	1.3990	2.5600	3.4160
117	4	0.0000	0.0000	0.0000	1.0900	1.7880	2.3460	2.3720
118	4	0.0000	0.0000	0.0000	0.9910	1.3760	2.1430	2.2380
119	4	0.0000	0.0000	0.0000	1.7460	0.5570	0.9590	3.0370
120	4	0.0000	0.0000	0.0000	0.6150	1.4290	1.8860	2.5640

Ek 2. BMIRT Örnek Betik Dosyası

```
5000 120 1 1.0 1.0 15000 3000 4 4 92389631 0.0 1.0 0.2 0.1 0.01 1.5 1.5 0.01 0.0 1.5 0.01
100 400 0.01 1.1 0.0 0.05 1.2 0.0 0.05 0.05 0.1

1111111111111111111111111111111111111111111111111111111111111114444444444444
4444444444444444444444444444444444444444444444

1 31 61 91

1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62
63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92
93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115
116 117 118 119 120

1111111111111111111111111111110000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000

0000000000000000000000000000011111111111111111111111111111110000000000000000
000000000000000000000000000000000000000000000000

0000000000000000000000000000000000000000000000000000000000000000001111111111111
1111111111111111110000000000000000000000000000000

0000000000000000000000000000000000000000000000000000000000000000000000000000000
000000000000000000111111111111111111111111111111
```

Ek 3. Simulasyon Analizi İçin Betik Dosyası

```
#!/usr/bin/python3
# -*- coding: utf-8 -*-

import os
import subprocess
from threading import Thread

# conditions are looped by reading from the file. Condition files should be written as kosul111, kosul112, kosul113 one under the other and saved to the file.
# Conditions(kosul111, kosul112)

# variables for this analysis
# analysis_path is the main path
analysis_path = "analiz1"
#c condition_name is name of analysis
condition_name = "kosul11"

main_folder_path = "{0}/{1}".format(analysis_path, condition_name)
conditions_number_file_path = "{0}/conditions".format(analysis_path)
bMIRT28_sh = "{0}/BMIRT28.sh".format(analysis_path)
bMIRTHO_sh = "{0}/BMIRTHO.sh".format(analysis_path)

# count for loop. default value is 1-100
count_start = 1
count_end = 100

# start for conditions
def start_condition():
    #conditions in file. file name contitions
    conditions = read_file_by_line(conditions_number_file_path)
    for condition in conditions:
        print(condition)
        print(get_Bifactor(condition))

        # Definition Thread for a, b and c analysis type
        cond_a = Thread(target=condition_run, args=(condition, 'a'))
        cond_b = Thread(target=condition_run, args=(condition, 'b'))
        cond_c = Thread(target=condition_run, args=(condition, 'c'))

        # Start Thread for cond_a, cond_b and cond_c
        cond_a.start()
        cond_b.start()
        cond_c.start()

def condition_run(cond, type):

    create_all_out_folder(cond, type)
    print("created out folder for {0} condition".format(type))

    ## analysis fot MIRT.ctl file
    MIRT_ctl_path = "{0}/{1}/ctl/{2}/MIRT.ctl".format(main_folder_path, cond, type)
    MIRT_out_path = "{0}/{1}/out/{2}/MIRTout".format(main_folder_path, cond, type)
```

```
    HOIRT_ctl_path = "{0}/{1}/ctl/{2}/HOIRT.ctl".format(main_folder_path, cond, type)
    HOIRT_out_path = "{0}/{1}/out/{2}/HOIRTout".format(main_folder_path, cond, type)

    Bifactor_ctl_path = get_Bifactor(cond)
    print(Bifactor_ctl_path)
    Bifactor_out_path = "{0}/{1}/out/{2}/Bifactor".format(main_folder_path, cond, type)

    ## Definition thread according to ctl file
    MIRT = Thread(target=condition_method, args=(cond, type, MIRT_ctl_path, MIRT_out_path,
"28"))
    HOIRT = Thread(target=condition_method, args=(cond, type, HOIRT_ctl_path,
HOIRT_out_path, "HO"))
    Bifactor = Thread(target=condition_method, args=(cond, type, Bifactor_ctl_path,
Bifactor_out_path, "28"))

    # # start thread for MIRT, HOIRT and Bifactor
    MIRT.start()
    HOIRT.start()
    Bifactor.start()

def condition_method(cond, type, ctl_path, out_path, bMIRT):
    print("---------------->>>>>>>>>>> " + str(ctl_path))
    count = count_end + 1
    for data in range(count_start, count):
        data_path = "{0}/{1}/data/{2}/{3}.rwo".format(main_folder_path, cond, type, data)
        if bMIRT == "HO":
            run_command("/bin/bash {param0} {param1} {param2} {param3}/{param4}
{param5}".format(param0=bMIRTHO_sh, param1=ctl_path,
                                                    param2=data_path, param3=out_path,
                                                    param4=data, param5=analysis_path))
        else:
            run_command("/bin/bash {param0} {param1} {param2} {param3}/{param4}
{param5}".format(param0=bMIRT28_sh, param1=ctl_path, param2=data_path, param3=out_path,
param4=data, param5=analysis_path))
        print("---------->>>> END ANALYSIS {0} for {1} method".format(cond, type))

        param_txt_path = "{0}/{1}-param.txt".format(out_path, data)
        theta_txt_path = "{0}/{1}-theta.txt".format(out_path, data)

        pop_txt_path = "{0}/{1}-pop.txt".format(out_path, data)
        likehood_txt_path = "{0}/{1}-likehood.txt".format(out_path, data)
        overall_txt_path = "{0}/{1}-overall.txt".format(out_path, data)

        # delete files param.txt, pop.txt, likehood.txt and theta.txt
        if is_exist(param_txt_path):
            delete_file(param_txt_path)

        if is_exist(theta_txt_path):
            delete_file(theta_txt_path)

        if is_exist(pop_txt_path):
            delete_file(pop_txt_path)

        if is_exist(likehood_txt_path):
            delete_file(likehood_txt_path)
```

```
        if is_exist(overall_txt_path):
            delete_file(overall_txt_path)

#create all out folder
def create_all_out_folder(cond, type):

    if not is_exist("{0}/{1}/out".format(main_folder_path, cond)):
        create_directory("{0}/{1}/out".format(main_folder_path, cond))
    if not is_exist("{0}/{1}/out/{2}".format(main_folder_path, cond, type)):
        create_directory("{0}/{1}/out/{2}".format(main_folder_path, cond, type))

    # out folder fot MIRTout
    if not is_exist("{0}/{1}/out/{2}/MIRTout".format(main_folder_path, cond, type)):
        create_directory("{0}/{1}/out/{2}/MIRTout".format(main_folder_path, cond, type))
    # out folder for HOIRTout
    if not is_exist("{0}/{1}/out/{2}/HOIRTout".format(main_folder_path, cond, type)):
        create_directory("{0}/{1}/out/{2}/HOIRTout".format(main_folder_path, cond, type))
    # out folder for MIRTout
    if not is_exist("{0}/{1}/out/{2}/Bifactor".format(main_folder_path, cond, type)):
        create_directory("{0}/{1}/out/{2}/Bifactor".format(main_folder_path, cond, type))
    # out folder for MIRToverallout
    if not is_exist("{0}/{1}/out/{2}/MIRToverallout".format(main_folder_path, cond, type)):
        create_directory("{0}/{1}/out/{2}/MIRToverallout".format(main_folder_path, cond, type))
        create_directory("{0}/{1}/out/{2}/MIRToverallout".format(main_folder_path, cond, type))

# return Bifactor_ctl path for kosul11
def get_Bifactor(condition):
    # Bifactor ctl file for kosul11 analysis
    Bifactor_ctl_111 = "{0}/Bifactor111.ctl".format(main_folder_path)
    Bifactor_ctl_121 = "{0}/Bifactor121.ctl".format(main_folder_path)
    Bifactor_ctl_131 = "{0}/Bifactor131.ctl".format(main_folder_path)
    Bifactor_ctl_141 = "{0}/Bifactor141.ctl".format(main_folder_path)

    Bifactor_ctl_path = None

    if condition == "kosul111" or condition == "kosul112" or condition == "kosul113" or condition ==
"kosul114" or condition == "kosul115":
        Bifactor_ctl_path = Bifactor_ctl_111

    if condition == "kosul121" or condition == "kosul122" or condition == "kosul123" or condition ==
"kosul124" or condition == "kosul125":
        Bifactor_ctl_path = Bifactor_ctl_121

    if condition == "kosul131" or condition == "kosul132" or condition == "kosul133" or condition ==
"kosul134" or condition == "kosul135":
        Bifactor_ctl_path = Bifactor_ctl_131

    if condition == "kosul141" or condition == "kosul142" or condition == "kosul143" or condition ==
"kosul144" or condition == "kosul145":
        Bifactor_ctl_path = Bifactor_ctl_141

    return Bifactor_ctl_path

# run command with subprocess
```

```
def run_command(cmd):
    p = subprocess.Popen(cmd, shell=True, stdout=subprocess.PIPE, stderr=subprocess.STDOUT)
    stdout = []
    while True:
        line = p.stdout.readline()
        stdout.append(line)
        print(str(line.strip()))
        if not line:
            break

# return full path status(true or false)
def is_exist(full_path):
    try:
        return os.path.exists(full_path)
    except:
        raise

# created directory by dir_path
def create_directory(dir_path):
    try:
        #return os.makedirs(dir_path)
        os.system("mkdir -p "+ str(dir_path))
    except:
        raise

# return lines for read by line
def read_file_by_line(full_path, mode='r'):
    with open(full_path) as f:
        lines = [line.rstrip() for line in f]
    return lines

#delete file by file_path
def delete_file(full_path):
    try:
        os.remove(full_path)
    except:
        raise

#create file by file_apth
def create_file(full_path):
    try:
        if os.path.exists(full_path):
            return None
        else:
            file = open(full_path, 'w')
            file.close()
            return True
    except:
        raise

start_condition()
```

Ek 4. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (*r*=0,2)

Korelasyon 0,2 - Çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25 ve %50) her bir model için test güçlüklerinde hesaplanan ortalama RMSE, güvenirlik ve sınıflama doğruluğu değerleri

		Ortalama RMSE				Ortalama güvenirlik				Ortalama sınıflama doğruluğu			
Model	Test güçlüğü	Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
		%5	%10	%25	%50	%5	%10	%25	%50	%5	%10	%25	%50
ÇBMTK	Çok zor	,314	,273	,255	,185	,897	,918	,939	,966	,877	,884	,891	,939
	Zor	,211	,197	,161	,128	,929	,944	,970	,978	,869	,889	,917	,951
	Orta güçlük	,164	,149	,121	,098	,944	,956	,978	,982	,875	,891	,937	,927
	Kolay	,164	,151	,128	,113	,942	,955	,972	,979	,880	,904	,942	,932
	Çok kolay	,202	,192	,179	,155	,934	,941	,957	,969	,893	,900	,922	,949
ÜDMTK	Çok zor	,605	,565	,571	,554	,815	,811	,799	,821	,722	,725	,726	,736
	Zor	,453	,450	,461	,459	,864	,855	,796	,825	,736	,737	,738	,742
	Orta güçlük	,366	,365	,414	,382	,882	,870	,798	,846	,741	,742	,743	,747
	Kolay	,374	,405	,427	,444	,893	,853	,816	,834	,743	,743	,745	,747
	Çok kolay	,471	,485	,525	,537	,854	,853	,809	,824	,738	,741	,742	,746
Bifaktör	Çok zor	,794	,617	,714	,677	,582	,739	,612	,666	,830	,754	,779	,768
	Zor	,478	,513	,763	,500	,831	,777	,486	,789	,746	,751	,828	,753
	Orta güçlük	,499	,698	,729	,480	,746	,571	,491	,744	,765	,840	,837	,762
	Kolay	,646	,434	,540	,466	,635	,819	,696	,818	,825	,767	,784	,790
	Çok kolay	,491	,700	,547	,575	,825	,614	,783	,789	,785	,847	,788	,815

Korelasyon 0,2 - Çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25 ve %50) her bir model için test güçlüklerinin ikili karşılaştırmaları

			RMSE				Güvenirlik				Sınıflama doğruluğu			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50	%5	%10	%25	%50
Model	(I) test güçlüğü	(J) test güçlüğü	Ortalama farkı (I-J)				Ortalama farkı (I-J)				Ortalama farkı (I-J)			
ÇBMTK	Çok zor	Zor	,103*	,077*	,095*	,057*	-,032*	-,026*	-,031*	-,012*	,008*	-,005*	-,026*	-,013*
		Orta güçlük	,150*	,125*	,135*	,087*	-,047*	-,038*	-,039*	-,016*	,002*	-,007*	-,046*	,012*
		Kolay	,151*	,123*	,127*	,072*	-,044*	-,037*	-,034*	-,012*	-,003*	-,019*	-,051*	,007*
		Çok kolay	,113*	,082*	,077*	,030*	-,037*	-,023*	-,019*	-,003*	-,016*	-,016*	-,031*	-,010*
	Zor	Orta güçlük	,047*	,048*	,040*	,030*	-,015*	-,012*	-,008*	-,004*	-,006*	-,002*	-,020*	,025*
		Kolay	,047*	,046*	,033*	,014*	-,013*	-,011*	-,003*	-,001*	-,011*	-,015*	-,025*	,020*
		Çok kolay	,009*	,005*	-,018*	-,027*	-,005*	,003*	,012*	,008*	-,024*	-,011*	-,005*	,003*
	Orta güçlük	Kolay	0,000	-,002*	-,007*	-,016*	,003*	,001*	,005*	,004*	-,005*	-,013*	-,006*	-,005*
		Çok kolay	-,038*	-,043*	-,058*	-,057*	,011*	,015*	,020*	,013*	-,018*	-,009*	,014*	-,022*
	Kolay	Çok kolay	-,038*	-,041*	-,051*	-,041*	,008*	,014*	,015*	,009*	-,013*	,003*	,020*	-,017*
ÜDMTK	Çok zor	Zor	,151*	,115*	,110*	,095*	-,049*	-,044*	0,004	-,004*	-,014*	-,013*	-,011*	-,006*
		Orta güçlük	,238*	,200*	,157*	,173*	-,068*	-,059*	0,001	-,024*	-,019*	-,017*	-,016*	-,011*
		Kolay	,231*	,160*	,144*	,110*	-,078*	-,042*	-,017*	-,013*	-,020*	-,018*	-,019*	-,011*
		Çok kolay	,133*	,080*	,046*	,017*	-,039*	-,042*	-,009*	-0,003	-,016*	-,017*	-,016*	-,010*
	Zor	Orta güçlük	,087*	,085*	,047*	,077*	-,018*	-,015*	-0,002	-,020*	-,005*	-,004*	-,005*	-,004*
		Kolay	,080*	,046*	,034*	,014*	-,029*	0,002	-,020*	-,008*	-,007*	-,006*	-,007*	-,005*
		Çok kolay	-,018*	-,035*	-,063*	-,079*	,010*	0,002	-,013*	0,002	-,002*	-,004*	-,004*	-,004*
	Orta güçlük	Kolay	-,007*	-,040*	-,013*	-,063*	-,011*	,017*	-,018*	,012*	-,001*	-,001*	-,003*	-0,001
		Çok kolay	-,105*	-,120*	-,110*	-,156*	,028*	,017*	-,010*	,022*	,003*	0,000	0,001	0,000
	Kolay	Çok kolay	-,098*	-,081*	-,098*	-,093*	,039*	0,001	,007*	,010*	,004*	,002*	,003*	0,001
Bifaktör	Çok zor	Zor	,316*	,104*	-,049*	,176*	-,250*	-,039*	,127*	-,123*	,084*	0,003	-,048*	,015*
		Orta güçlük	,295*	-,081*	-,015*	,197*	-,165*	,168*	,121*	-,077*	,065*	-,087*	-,058*	,006*
		Kolay	,148*	,182*	,174*	,211*	-,053*	-,080*	-,084*	-,152*	,005*	-,013*	-,005*	-,021*
		Çok kolay	,303*	-,084*	,167*	,102*	-,243*	,125*	-,171*	-,122*	,045*	-,093*	-,009*	-,047*
	Zor	Orta güçlük	-,021*	-,185*	,034*	,020*	,085*	,206*	-0,005	,045*	-,019*	-,090*	-,009*	-,009*
		Kolay	-,168*	,078*	,223*	,034*	,197*	-,042*	-,210*	-,029*	-,079*	-,016*	,044*	-,037*
		Çok kolay	-,013*	-,187*	,216*	-,074*	,006	,163*	-,297*	0,000	-,039*	-,096*	,039*	-,062*
	Orta güçlük	Kolay	-,147*	,264*	,189*	,014*	,112*	-,248*	-,205*	-,075*	-,060*	,074*	,053*	-,028*
		Çok kolay	0,008	-0,002	,182*	-,094*	-,079*	-,043*	-,292*	-,045*	-,020*	-,007*	,049*	-,053*
	Kolay	Çok kolay	,155*	-,266*	-0,007	-,109*	-,190*	,205*	-,087*	,030*	,040*	-,080*	-,004*	-,026*

*p<,001

Korelasyon 0,2 - Çoklu puanlanan madde yüzdesi düzeylerinde her bir test güçlüğü için modellerin ikili karşılaştırmaları

			RMSE				Güvenirlik				Sınıflama doğruluğu			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50	%5	%10	%25	%50
Test güçlüğü	(I) Model	(J) Model	Ortalama farkı (I-J)				Ortalama farkı (I-J)				Ortalama farkı (I-J)			
Çok zor	ÇBMTK	ÜDMTK	-,290*	-,292*	-,315*	-,369*	,082*	,107*	,139*	,145*	,155*	,160*	,165*	,203*
		Bifaktör	-,480*	-,343*	-,458*	-,492*	,315*	,179*	,326*	,300*	,047*	,131*	,112*	,171*
	ÜDMTK	Bifaktör	-,190*	-,051*	-,143*	-,122*	,233*	,072*	,187*	,155*	-,108*	-,029*	-,053*	-,032*
Zor	ÇBMTK	ÜDMTK	-,242*	-,254*	-,301*	-,331*	,065*	,089*	,174*	,152*	,134*	,152*	,179*	,209*
		Bifaktör	-,267*	-,316*	-,602*	-,373*	,098*	,167*	,484*	,189*	,123*	,138*	,089*	,199*
	ÜDMTK	Bifaktör	-,025*	-,062*	-,302*	-,042*	,033*	,078*	,310*	,037*	-,010*	-,013*	-,090*	-,011*
Orta güçlük	ÇBMTK	ÜDMTK	-,203*	-,217*	-,293*	-,284*	,062*	,086*	,179*	,137*	,134*	,149*	,194*	,180*
		Bifaktör	-,335*	-,549*	-,608*	-,382*	,198*	,385*	,487*	,238*	,110*	,051*	,100*	,165*
	ÜDMTK	Bifaktör	-,133*	-,333*	-,315*	-,099*	,136*	,299*	,307*	,102*	-,024*	-,099*	-,094*	-,015*
Kolay	ÇBMTK	ÜDMTK	-,210*	-,254*	-,299*	-,331*	,049*	,102*	,156*	,145*	,138*	,161*	,197*	,184*
		Bifaktör	-,483*	-,284*	-,412*	-,353*	,307*	,136*	,276*	,160*	,055*	,137*	,158*	,142*
	ÜDMTK	Bifaktör	-,272*	-,029*	-,113*	-,022*	,258*	,034*	,120*	,015*	-,083*	-,024*	-,039*	-,043*
Çok kolay	ÇBMTK	ÜDMTK	-,270*	-,294*	-,346*	-,383*	,080*	,088*	,149*	,146*	,155*	,159*	,180*	,202*
		Bifaktör	-,290*	-,509*	-,368*	-,420*	,109*	,327*	,174*	,181*	,108*	,053*	,134*	,133*
	ÜDMTK	Bifaktör	-,020*	-,215*	-,022*	-,037*	,029*	,239*	,025*	,035*	-,047*	-,106*	-,046*	-,069*

*p<,001

Ek 5. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (*r*=0,5)

Korelasyon 0,5 - Çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25 ve %50) her bir yöntem için test güçlüklerinde hesaplanan ortalama RMSE değerleri

		Ortalama RMSE				Ortalama güvenirlik				Ortalama sınıflama doğruluğu			
Model	Test güçlüğü	Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
		%5	%10	%25	%50	%5	%10	%25	%50	%5	%10	%25	%50
ÇBMTK	Çok zor	,343	,303	,287	,213	,911	,928	,942	,969	,898	,907	,915	,947
	Zor	,235	,217	,178	,148	,946	,956	,973	,981	,903	,914	,938	,956
	Orta güçlük	,181	,165	,135	,110	,962	,969	,982	,987	,912	,923	,950	,950
	Kolay	,185	,172	,149	,136	,961	,968	,979	,974	,916	,932	,953	,947
	Çok kolay	,235	,226	,217	,192	,951	,955	,965	,974	,924	,929	,944	,956
ÜDMTK	Çok zor	,521	,487	,493	,478	,879	,887	,879	,905	,825	,830	,831	,842
	Zor	,415	,412	,412	,417	,924	,922	,899	,913	,846	,848	,850	,854
	Orta güçlük	,365	,355	,376	,366	,940	,939	,910	,928	,857	,858	,860	,863
	Kolay	,362	,379	,384	,404	,945	,932	,920	,916	,863	,863	,866	,868
	Çok kolay	,422	,430	,459	,469	,925	,926	,908	,916	,864	,866	,868	,872
Bifaktör	Çok zor	,542	,495	,512	,512	,861	,879	,866	,893	,829	,825	,827	,837
	Zor	,420	,423	,463	,438	,920	,914	,865	,906	,832	,836	,841	,841
	Orta güçlük	,378	,400	,424	,385	,931	,912	,880	,915	,840	,847	,852	,847
	Kolay	,385	,386	,403	,419	,932	,927	,909	,912	,852	,851	,856	,864
	Çok kolay	,425	,451	,467	,492	,921	,911	,904	,911	,858	,866	,863	,877

Korelasyon 0,5 - Çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25 ve %50) her bir yöntem için test güçlüklerinin ikili karşılaştırmaları

			RMSE				Güvenirlik				Sınıflama doğruluğu			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50	%5	%10	%25	%50
Model	(I) test güçlüğü	(J) test güçlüğü	Ortalama farkı (I-J)				Ortalama farkı (I-J)				Ortalama farkı (I-J)			
ÇBMTK	Çok zor	Zor	,109*	,086*	,108*	,065*	-,035*	-,028*	-,031*		-,005*	-,007*	-,023*	-,009*
		Orta güçlük	,163*	,139*	,151*	,103*	-,051*	-,041*	-,040*		-,014*	-,016*	-,035*	-,004*
		Kolay	,158*	,132*	,138*	,077*	-,051*	-,040*	-,037*		-,018*	-,025*	-,038*	0,000
		Çok kolay	,109*	,077*	,069*	,021*	-,040*	-,027*	-,023*		-,026*	-,021*	-,030*	-,010*
	Zor	Orta güçlük	,054*	,052*	,043*	,038*	-,015*	-,013*	-,009*		-,009*	-,009*	-,012*	,005*
		Kolay	,049*	,045*	,029*	,012*	-,015*	-,012*	-,006*		-,014*	-,018*	-,015*	,009*
		Çok kolay	0,000	-,009*	-,039*	-,044*	-,004*	,001*	,008*		-,021*	-,014*	-,007*	-,001
	Orta güçlük	Kolay	-,004*	-,007*	-,014*	-,026*	0,000	,001*	,003*		-,005*	-,009*	-,003*	,003*
		Çok kolay	-,054*	-,061*	-,082*	-,082*	,011*	,014*	,017*		-,012*	-,006*	,006*	-,006*
	Kolay	Çok kolay	-,050*	-,054*	-,068*	-,055*	,011*	,013*	,014*		-,007*	,003*	,008*	-,010*
ÜDMTK	Çok zor	Zor	,106*	,074*	,081*	,061*	-,045*	-,036*	-,019*		-,021*	-,018*	-,019*	-,012*
		Orta güçlük	,156*	,132*	,117*	,112*	-,061*	-,052*	-,031*		-,032*	-,028*	-,029*	-,021*
		Kolay	,159*	,108*	,109*	,074*	-,066*	-,046*	-,041*		-,038*	-,033*	-,035*	-,026*
		Çok kolay	,099*	,056*	,034*	,008*	-,046*	-,039*	-,029*		-,039*	-,036*	-,036*	-,030*
	Zor	Orta güçlük	,050*	,058*	,036*	,051*	-,016*	-,016*	-,012*		-,010*	-,010*	-,011*	-,010*
		Kolay	,053*	,034*	,028*	,013*	-,021*	-,010*	-,022*		-,016*	-,015*	-,016*	-,015*
		Çok kolay	-,007*	-,018*	-,047*	-,052*	0,000	-,003*	-,010*		-,017*	-,018*	-,018*	-,018*
	Orta güçlük	Kolay	,003*	-,024*	-,008*	-,038*	-,005*	,007*	-,010*		-,006*	-,005*	-,006*	-,005*
		Çok kolay	-,057*	-,076*	-,083*	-,104*	,015*	,013*	,002*		-,007*	-,007*	-,007*	-,009*
	Kolay	Çok kolay	-,060*	-,052*	-,075*	-,065*	,020*	,007*	,012*		-0,001	-,002*	-,001*	-,004*
Bifaktör	Çok zor	Zor	,122*	,073*	,048*	,074*	-,059*	-,035*	0,000		-,003*	-,011*	-,014*	-,004*
		Orta güçlük	,164*	,096*	,088*	,127*	-,070*	-,033*	-,015*		-,011*	-,022*	-,025*	-,010*
		Kolay	,158*	,110*	,109*	,093*	-,071*	-,048*	-,043*		-,023*	-,026*	-,029*	-,027*
		Çok kolay	,117*	,044*	,045*	,019*	-,060*	-,032*	-,038*		-,029*	-,041*	-,036*	-,040*
	Zor	Orta güçlük	,042*	,023*	,040*	,053*	-,011*	,003*	-,015*		-,008*	-,011*	-,011*	-,006*
		Kolay	,035*	,037*	,060*	,019*	-,012*	-,012*	-,044*		-,020*	-,016*	-,015*	-,023*
		Çok kolay	-,005*	-,028*	-,004*	-,055*	-0,001	,004*	-,038*		-,026*	-,030*	-,022*	-,036*
	Orta güçlük	Kolay	-,006*	,014*	,021*	-,033*	-0,001	-,015*	-,028*		-,012*	-,004*	-,004*	-,017*
		Çok kolay	-,047*	-,051*	-,043*	-,107*	,010*	0,001	-,023*		-,018*	-,019*	-,011*	-,030*
	Kolay	Çok kolay	-,040*	-,065*	-,064*	-,074*	,011*	,016*	,005*		-,006*	-,015*	-,007*	-,013*

*p<,001

Korelasyon 0,5 - Çoklu puanlanan madde yüzdesi düzeylerinde her bir test güçlüğü için modellerin ikili karşılaştırmaları

			RMSE				Güvenirlik				Sınıflama doğruluğu			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50	%5	%10	%25	%50
Test güçlüğü	(I) Model	(J) Model	Ortalama farkı (I-J)				Ortalama farkı (I-J)				Ortalama farkı (I-J)			
Çok zor	ÇBMTK	ÜDMTK	-,177*	-,183*	-,206*	-,265*	,032*	,042*	,063*	,064*	,073*	,077*	,084*	,104*
		Bifaktör	-,199*	-,192*	-,225*	-,299*	,050*	,049*	,076*	,077*	,069*	,083*	,088*	,109*
	ÜDMTK	Bifaktör	-,021*	-,009*	-,019*	-,034*	,018*	,007*	,014*	,013*	-,004*	,005*	,004*	,005*
Zor	ÇBMTK	ÜDMTK	-,180*	-,195*	-,234*	-,269*	,022*	,034*	,074*	,068*	,056*	,066*	,088*	,102*
		Bifaktör	-,185*	-,205*	-,285*	-,290*	,027*	,042*	,107*	,074*	,070*	,079*	,097*	,115*
	ÜDMTK	Bifaktör	-,005*	-,010*	-,051*	-,021*	,005*	,008*	,033*	,007*	,014*	,012*	,009*	,013*
Orta güçlük	ÇBMTK	ÜDMTK	-,184*	-,190*	-,241*	-,255*	,022*	,030*	,072*	,059*	,055*	,065*	,090*	,087*
		Bifaktör	-,198*	-,235*	-,288*	-,275*	,031*	,057*	,101*	,072*	,071*	,076*	,098*	,103*
	ÜDMTK	Bifaktör	-,013*	-,045*	-,048*	-,020*	,009*	,027*	,030*	,013*	,016*	,012*	,008*	,017*
Kolay	ÇBMTK	ÜDMTK	-,177*	-,207*	-,235*	-,268*	,016*	,036*	,058*	,058*	,053*	,069*	,087*	,078*
		Bifaktör	-,200*	-,214*	-,254*	-,282*	,030*	,041*	,070*	,062*	,064*	,081*	,097*	,083*
	ÜDMTK	Bifaktör	-,023*	-,007*	-,019*	-,015*	,014*	,005*	,012*	,004*	,011*	,012*	,010*	,005*
Çok kolay	ÇBMTK	ÜDMTK	-,187*	-,204*	-,241*	-,278*	,026*	,030*	,056*	,058*	,060*	,063*	,077*	,085*
		Bifaktör	-,190*	-,225*	-,250*	-,301*	,030*	,044*	,061*	,063*	,065*	,063*	,082*	,080*
	ÜDMTK	Bifaktör	-,003*	-,021*	-,008*	-,023*	,004*	,015*	,005*	,005*	,005*	0,000	,005*	-,005*

*p<,001

Ek 6. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (*r*=0,8)

Korelasyon 0,8 - Çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25) her bir yöntem için test güçlüklerinde hesaplanan ortalama RMSE değerleri

		Ortalama RMSE				Ortalama güvenirlik				Ortalama sınıflama doğruluğu			
Model	Test güçlüğü	Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
		%5	%10	%25	%50	%5	%10	%25	%50	%5	%10	%25	%50
ÇBMTK	Çok zor	,352	,328	,296	,231	,927	,935	,953	,976	,904	,909	,928	,954
	Zor	,252	,242	,191	,167	,960	,962	,978	,985	,931	,936	,952	,963
	Orta güçlük	,201	,196	,151	,123	,972	,973	,985	,990	,943	,943	,960	,966
	Kolay	,208	,208	,168	,157	,972	,972	,983	,987	,946	,950	,962	,968
	Çok kolay	,255	,256	,239	,219	,962	,962	,970	,978	,946	,948	,960	,968
ÜDMTK	Çok zor	,399	,376	,365	,323	,920	,927	,935	,963	,874	,880	,885	,901
	Zor	,300	,298	,284	,265	,956	,957	,958	,971	,898	,901	,904	,910
	Orta güçlük	,257	,253	,250	,223	,968	,968	,967	,977	,909	,909	,912	,916
	Kolay	,258	,271	,257	,253	,969	,966	,969	,975	,911	,912	,915	,918
	Çok kolay	,307	,313	,322	,313	,957	,957	,956	,967	,908	,910	,913	,918
Bifaktör	Çok zor	,400	,366	,369	,343	,920	,933	,934	,961	,878	,884	,886	,901
	Zor	,301	,295	,296	,279	,956	,958	,954	,970	,899	,903	,905	,910
	Orta güçlük	,260	,254	,261	,228	,967	,968	,964	,976	,910	,912	,915	,918
	Kolay	,260	,265	,262	,264	,968	,967	,967	,974	,916	,916	,920	,923
	Çok kolay	,307	,308	,324	,329	,956	,958	,956	,966	,915	,919	,920	,925

Korelasyon 0,8 - Çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25) her bir yöntem için test güçlüklerinin ikili karşılaştırmaları

			RMSE				Güvenirlik				Sınıflama doğruluğu			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50	%5	%10	%25	%50
Model	(I) test güçlüğü	(J) test güçlüğü	Ortalama farkı (I-J)				Ortalama farkı (I-J)				Ortalama farkı (I-J)			
ÇBMTK	Çok zor	Zor	,100*	,086*	,105*	,064*	-,032*	-,027*	-,025*	-,009*	-,027*	-,027*	-,025*	-,008*
		Orta güçlük	,151*	,132*	,144*	,107*	-,045*	-,038*	-,033*	-,014*	-,039*	-,034*	-,033*	-,012*
		Kolay	,144*	,120*	,128*	,074*	-,044*	-,037*	-,030*	-,010*	-,042*	-,041*	-,034*	-,013*
		Çok kolay	,097*	,072*	,056*	,012*	-,034*	-,027*	-,018*	-,002*	-,042*	-,039*	-,032*	-,013*
	Zor	Orta güçlük	,051*	,046*	,040*	,044*	-,012*	-,011*	-,007*	-,004*	-,012*	-,007*	-,008*	-,003*
		Kolay	,044*	,035*	,023*	,010*	-,012*	-,010*	-,005*	-,001*	-,015*	-,014*	-,010*	-,005*
		Çok kolay	-,003*	-,014*	-,048*	-,052*	-,002*	0,00	,008*	,007*	-,015*	-,012*	-,007*	-,005*
	Orta güçlük	Kolay	-,007*	-,011*	-,016*	-,034*	0,000	,001*	,002*	,003*	-,003*	-,007*	-,002*	-,002*
		Çok kolay	-,054*	-,060*	-,088*	-,096*	,010*	,011*	,015*	,012*	-,003*	-,005*	,001*	-,002*
	Kolay	Çok kolay	-,047*	-,048*	-,072*	-,062*	,010*	,010*	,013*	,009*	0,000	,002*	,003*	0,00
ÜDMTK	Çok zor	Zor	,098*	,078*	,082*	,058*	-,036*	-,029*	-,024*	-,008*	-,024*	-,020*	-,019*	-,009*
		Orta güçlük	,142*	,123*	,115*	,100*	-,048*	-,041*	-,032*	-,015*	-,035*	-,028*	-,027*	-,015*
		Kolay	,141*	,105*	,109*	,070*	-,049*	-,038*	-,034*	-,013*	-,038*	-,031*	-,031*	-,017*
		Çok kolay	,091*	,063*	,043*	,011*	-,036*	-,030*	-,021*	-,004*	-,034*	-,030*	-,028*	-,017*
	Zor	Orta güçlük	,043*	,045*	,033*	,042*	-,012*	-,012*	-,008*	-,006*	-,011*	-,008*	-,008*	-,006*
		Kolay	,043*	,028*	,027*	,012*	-,013*	-,009*	-,010*	-,004*	-,013*	-,011*	-,012*	-,008*
		Çok kolay	-,007*	-,015*	-,039*	-,047*	0,000	0,00	,002*	,004*	-,010*	-,009*	-,009*	-,007*
	Orta güçlük	Kolay	0,000	-,018*	-,007*	-,030*	-,001*	,003*	-,002*	,002*	-,002*	-,003*	-,004*	-,002*
		Çok kolay	-,050*	-,060*	-,072*	-,089*	,011*	,011*	,011*	,010*	,001*	-,001*	-,001*	-,001*
	Kolay	Çok kolay	-,050*	-,043*	-,066*	-,059*	,013*	,009*	,013*	,008*	,004*	,001*	,003*	,001
Bifaktör	Çok zor	Zor	,099*	,070*	,073*	,063*	-,036*	-,025*	-,021*	-,009*	-,021*	-,019*	-,019*	-,009*
		Orta güçlük	,139*	,112*	,108*	,114*	-,048*	-,035*	-,030*	-,014*	-,032*	-,028*	-,028*	-,016*
		Kolay	,139*	,100*	,107*	,079*	-,049*	-,034*	-,034*	-,013*	-,038*	-,032*	-,033*	-,022*
		Çok kolay	,093*	,057*	,044*	,014*	-,037*	-,025*	-,022*	-,005*	-,037*	-,036*	-,034*	-,024*
	Zor	Orta güçlük	,041*	,042*	,035*	,051*	-,011*	-,010*	-,009*	-,006*	-,011*	-,009*	-,010*	-,007*
		Kolay	,041*	,030*	,034*	,015*	-,013*	-,009*	-,013*	-,004*	-,017*	-,013*	-,014*	-,013*
		Çok kolay	-,006*	-,013*	-,028*	-,050*	-0,001	0,00	-,001*	,004*	-,017*	-,017*	-,015*	-,015*
	Orta güçlük	Kolay	-,000	-,012*	-0,001	-,036*	-,001*	,001*	-,004*	,001*	-,006*	-,004*	-,005*	-,005*
		Çok kolay	-,046*	-,054*	-,064*	-,101*	,011*	,010*	,008*	,009*	-,006*	-,008*	-,005*	-,008*
	Kolay	Çok kolay	-,046*	-,043*	-,062*	-,065*	,012*	,009*	,012*	,008*	0,000	-,004*	-0,001	-,002*

*p<,001

Korelasyon 0,8 - Çoklu puanlanan madde yüzdesi düzeylerinde her bir test güçlüğü için modellerin ikili karşılaştırmaları

			RMSE				Güvenirlik				Sınıflama doğruluğu			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50	%5	%10	%25	%50
Test güçlüğü	(I) Model	(J) Model	Ortalama farkı (I-J)				Ortalama farkı (I-J)				Ortalama farkı (I-J)			
Çok zor	ÇBMTK	ÜDMTK	-,047*	-,048*	-,070*	-,092*	,007*	,008*	,018*	,014*	,031*	,028*	,043*	,053*
		Bifaktör	-,047*	-,038*	-,073*	-,112*	,008*	,002*	,019*	,015*	,026*	,025*	,041*	,053*
	ÜDMTK	Bifaktör	-,001*	,010*	-,003*	-,019*	,001*	-,006*	,001*	,001*	-,004*	-,003*	-,002*	0,000
Zor	ÇBMTK	ÜDMTK	-,048*	-,056*	-,092*	-,098*	,003*	,006*	,019*	,014*	,033*	,035*	,049*	,052*
		Bifaktör	-,049*	-,053*	-,105*	-,112*	,004*	,004*	,024*	,015*	,032*	,033*	,047*	,052*
	ÜDMTK	Bifaktör	-0,001	,003*	-,012*	-,014*	,000*	-,001*	,004*	,001*	-,001*	-,002*	-,001*	0,000
Orta güçlük	ÇBMTK	ÜDMTK	-,056*	-,057*	-,099*	-,100*	,004*	,005*	,018*	,012*	,034*	,034*	,049*	,050*
		Bifaktör	-,059*	-,058*	-,109*	-,105*	,005*	,005*	,022*	,014*	,033*	,031*	,046*	,048*
	ÜDMTK	Bifaktör	-,003*	-0,001	-,010*	-,005*	,001*	0,00	,003*	,002*	-,001*	-,003*	-,003*	-,001*
Kolay	ÇBMTK	ÜDMTK	-,050*	-,063*	-,089*	-,096*	,002*	,007*	,014*	,011*	,035*	,038*	,047*	,049*
		Bifaktör	-,053*	-,058*	-,094*	-,107*	,003*	,005*	,016*	,012*	,030*	,034*	,043*	,045*
	ÜDMTK	Bifaktör	-,003*	,005*	-,005*	-,011*	,001*	-,001*	,002*	,001*	-,004*	-,004*	-,004*	-,005*
Çok kolay	ÇBMTK	ÜDMTK	-,053*	-,057*	-,083*	-,094*	,005*	,005*	,014*	,011*	,039*	,038*	,047*	,050*
		Bifaktör	-,052*	-,052*	-,085*	-,110*	,005*	,004*	,015*	,012*	,031*	,028*	,039*	,042*
	ÜDMTK	Bifaktör	,001*	,005*	-,002*	-,017*	,000*	-,001*	,001*	,001*	-,008*	-,009*	-,008*	-,008*

*p<,001

Ek 7. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,2; ÇPMY basit ana etkisi)

Tablo.. Toplam Test İçin Korelasyon 0,2 Olduğunda Test Güçlüğü Düzeylerinde Her Bir Yöntem İçin Çoklu Puanlanan Madde Yüzdelerinin Karşılaştırılması (RMSE)

		RMSE		Güvenirlik		Sınıflama doğruluğu	
Test güçlüğü	Model	F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	ÇBMTK	27755,550	,995	27623,293	,995	34473,325	,996
	ÜDMTK	1921,741	,936	197,755	,600	563,192	,810
	Bifaktör	2115,847	,941	1042,975	,888	3821,054	,967
Zor	ÇBMTK	24384,482	,995	39546,892	,997	76383,376	,998
	ÜDMTK	74,434	,361	2602,503	,952	182,287	,580
	Bifaktör	14285,968	,991	15900,797	,992	8371,196	,984
Orta güçlük	ÇBMTK	29197,949	,995	42701,827	,997	40069,914	,997
	ÜDMTK	1738,389	,929	4781,781	,973	202,588	,605
	Bifaktör	3980,536	,968	3799,204	,966	4093,148	,969
Kolay	ÇBMTK	13347,444	,990	29163,737	,995	43555,851	,997
	ÜDMTK	3504,816	,964	3821,623	,967	169,621	,562
	Bifaktör	2951,919	,957	2715,514	,954	1631,590	,925
Çok kolay	ÇBMTK	6037,286	,979	14881,128	,991	36320,350	,996
	ÜDMTK	4722,528	,973	1830,185	,933	289,492	,687
	Bifaktör	5648,160	,977	5061,623	,975	6209,063	,979

*p<,001

Toplam test için korelasyon 0,2 olduğunda her bir yöntem için çoklu puanlanan madde yüzdelerinde hesaplanan ortalama RMSE değerleri

		Ortalama RMSE					Ortalama güvenirlik					Ortalama sınıflama doğruluğu				
Model	Çoklu puanlanan madde yüzdesi	Test güçlüğü					Test güçlüğü					Test güçlüğü				
		Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
ÇBMTK	%5	,314	,211	,164	,164	,202	,897	,929	,944	,942	,934	,877	,869	,875	,880	,893
	%10	,273	,197	,149	,151	,192	,918	,944	,956	,955	,941	,884	,889	,891	,904	,900
	%25	,255	,161	,121	,128	,179	,939	,970	,978	,972	,957	,891	,917	,937	,942	,922
	%50	,185	,128	,098	,113	,155	,966	,978	,982	,979	,969	,939	,951	,927	,932	,949
ÜDMTK	%5	,605	,453	,366	,374	,471	,815	,864	,882	,893	,854	,722	,736	,741	,743	,738
	%10	,565	,450	,365	,405	,485	,811	,855	,870	,853	,853	,725	,737	,742	,743	,741
	%25	,571	,461	,414	,427	,525	,799	,796	,798	,816	,809	,726	,738	,743	,745	,742
	%50	,554	,459	,382	,444	,537	,821	,825	,846	,834	,824	,736	,742	,747	,747	,746
Bifaktör	%5	,794	,478	,499	,646	,491	,582	,831	,746	,635	,825	,830	,746	,765	,825	,785
	%10	,617	,513	,698	,434	,700	,739	,777	,571	,819	,614	,754	,751	,840	,767	,847
	%25	,714	,763	,729	,540	,547	,612	,486	,491	,696	,783	,779	,828	,837	,784	,788
	%50	,677	,500	,480	,466	,575	,666	,789	,744	,818	,789	,768	,753	,762	,790	,815

Toplam test için korelasyon 0,2 olduğunda her bir yöntem için çoklu puanlanan madde yüzdelerinin ikili karşılaştırmaları

Model	(I) ÇPMY	(J) ÇPMY	RMSE					Güvenirlik					Sınıflama doğruluğu				
			Test güçlüğü					Test güçlüğü					Test güçlüğü				
			Çok zor	Zor	Orta	Kolay	Çok kolay	Çok zor	Zor	Orta	Kolay	Çok kolay	Çok zor	Zor	Orta	Kolay	Çok kolay
			Ortalama fark (I-J)					Ortalama fark (I-J)					Ortalama fark (I-J)				
ÇBMTK	%5	%10	,041*	,014*	,015*	,013*	,010*	-,021*	-,015*	-,012*	-,013*	-,007*	-,007*	-,020*	-,016*	-,023*	-,007*
		%25	,059*	,050*	,043*	,036*	,023*	-,042*	-,041*	-,033*	-,031*	-,024*	-,014*	-,048*	-,062*	-,062*	-,029*
		%50	,129*	,083*	,066*	,050*	,047*	-,069*	-,049*	-,038*	-,037*	-,036*	-,062*	-,082*	-,052*	-,051*	-,055*
	%10	%25	,018*	,036*	,028*	,023*	,013*	-,021*	-,026*	-,022*	-,018*	-,017*	-,007*	-,028*	-,046*	-,039*	-,022*
		%50	,089*	,069*	,051*	,038*	,037*	-,049*	-,034*	-,026*	-,024*	-,029*	-,055*	-,062*	-,036*	-,028*	-,048*
	%25	%50	,071*	,033*	,023*	,015*	,024*	-,028*	-,008*	-,005*	-,006*	-,012*	-,048*	-,034*	,010*	,011*	-,026*
ÜDMTK	%5	%10	,039*	0,003	0,001	-,031*	-,014*	,004*	,009*	,012*	,040*	0,001	-,002*	-,001*	0,000	0,000	-,003*
		%25	,034*	-,008*	-,048*	-,053*	-,053*	,015*	,068*	,084*	,077*	,045*	-,004*	-,002*	-,001*	-,003*	-,004*
		%50	,050*	-,006*	-,015*	-,071*	-,066*	-,007*	,039*	,037*	,059*	,030*	-,014*	-,006*	-,005*	-,004*	-,008*
	%10	%25	-,006*	-,011*	-,049*	-,022*	-,039*	,012*	,059*	,072*	,037*	,044*	-,002*	-0,001	-,001*	-,002*	-0,001
		%50	,011*	-,008*	-,016*	-,039*	-,052*	-,010*	,029*	,024*	,019*	,029*	-,011*	-,005*	-,005*	-,004*	-,005*
	%25	%50	,017*	0,002	,033*	-,018*	-,013*	-,022*	-,030*	-,047*	-,018*	-,015*	-,010*	-,004*	-,004*	-,002*	-,004*
Bifaktör	%5	%10	,178*	-,035*	-,199*	,212*	-,209*	-,157*	,054*	,176*	-,184*	,211*	,076*	-,005*	-,075*	,058*	-,061*
		%25	,080*	-,285*	-,230*	,107*	-,055*	-,031*	,345*	,255*	-,061*	,042*	,051*	-,082*	-,072*	,041*	-,003*
		%50	,118*	-,022*	,019*	,180*	-,083*	-,085*	,042*	0,003	-,184*	,036*	,062*	-,007*	0,003	,036*	-,030*
	%10	%25	-,097*	-,250*	-,031*	-,105*	,154*	,126*	,291*	,080*	,123*	-,170*	-,026*	-,077*	0,003	-,017*	,059*
		%50	-,060*	,012*	,218*	-,032*	,126*	,072*	-,012*	-,173*	0,000	-,175*	-,014*	-0,002	,079*	-,023*	,032*
	%25	%50	,037*	,263*	,249*	,074*	-,028*	-,054*	-,303*	-,253*	-,122*	-0,005	,011*	,075*	,075*	-,006*	-,027*

*p<,001

Ek 8. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,5; ÇPMY basit ana etkisi)

Toplam test için korelasyon 0,5 olduğunda test güçlüğü düzeylerinde her bir yöntem için çoklu puanlanan madde yüzdelerinin karşılaştırılması (RMSE)

		RMSE		Güvenirlik		Sınıflama doğruluğu	
Test güçlüğü	Model	F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	ÇBMTK	24282,885	,995	22681,308*	,994	27644,584	,995
	ÜDMTK	1894,056	,935	1780,389*	,931	1632,088	,925
	Bifaktör	1487,072	,918	1459,279*	,917	1169,776	,899
Zor	ÇBMTK	19879,531	,993	27070,562*	,995	32131,040	,996
	ÜDMTK	34,820	,209	2533,382*	,950	592,207	,818
	Bifaktör	1503,911	,919	5590,055*	,977	833,554	,863
Orta güçlük	ÇBMTK	25764,999	,995	28308,705*	,995	31864,830	,996
	ÜDMTK	768,041	,853	5164,326*	,975	593,002	,818
	Bifaktör	1526,938	,920	3937,225*	,968	1093,831	,892
Kolay	ÇBMTK	7834,637	,983	2,263*	-	15301,475	,991
	ÜDMTK	2834,569	,956	8,042*	,057	42,929	,245
	Bifaktör	1635,826	,925	5,757*	-	213,463	,618
Çok kolay	ÇBMTK	3277,725	,961	9943,820*	,987	18867,980	,993
	ÜDMTK	4032,106	,968	1600,858*	,924	960,924	,879
	Bifaktör	3846,638	,967	745,621*	,850	3184,025	,960

*p<,001

Toplam test için korelasyon 0,5 olduğunda her bir yöntem için çoklu puanlanan madde yüzdelerinde hesaplanan ortalama RMSE değerleri

		Ortalama RMSE					Ortalama güvenirlik					Ortalama sınıflama doğruluğu				
		Test güçlüğü					Test güçlüğü					Test güçlüğü				
Model	Çoklu puanlanan madde yüzdesi	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
ÇBMTK	%5	,343	,235	,181	,185	,235	,911	,946	,962	,961	,951	,898	,903	,912	,916	,924
	%10	,303	,217	,165	,172	,226	,928	,956	,969	,968	,955	,907	,914	,923	,932	,929
	%25	,287	,178	,135	,149	,217	,942	,973	,982	,979	,965	,915	,938	,950	,953	,944
	%50	,213	,148	,110	,136	,192	,969	,981	,987	,974	,974	,947	,956	,950	,947	,956
ÜDMTK	%5	,521	,415	,365	,362	,422	,879	,924	,940	,945	,925	,825	,846	,857	,863	,864
	%10	,487	,412	,355	,379	,430	,887	,922	,939	,932	,926	,830	,848	,858	,863	,866
	%25	,493	,412	,376	,384	,459	,879	,899	,910	,920	,908	,831	,850	,860	,866	,868
	%50	,478	,417	,366	,404	,469	,905	,913	,928	,916	,916	,842	,854	,863	,868	,872
Bifaktör	%5	,542	,420	,378	,385	,425	,861	,920	,931	,932	,921	,829	,832	,840	,852	,858
	%10	,495	,423	,400	,386	,451	,879	,914	,912	,927	,911	,825	,836	,847	,851	,866
	%25	,512	,463	,424	,403	,467	,866	,865	,880	,909	,904	,827	,841	,852	,856	,863
	%50	,512	,438	,385	,419	,492	,893	,906	,915	,912	,911	,837	,841	,847	,864	,877

Toplam test için korelasyon 0,5 olduğunda her bir yöntem için çoklu puanlanan madde yüzdelerinin ikili karşılaştırmaları

			RMSE					Güvenirlik					Sınıflama doğruluğu				
			Test güçlüğü					Test güçlüğü					Test güçlüğü				
			Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
Model	(I) Çoklu puanlanan madde yüzdesi	(J) Çoklu puanlanan madde yüzdesi	Ortalama fark (I-J)					Ortalama fark (I-J)					Ortalama fark (I-J)				
ÇBMTK	%5	%10	,040*	,017*	,016*	,013*	,009*	-,018*	-,010*	-,007*		-,004*	-,010*	-,012*	-,011*	-,016*	-,005*
		%25	,057*	,056*	,046*	,036*	,018*	-,031*	-,026*	-,020*		-,014*	-,017*	-,035*	-,038*	-,037*	-,021*
		%50	,131*	,087*	,071*	,049*	,043*	-,059*	-,035*	-,025*		-,024*	-,049*	-,053*	-,038*	-,031*	-,033*
	%10	%25	,017*	,039*	,030*	,023*	,009*	-,014*	-,017*	-,013*		-,010*	-,007*	-,024*	-,027*	-,021*	-,016*
		%50	,091*	,069*	,055*	,035*	,034*	-,041*	-,025*	-,017*		-,019*	-,039*	-,041*	-,027*	-,015*	-,028*
	%25	%50	,074*	,031*	,025*	,013*	,026*	-,027*	-,008*	-,005*		-,010*	-,032*	-,018*	0,00	,006*	-,012*
ÜDMTK	%5	%10	,034*	,002*	,010*	-,017*	-,008*	-,008*	,002*	0,001	,013	0,00	-,005*	-,002*	-,002*	-0,001	-,002*
		%25	,028*	,003*	-,011*	-,022*	-,036*	-0,001	,026*	,029*	,025	,016*	-,006*	-,003*	-,003*	-,003*	-,004*
		%50	,043*	-,002*	-0,001	-,042*	-,047*	-,027*	,011*	,012*	,030*	,008*	-,017*	-,007*	-,007*	-,006*	-,008*
	%10	%25	-,006*	0,000	-,021*	-,005*	-,028*	,007*	,024*	,028*	,012	,017*	-0,001	-,002*	-,002*	-,003*	-,002*
		%50	,009*	-,004*	-,011*	-,025*	-,039*	-,019*	,009*	,011*	,017	,009*	-,012*	-,006*	-,005*	-,005*	-,006*
	%25	%50	,015*	-,005*	,010*	-,020*	-,011*	-,026*	-,014*	-,017*	,005	-,008*	-,011*	-,004*	-,003*	-,002*	-,004*
Bifaktör	%5	%10	,047*	-0,003	-,021*	-0,001	-,026*	-,018*	,005*	,019*		,010*	,004*	-,003*	-,007*	0,001	-,008*
		%25	,031*	-,043*	-,045*	-,018*	-,042*	-,005*	,054*	,050*		,017*	,002*	-,008*	-,012*	-,004*	-,004*
		%50	,030*	-,018*	-,007*	-,034*	-,067*	-,032*	,013*	,016*		,010*	-,008*	-,008*	-,006*	-,012*	-,018*
	%10	%25	-,016*	-,041*	-,024*	-,017*	-,016*	,014*	,049*	,031*		,007*	-,002*	-,005*	-,005*	-,005*	,003*
		%50	-,017*	-,015*	,015*	-,033*	-,041*	-,014*	,008*	-,003*		0,00	-,012*	-,005*	0,00	-,013*	-,011*
	%25	%50	0,000	,026*	,038*	-,016*	-,025*	-,027*	-,041*	-,034*		-,007*	-,010*	0,00	,005*	-,008*	-,014*

*p<,001

Ek 9. Toplam Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,8; ÇPMY basit ana etkisi)

Toplam test için korelasyon 0,8 olduğunda test güçlüğü düzeylerinde her bir yöntem için çoklu puanlanan madde yüzdelerinin karşılaştırılması (RMSE)

		RMSE		Güvenirlik		Sınıflama doğruluğu	
Test güçlüğü	Model	F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	ÇBMTK	19570,400	,993	24326,535	,995	30445,921	,996
	ÜDMTK	5666,071	,977	12537,178	,990	6485,727	,980
	Bifaktör	2905,749	,957	11080,537	,988	4586,080	,972
Zor	ÇBMTK	18277,130	,993	26050,192	,995	21194,643	,994
	ÜDMTK	2260,588	,945	4817,214	,973	2287,358	,945
	Bifaktör	680,488	,838	4898,244	,974	1666,256	,927
Orta güçlük	ÇBMTK	26190,355	,995	30685,655	,996	14833,514	,991
	ÜDMTK	2524,626	,950	3256,622	,961	1275,037	,906
	Bifaktör	2216,022	,944	3047,682	,958	1091,301	,892
Kolay	ÇBMTK	8460,816	,985	17296,084	,992	12517,044	,990
	ÜDMTK	486,623	,787	2597,578	,952	1004,780	,884
	Bifaktör	42,143	,242	1900,063	,935	1096,942	,893
Çok kolay	ÇBMTK	2856,610	,956	9880,644	,987	10911,733	,988
	ÜDMTK	286,024	,684	3161,870	,960	1687,812	,927
	Bifaktör	988,615	,882	2649,009	,953	1288,110	,907

*p<,001

Toplam test için korelasyon 0,8 olduğunda her bir yöntem, çoklu puanlanan madde yüzdesi ve test güçlüğü düzeylerinde elde edilen ortalama RMSE değerleri

		Ortalama RMSE					Ortalama güvenirlik					Ortalama sınıflama doğruluğu				
		Test güçlüğü					Test güçlüğü					Test güçlüğü				
Model	Çoklu puanlanan madde yüzdesi	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
ÇBMTK	%5	,352	,252	,201	,208	,255	,927	,960	,972	,972	,962	,904	,931	,943	,946	,946
	%10	,328	,242	,196	,208	,256	,935	,962	,973	,972	,962	,909	,936	,943	,950	,948
	%25	,296	,191	,151	,168	,239	,953	,978	,985	,983	,970	,928	,952	,960	,962	,960
	%50	,231	,167	,123	,157	,219	,976	,985	,990	,987	,978	,954	,963	,966	,968	,968
ÜDMTK	%5	,399	,300	,257	,258	,307	,920	,956	,968	,969	,957	,874	,898	,909	,911	,908
	%10	,376	,298	,253	,271	,313	,927	,957	,968	,966	,957	,880	,901	,909	,912	,910
	%25	,365	,284	,250	,257	,322	,935	,958	,967	,969	,956	,885	,904	,912	,915	,913
	%50	,323	,265	,223	,253	,313	,963	,971	,977	,975	,967	,901	,910	,916	,918	,918
Bifaktör	%5	,400	,301	,260	,260	,307	,920	,956	,967	,968	,956	,878	,899	,910	,916	,915
	%10	,366	,295	,254	,265	,308	,933	,958	,968	,967	,958	,884	,903	,912	,916	,919
	%25	,369	,296	,261	,262	,324	,934	,954	,964	,967	,956	,886	,905	,915	,920	,920
	%50	,343	,279	,228	,264	,329	,961	,970	,976	,974	,966	,901	,910	,918	,923	,925

Toplam test için korelasyon 0,8 olduğunda her bir yöntem için çoklu puanlanan madde yüzdelerinin ikili karşılaştırmaları

			RMSE					Güvenirlik					Sınıflama doğruluğu				
			Test güçlüğü					Test güçlüğü					Test güçlüğü				
			Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
Model	(I) ÇPMY	(J) ÇPMY	Ortalama fark (I-J)					Ortalama fark (I-J)					Ortalama fark (I-J)				
ÇBMTK	%5	%10	,024*	,010*	,005*	0,000	-0,001	-,008*	-,003*	-,001*	-,001*	,000*	-,004*	-,005*	0,00	-,004*	-,002*
		%25	,056*	,061*	,050*	,040*	,015*	-,025*	-,018*	-,013*	-,011*	-,009*	-,023*	-,022*	-,017*	-,016*	-,013*
		%50	,121*	,085*	,077*	,051*	,036*	-,049*	-,026*	-,018*	-,015*	-,016*	-,050*	-,032*	-,023*	-,022*	-,022*
	%10	%25	,032*	,051*	,045*	,040*	,017*	-,018*	-,016*	-,012*	-,011*	-,008*	-,019*	-,017*	-,017*	-,013*	-,012*
		%50	,097*	,075*	,073*	,050*	,037*	-,041*	-,023*	-,017*	-,014*	-,016*	-,045*	-,027*	-,023*	-,018*	-,020*
	%25	%50	,065*	,024*	,028*	,010*	,020*	-,024*	-,007*	-,005*	-,004*	-,008*	-,027*	-,010*	-,006*	-,006*	-,008*
ÜDMTK	%5	%10	,023*	,002*	,004*	-,013*	-,006*	-,007*	0,000	0,00	,004*	,000	-,007*	-,003*	0,00	0,00	-,003*
		%25	,033*	,017*	,007*	0,001	-,015*	-,015*	-,002	,001	,001*	,000	-,011*	-,006*	-,003*	-,004*	-,005*
		%50	,075*	,035*	,034*	,004*	-,005*	-,042*	-,015*	-,009*	-,006*	-,010*	-,027*	-,012*	-,007*	-,007*	-,010*
	%10	%25	,010*	,014*	,003*	,014*	-,009*	-,008*	-,002	,001	-,003*	,001	-,004*	-,003*	-,003*	-,004*	-,002*
		%50	,053*	,033*	,030*	,017*	0,001	-,035*	-,014*	-,009*	-,010*	-,010*	-,021*	-,009*	-,007*	-,007*	-,007*
	%25	%50	,042*	,018*	,027*	,003*	,010*	-,028*	-,012*	-,010*	-,006*	-,011*	-,016*	-,007*	-,005*	-,003*	-,005*
Bifaktör	%5	%10	,034*	,006*	,007*	-,005*	-0,002	-,013*	-,002*	-,001*	,001*	-,002*	-,006*	-,004*	-,002*	0,00	-,004*
		%25	,031*	,005*	0,000	-0,002	-,018*	-,014*	,001*	,004*	,001*	,001*	-,008*	-,006*	-,005*	-,004*	-,005*
		%50	,057*	,022*	,032*	-,004*	-,023*	-,042*	-,014*	-,009*	-,006*	-,010*	-,023*	-,011*	-,008*	-,007*	-,010*
	%10	%25	-,003*	-0,001	-,007*	,003*	-,016*	-0,001	,004*	,005*	0,000	,003*	-,003*	-,003*	-,003*	-,004*	-,001*
		%50	,023*	,016*	,025*	0,001	-,021*	-,028*	-,012*	-,008*	-,008*	-,008*	-,018*	-,008*	-,006*	-,007*	-,006*
	%25	%50	,026*	,017*	,032*	-,002*	-,005*	-,028*	-,016*	-,012*	-,007*	-,011*	-,015*	-,005*	-,003*	-,003*	-,005*

*p<,001

Ek 10. Toplam Test - Model x Test Güçlüğü x Korelasyon Etkileşimi (Korelasyon basit ana etkisi)

Toplam test için çoklu puanlanan madde yüzdesinin ve test güçlüğünün her bir düzeyi için model ve korelasyon değişkenleri ile iki yönlü karma ANOVA (RMSE)

			Çoklu puanlanan madde yüzdesi							
			%5		%10		%25		%50	
Test güçlüğü	Kaynak		F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	Gruplar içi	M	177950,757	,998	150304,012	,998	85016,501	,997	165790,517	,998
		M *x* K	47080,169	,997	29342,182	,995	15788,631	,991	21295,156	,993
	Gruplar arası	K	64310,931	,998	26429,056	,994	24927,680	,994	46238,872	,997
Zor	Gruplar içi	M	393101,207	,999	141301,770	,998	193866,052	,998	346432,107	,999
		M *x* K	54828,202	,997	20788,020	,993	38297,374	,996	30819,048	,995
	Gruplar arası	K	36531,464	,996	25363,419	,994	53079,642	,997	43508,535	,997
Orta güçlük	Gruplar içi	M	31860,483	,991	170092,047	,998	172660,926	,998	142767,285	,998
		M *x* K	5129,285	,972	48717,413	,997	33850,876	,996	13626,778	,989
	Gruplar arası	K	8149,820	,982	57868,832	,997	50587,987	,997	26330,455	,994
Kolay	Gruplar içi	M	76780,105	,996	333237,353	,999	73829,150	,996	449293,771	,999
		M *x* K	21793,642	,993	40399,772	,996	9026,516	,984	40117,188	,996
	Gruplar arası	K	26644,395	,994	21447,028	,993	14461,441	,990	48935,389	,997
Çok kolay	Gruplar içi	M	367123,279	,999	108232,144	,997	355856,125	,999	378407,723	,999
		M *x* K	53526,527	,997	27124,454	,995	42668,732	,997	41265,872	,996
	Gruplar arası	K	32863,936	,996	36560,636	,996	36291,105	,996	44290,634	,997

*p<,001

Toplam test için çoklu puanlanan madde yüzdesinin ve test güçlüğünün her bir düzeyi için model ve korelasyon değişkenleri ile iki yönlü karma ANOVA (güvenirlik)

			Çoklu puanlanan madde yüzdesi							
			%5		%10		%25		%50	
Test güçlüğü	Kaynak		F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	Gruplar içi	M	52086,570	,994	13760,678	,979	17430,882	,983	17657,756	,983
		M *x* K	33054,193	,996	6290,128	,977	8089,292	,982	7828,854	,981
	Gruplar arası	K	50285,818	,997	17381,408	,992	17431,576	,992	23556,778	,994
Zor	Gruplar içi	M	28901,059	,990	13291,212	,978	92626,550	,997	57995,922	,995
		M *x* K	12269,585	,988	6342,844	,977	46154,962	,997	16777,572	,991
	Gruplar arası	K	44233,545	,997	20283,512	,993	73520,168	,998	40563,789	,996
Orta güçlük	Gruplar içi	M	4548,093	,939	99054,893	,997	83202,147	,996	20163,258	,985
		M *x* K	2871,862	,951	66162,468	,998	43202,352	,997	7648,885	,981
	Gruplar arası	K	9122,698	,984	124411,339	,999	85495,905	,998	22485,138	,993
Kolay	Gruplar içi	M	31544,623	,991	32219,305	,991	12605,152	,977	43507,489	,993
		M *x* K	24261,582	,994	12626,318	,988	5361,636	,973	13352,906	,989
	Gruplar arası	K	49011,419	,997	33780,007	,996	15202,433	,990	86,569	,368
Çok kolay	Gruplar içi	M	26412,501	,989	28302,817	,990	34048,997	,991	45871,317	,994
		M *x* K	10814,420	,986	19658,554	,993	10668,044	,986	15192,259	,990
	Gruplar arası	K	33692,464	,996	40741,735	,996	28861,701	,995	36706,240	,996

*p<,001

Toplam test için çoklu puanlanan madde yüzdesinin ve test güçlüğünün her bir düzeyi için model ve korelasyon değişkenleri ile iki yönlü karma ANOVA (sınıflama doğruluğu)

			Çoklu puanlanan madde yüzdesi							
			%5		%10		%25		%50	
Test güçlüğü	Kaynak		F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	Gruplar içi	M	89079,909	,997	229470,869	,999	113181,522	,997	483387,741	,999
		M *x* K	21321,698	,993	40415,363	,996	14312,716	,990	57900,497	,997
	Gruplar arası	K	53747,189	,997	121567,731	,999	78093,774	,998	182930,834	,999
Zor	Gruplar içi	M	406749,310	,999	350627,579	,999	136077,363	,998	1066399,354	1,000
		M *x* K	61084,831	,998	54790,766	,997	20602,283	,993	143920,046	,999
	Gruplar arası	K	201615,299	,999	247996,882	,999	80378,820	,998	359172,857	1,000
Orta güçlük	Gruplar içi	M	82207,162	,996	47078,699	,994	175618,026	,998	668053,126	1,000
		M *x* K	12014,297	,988	11422,032	,987	28922,871	,995	81728,889	,998
	Gruplar arası	K	155652,804	,999	57670,555	,997	106030,892	,999	439163,228	1,000
Kolay	Gruplar içi	M	30253,658	,990	719541,906	1,000	404024,589	,999	195567,029	,998
		M *x* K	6701,281	,978	113478,779	,999	61263,847	,998	27597,841	,995
	Gruplar arası	K	52913,829	,997	355636,699	1,000	244003,177	,999	29679,389	,995
Çok kolay	Gruplar içi	M	508449,822	,999	93692,256	,997	659854,252	1,000	864048,812	1,000
		M *x* K	85526,253	,998	21745,090	,993	97408,508	,998	137373,895	,999
	Gruplar arası	K	237367,770	,999	116342,682	,999	312045,634	1,000	357308,709	1,000

*p<,001

Toplam test için çoklu puanlanan madde yüzdesinin ve test güçlüğünün bütün düzeylerinde her bir yöntem için korelasyon düzeylerinin karşılaştırılması (RMSE)

		Çoklu puanlanan madde yüzdesi							
		%5		%10		%25		%50	
Test güçlüğü	Model	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	ÇBMTK	3246,917	,956	5665,933	,974	3976,001	,964	4292,621	,967
	ÜDMTK	72361,740	,998	44014,811	,997	34964,064	,996	95266,125	,998
	Bifaktör	60173,593	,998	27162,164	,995	17796,128	,992	24465,299	,994
Zor	ÇBMTK	5483,314	,974	8283,038	,982	3378,889	,958	4330,234	,967
	ÜDMTK	54830,981	,997	48591,798	,997	23217,617	,994	59851,458	,998
	Bifaktör	54768,849	,997	18921,742	,992	54530,091	,997	33688,925	,996
Orta güçlük	ÇBMTK	6197,849	,977	13378,722	,989	6456,052	,978	6350,341	,977
	ÜDMTK	33144,169	,996	30323,028	,995	26427,760	,994	59841,793	,998
	Bifaktör	5028,725	,971	56818,034	,997	44024,073	,997	14271,248	,990
Kolay	ÇBMTK	8855,062	,984	12357,410	,988	7265,669	,980	7174,132	,980
	ÜDMTK	42895,448	,997	32497,615	,995	36003,966	,996	56244,327	,997
	Bifaktör	22911,510	,994	32414,498	,995	8815,386	,983	45777,580	,997
Çok kolay	ÇBMTK	8482,340	,983	11743,335	,988	10254,636	,986	9753,785	,985
	ÜDMTK	57807,747	,997	55421,251	,997	60531,172	,998	71360,991	,998
	Bifaktör	45369,113	,997	29192,183	,995	36290,367	,996	38251,170	,996

*p<,001

Toplam test için çoklu puanlanan madde yüzdesinin ve test güçlüğünün bütün düzeylerinde her bir yöntem için korelasyon düzeylerinin karşılaştırılması (güvenirlik)

		Çoklu puanlanan madde yüzdesi							
		%5		%10		%25		%50	
Test güçlüğü	Model	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	ÇBMTK	5518,010	,974	2185,748	,936	2530,466	,945	3161,224	,955
	ÜDMTK	23580,770	,994	18922,935	,992	14036,531	,990	56576,633	,997
	Bifaktör	46105,678	,997	10121,884	,986	11581,697	,987	11333,244	,987
Zor	ÇBMTK	13617,710	,989	7635,975	,981	3125,514	,955	3998,778	,964
	ÜDMTK	36334,789	,996	31166,320	,995	20913,054	,993	51811,835	,997
	Bifaktör	28820,320	,995	10192,568	,986	71973,713	,998	21915,380	,993
Orta güçlük	ÇBMTK	18498,203	,992	14193,446	,990	6344,191	,977	13285,933	,989
	ÜDMTK	33515,057	,996	32873,250	,996	32513,966	,995	53982,415	,997
	Bifaktör	4521,347	,968	96586,603	,998	65468,462	,998	10729,900	,986
Kolay	ÇBMTK	23430,809	,994	9795,797	,985	8544,062	,983	1,302	-
	ÜDMTK	47422,189	,997	30124,572	,995	34058,022	,996	174,486	,540
	Bifaktör	32729,020	,995	21264,144	,993	7519,823	,981	214,456	,591
Çok kolay	ÇBMTK	11948,068	,988	8672,080	,983	4454,176	,968	2735,048	,949
	ÜDMTK	31113,017	,995	36418,671	,996	34889,355	,996	54332,600	,997
	Bifaktör	21294,484	,993	27179,546	,995	14845,759	,990	19375,681	,992

*p<,001

Toplam test için çoklu puanlanan madde yüzdesinin ve test güçlüğünün bütün düzeylerinde her bir yöntem için korelasyon düzeylerinin karşılaştırılması (sınıflama doğruluğu)

		Çoklu puanlanan madde yüzdesi							
		%5		%10		%25		%50	
Test güçlüğü	Model	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	ÇBMTK	8837,764	,983	8306,915	,982	18966,261	,992	5632,749	,974
	ÜDMTK	118960,753	,999	143569,597	,999	174779,516	,999	276775,003	,999
	Bifaktör	5927,723	,976	67261,214	,998	15531,739	,991	68041,591	,998
Zor	ÇBMTK	45875,915	,997	35523,864	,996	25102,674	,994	3076,571	,954
	ÜDMTK	218398,507	,999	339757,157	1,000	289197,995	,999	457541,546	1,000
	Bifaktör	167805,764	,999	118421,261	,999	9834,086	,985	221816,949	,999
Orta güçlük	ÇBMTK	64567,831	,998	41069,520	,996	14386,556	,990	30960,548	,995
	ÜDMTK	335078,171	1,000	415561,857	1,000	437521,249	1,000	591485,745	1,000
	Bifaktör	31157,370	,995	5014,767	,971	11875,009	,988	180670,721	,999
Kolay	ÇBMTK	68193,425	,998	35722,351	,996	9166,996	,984	19077,927	,992
	ÜDMTK	401391,894	1,000	463164,903	1,000	476709,127	1,000	37866,662	,996
	Bifaktör	5099,324	,972	261726,660	,999	70687,104	,998	21502,275	,993
Çok kolay	ÇBMTK	42034,536	,996	41303,722	,996	28963,052	,995	10213,557	,986
	ÜDMTK	308888,945	1,000	318104,433	1,000	390957,992	1,000	590595,999	1,000
	Bifaktör	157966,794	,999	9840,793	,985	163528,182	,999	131080,810	,999

*p<,001

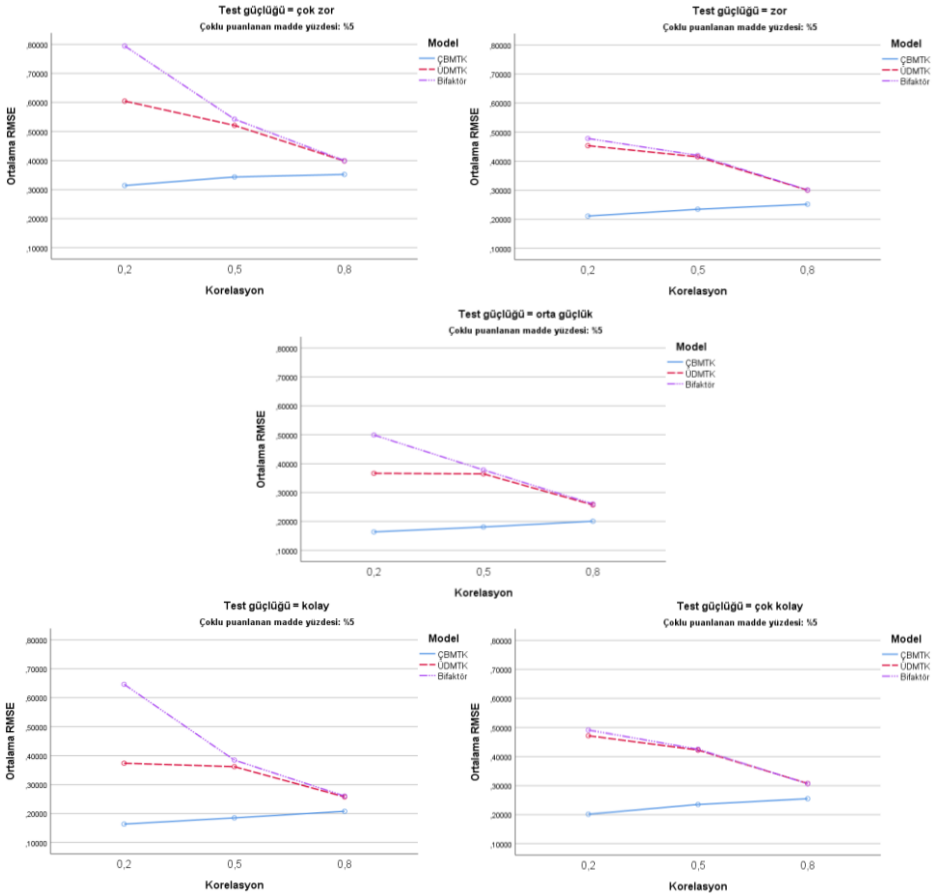
Toplam test için çoklu puanlanan madde yüzdesi, test güçlüğü, model ve korelasyon düzeylerinde elde edilen ortalama RMSE, güvenirlik ve sınıflama doğruluğu değerleri

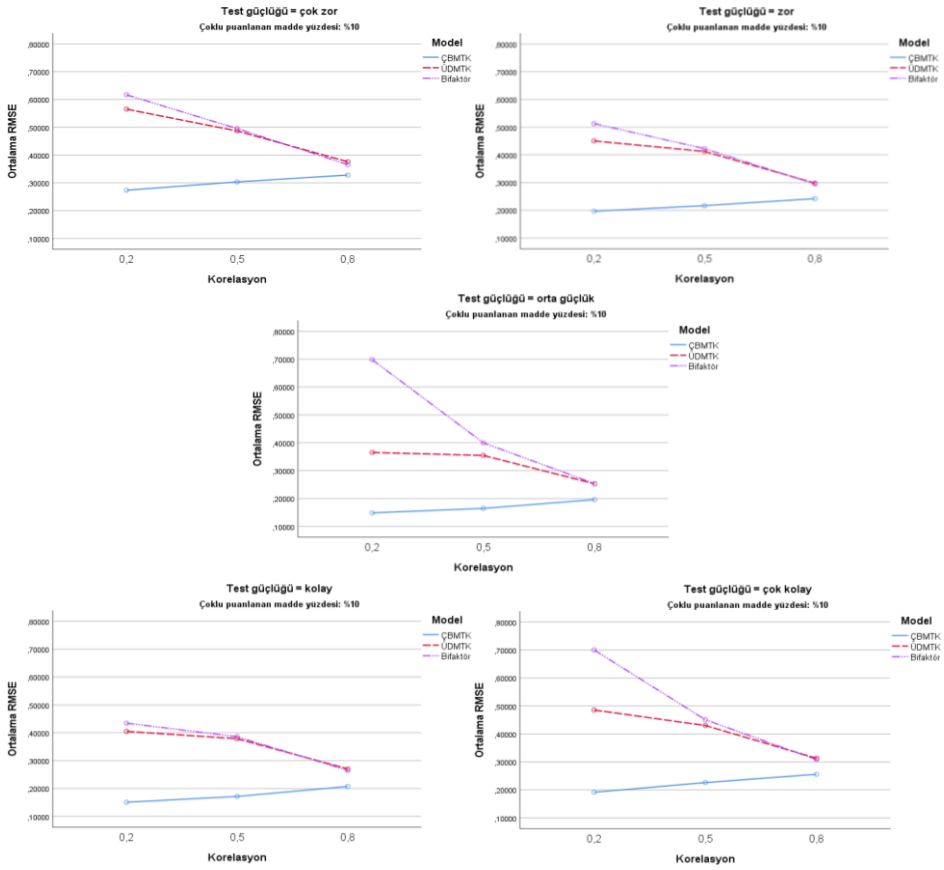
			Ortalama RMSE					Ortalama güvenirlik					Ortalama sınıflama doğruluğu				
			Test güçlüğü					Test güçlüğü					Test güçlüğü				
ÇPMY	Model	Korelasyon	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
%5	ÇBMTK	0,2	,314	,211	,164	,164	,202	,897	,929	,944	,942	,934	,877	,869	,875	,880	,893
		0,5	,343	,235	,181	,185	,235	,911	,946	,962	,961	,951	,898	,903	,912	,916	,924
		0,8	,352	,252	,201	,208	,255	,927	,960	,972	,972	,962	,904	,931	,943	,946	,946
	ÜDMTK	0,2	,605	,453	,366	,374	,471	,815	,864	,882	,893	,854	,722	,736	,741	,743	,738
		0,5	,521	,415	,365	,362	,422	,879	,924	,940	,945	,925	,825	,846	,857	,863	,864
		0,8	,399	,300	,257	,258	,307	,920	,956	,968	,969	,957	,874	,898	,909	,911	,908
	Bifaktör	0,2	,794	,478	,499	,646	,491	,582	,831	,746	,635	,825	,830	,746	,765	,825	,785
		0,5	,542	,420	,378	,385	,425	,861	,920	,931	,932	,921	,829	,832	,840	,852	,858
		0,8	,400	,301	,260	,260	,307	,920	,956	,967	,968	,956	,878	,899	,910	,916	,915
%10	ÇBMTK	0,2	,273	,197	,197	,149	,151	,918	,944	,956	,955	,941	,884	,889	,891	,904	,900
		0,5	,303	,217	,165	,172	,192	,928	,956	,969	,968	,955	,907	,914	,923	,932	,929
		0,8	,328	,242	,196	,208	,226	,935	,962	,973	,972	,962	,909	,936	,943	,950	,948
	ÜDMTK	0,2	,565	,450	,365	,405	,256	,811	,855	,870	,853	,853	,725	,737	,742	,743	,741
		0,5	,487	,412	,355	,379	,485	,887	,922	,939	,932	,926	,830	,848	,858	,863	,866
		0,8	,376	,298	,253	,271	,430	,927	,957	,968	,966	,957	,880	,901	,909	,912	,910
	Bifaktör	0,2	,617	,513	,698	,434	,313	,739	,777	,571	,819	,614	,754	,751	,840	,767	,847
		0,5	,495	,423	,400	,386	,700	,879	,914	,912	,927	,911	,825	,836	,847	,851	,866
		0,8	,366	,295	,254	,265	,451	,933	,958	,968	,967	,958	,884	,903	,912	,916	,919
%25	ÇBMTK	0,2	,255	,161	,121	,128	,179	,939	,970	,978	,972	,957	,891	,917	,937	,942	,922
		0,5	,287	,178	,135	,149	,217	,942	,973	,982	,979	,965	,915	,938	,950	,953	,944
		0,8	,296	,191	,151	,168	,239	,953	,978	,985	,983	,970	,928	,952	,960	,962	,960
	ÜDMTK	0,2	,571	,461	,414	,427	,525	,799	,796	,798	,816	,809	,726	,738	,743	,745	,742
		0,5	,493	,412	,376	,384	,459	,879	,899	,910	,920	,908	,831	,850	,860	,866	,868
		0,8	,365	,284	,250	,257	,322	,935	,958	,967	,969	,956	,885	,904	,912	,915	,913
	Bifaktör	0,2	,714	,763	,729	,540	,547	,612	,486	,491	,696	,783	,779	,828	,837	,784	,788
		0,5	,512	,463	,424	,403	,467	,866	,865	,880	,909	,904	,827	,841	,852	,856	,863
		0,8	,369	,296	,261	,262	,324	,934	,954	,964	,967	,956	,886	,905	,915	,920	,920
%50	ÇBMTK	0,2	,185	,128	,098	,113	,155	,966	,978	,982	,979	,969	,939	,951	,927	,932	,949
		0,5	,213	,148	,110	,136	,192	,969	,981	,987	,974	,974	,947	,956	,950	,947	,956
		0,8	,231	,167	,123	,157	,219	,976	,985	,990	,987	,978	,954	,963	,966	,968	,968
	ÜDMTK	0,2	,554	,459	,382	,444	,537	,821	,825	,846	,834	,824	,736	,742	,747	,747	,746
		0,5	,478	,417	,366	,404	,469	,905	,913	,928	,916	,916	,842	,854	,863	,868	,872
		0,8	,323	,265	,223	,253	,313	,963	,971	,977	,975	,967	,901	,910	,916	,918	,918
	Bifaktör	0,2	,677	,500	,480	,466	,575	,666	,789	,744	,818	,789	,768	,753	,762	,790	,815
		0,5	,512	,438	,385	,419	,492	,893	,906	,915	,912	,911	,837	,841	,847	,864	,877
		0,8	,343	,279	,228	,264	,329	,961	,970	,976	,974	,966	,901	,910	,918	,923	,925

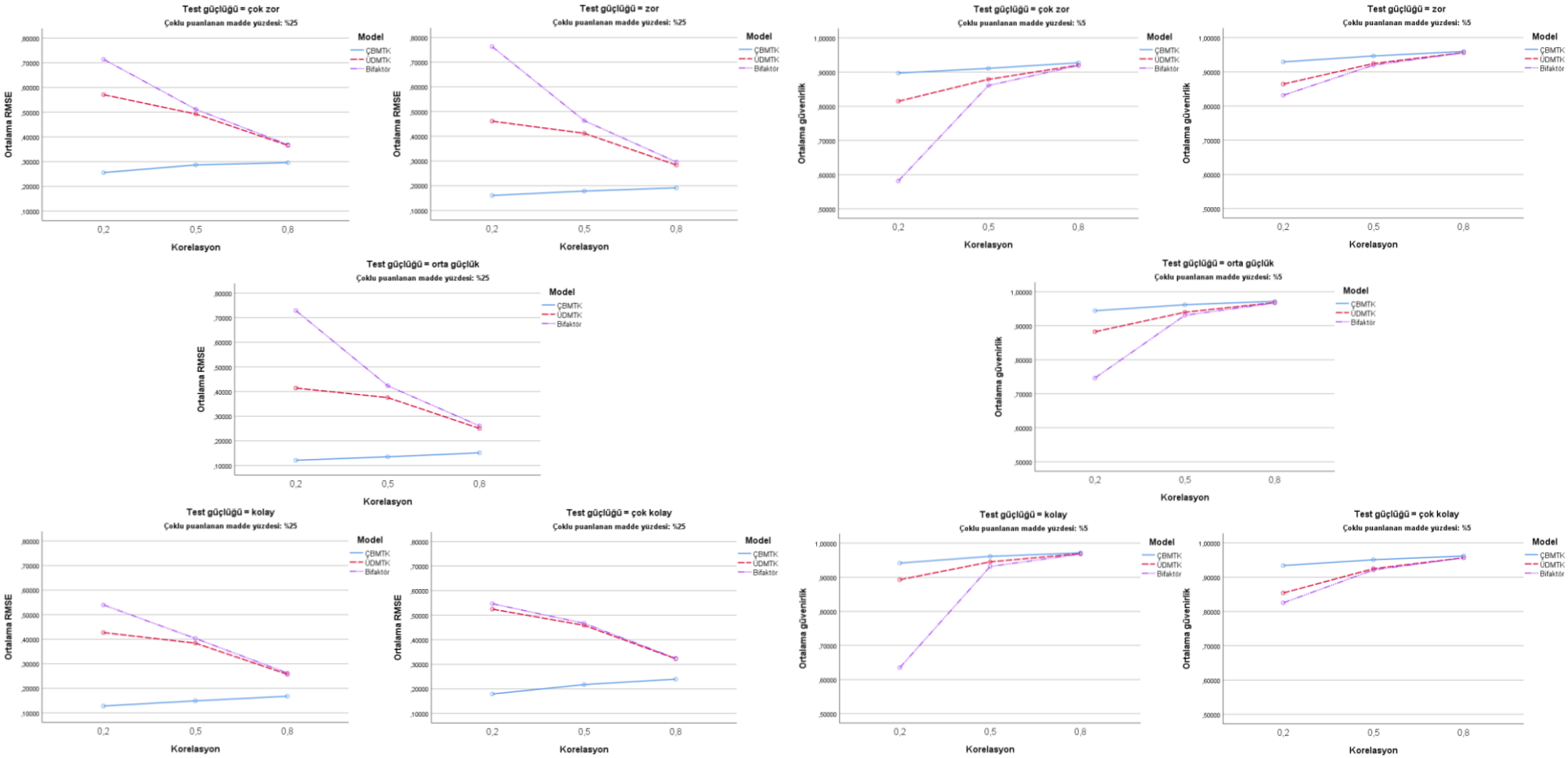
Toplam test için çoklu puanlanan madde yüzdesinin ve test güçlüğünün bütün düzeylerinde her bir yöntem için korelasyon düzeylerinin ikili karşılaştırmaları (RMSE)

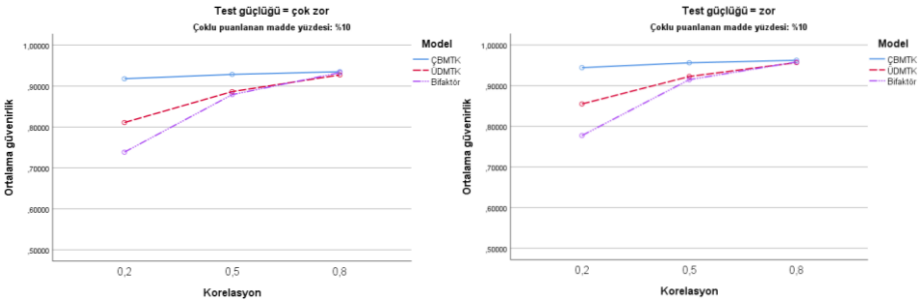
				RMSE					Güvenirlik					Sınıflama doğruluğu				
				Test güçlüğü					Test güçlüğü					Test güçlüğü				
				Çok zor	Zor	Orta	Kolay	Çok kolay	Çok zor	Zor	Orta	Kolay	Çok kolay	Çok zor	Zor	Orta	Kolay	Çok kolay
ÇPMY	Model	(I)K	(J)K	Ortalama fark (I-J)					Ortalama fark (I-J)					Ortalama fark (I-J)				
%5	ÇBMTK	0,2	0,5	-,029*	-,024*	-,017*	-,021*	-,033*	-,014*	-,017*	-,017*	-,020*	-,017*	-,021*	-,033*	-,037*	-,036*	-,030*
			0,8	-,038*	-,041*	-,037*	-,044*	-,053*	-,030*	-,031*	-,028*	-,030*	-,028*	-,027*	-,061*	-,068*	-,066*	-,053*
		0,5	0,8	-,009*	-,017*	-,020*	-,023*	-,020*	-,016*	-,013*	-,010*	-,010*	-,011*	-,007*	-,028*	-,031*	-,030*	-,022*
	ÜDMTK	0,2	0,5	,084*	,039*	0,002	,012*	,049*	-,064*	-,060*	-,057*	-,052*	-,071*	-,103*	-,111*	-,116*	-,120*	-,125*
			0,8	,206*	,153*	,109*	,116*	,164*	-,105*	-,092*	-,086*	-,077*	-,103*	-,151*	-,162*	-,168*	-,169*	-,169*
		0,5	0,8	,122*	,115*	,108*	,104*	,115*	-,042*	-,032*	-,028*	-,024*	-,032*	-,049*	-,052*	-,052*	-,049*	-,044*
	Bifaktör	0,2	0,5	,252*	,058*	,121*	,262*	,066*	-,279*	-,088*	-,185*	-,297*	-,096*	0,001	-,086*	-,075*	-,027*	-,073*
			0,8	,395*	,177*	,239*	,386*	,185*	-,338*	-,124*	-,221*	-,334*	-,131*	-,048*	-,153*	-,145*	-,090*	-,130*
		0,5	0,8	,143*	,119*	,118*	,124*	,119*	-,059*	-,036*	-,036*	-,037*	-,035*	-,049*	-,066*	-,069*	-,064*	-,057*
%10	ÇBMTK	0,2	0,5	-,030*	-,020*	-,016*	-,021*	-,035*	-,011*	-,012*	-,013*	-,013*	-,014*	-,023*	-,025*	-,032*	-,028*	-,028*
			0,8	-,055*	-,045*	-,047*	-,057*	-,064*	-,017*	-,018*	-,017*	-,017*	-,021*	-,025*	-,047*	-,052*	-,046*	-,047*
		0,5	0,8	-,025*	-,025*	-,031*	-,036*	-,030*	-,007*	-,006*	-,004*	-,004*	-,007*	-,001*	-,022*	-,020*	-,018*	-,019*
	ÜDMTK	0,2	0,5	,079*	,038*	,011*	,026*	,055*	-,076*	-,067*	-,069*	-,079*	-,073*	-,106*	-,111*	-,117*	-,120*	-,125*
			0,8	,189*	,152*	,112*	,134*	,172*	-,116*	-,102*	-,098*	-,112*	-,104*	-,156*	-,164*	-,167*	-,169*	-,169*
		0,5	0,8	,111*	,114*	,102*	,108*	,117*	-,041*	-,034*	-,029*	-,033*	-,032*	-,050*	-,053*	-,050*	-,048*	-,045*
	Bifaktör	0,2	0,5	,121*	,090*	,298*	,049*	,249*	-,141*	-,137*	-,341*	-,108*	-,297*	-,071*	-,085*	-,006*	-,084*	-,019*
			0,8	,251*	,218*	,444*	,169*	,392*	-,194*	-,181*	-,397*	-,148*	-,344*	-,130*	-,152*	-,071*	-,149*	-,072*
		0,5	0,8	,130*	,127*	,146*	,121*	,143*	-,054*	-,044*	-,056*	-,040*	-,047*	-,059*	-,067*	-,065*	-,065*	-,053*
%25	ÇBMTK	0,2	0,5	-,031*	-,018*	-,014*	-,021*	-,038*	-,003*	-,003*	-,004*	-,006*	-,007*	-,024*	-,021*	-,013*	-,010*	-,022*
			0,8	-,040*	-,031*	-,030*	-,040*	-,060*	-,014*	-,008*	-,007*	-,010*	-,013*	-,037*	-,035*	-,023*	-,020*	-,037*
		0,5	0,8	-,009*	-,013*	-,016*	-,019*	-,022*	-,011*	-,005*	-,003*	-,004*	-,006*	-,013*	-,015*	-,010*	-,009*	-,015*
	ÜDMTK	0,2	0,5	,078*	,049*	,039*	,043*	,066*	-,080*	-,103*	-,112*	-,104*	-,100*	-,105*	-,112*	-,118*	-,121*	-,126*
			0,8	,205*	,177*	,164*	,170*	,202*	-,136*	-,163*	-,169*	-,153*	-,148*	-,158*	-,166*	-,169*	-,170*	-,171*
		0,5	0,8	,127*	,128*	,125*	,127*	,136*	-,056*	-,060*	-,057*	-,048*	-,048*	-,054*	-,054*	-,051*	-,049*	-,045*
	Bifaktör	0,2	0,5	,202*	,300*	,305*	,137*	,080*	-,253*	-,379*	-,389*	-,213*	-,120*	-,048*	-,013*	-,015*	-,072*	-,074*
			0,8	,345*	,467*	,468*	,278*	,222*	-,321*	-,468*	-,472*	-,271*	-,172*	-,107*	-,078*	-,078*	-,136*	-,132*
		0,5	0,8	,143*	,167*	,163*	,141*	,143*	-,068*	-,089*	-,083*	-,058*	-,052*	-,059*	-,065*	-,063*	-,064*	-,057*
%50	ÇBMTK	0,2	0,5	-,028*	-,020*	-,012*	-,023*	-,037*	-,003*	-,003*	-,004*	,005	-,005*	-,008*	-,004*	-,024*	-,015*	-,008*
			0,8	-,046*	-,039*	-,026*	-,044*	-,064*	-,010*	-,007*	-,008*	-,008	-,009*	-,015*	-,011*	-,039*	-,036*	-,019*
		0,5	0,8	-,018*	-,019*	-,013*	-,021*	-,027*	-,007*	-,005*	-,003*	-,013	-,004*	-,008*	-,007*	-,016*	-,021*	-,011*
	ÜDMTK	0,2	0,5	,077*	,042*	,016*	,040*	,068*	-,084*	-,087*	-,082*	-,082*	-,092*	-,106*	-,112*	-,117*	-,121*	-,126*
			0,8	,231*	,194*	,159*	,191*	,225*	-,141*	-,145*	-,132*	-,141*	-,143*	-,165*	-,168*	-,170*	-,171*	-,172*
		0,5	0,8	,154*	,152*	,142*	,151*	,157*	-,057*	-,058*	-,050*	-,060*	-,051*	-,059*	-,056*	-,053*	-,050*	-,046*
	Bifaktör	0,2	0,5	,165*	,063*	,095*	,047*	,082*	-,226*	-,117*	-,171*	-,093*	-,122*	-,069*	-,088*	-,085*	-,074*	-,062*
			0,8	,334*	,221*	,252*	,202*	,246*	-,295*	-,181*	-,232*	-,156*	-,178*	-,133*	-,157*	-,156*	-,133*	-,110*
		0,5	0,8	,169*	,158*	,157*	,154*	,163*	-,069*	-,064*	-,061*	-,063*	-,055*	-,064*	-,069*	-,071*	-,059*	-,049*

*p<,001

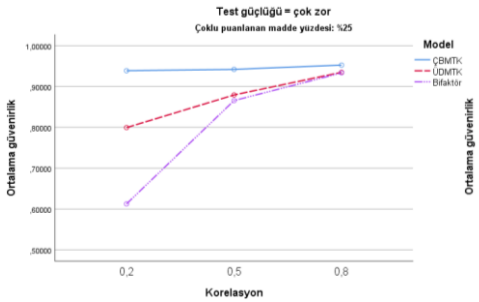

Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %5
Model
ÇBMTK
ÜDMTK
Bifaktör
Ortalama RMSE
Korelasyon
0,2
0,5
0,8

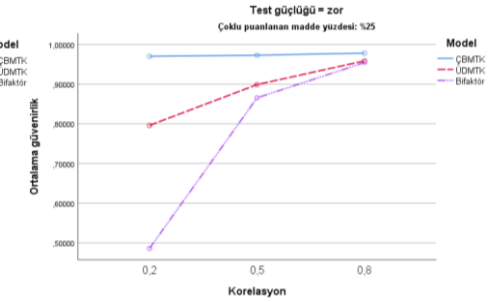

Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %10
Model
ÇBMTK
ÜDMTK
Bifaktör
Ortalama RMSE
Korelasyon
0,2
0,5
0,8

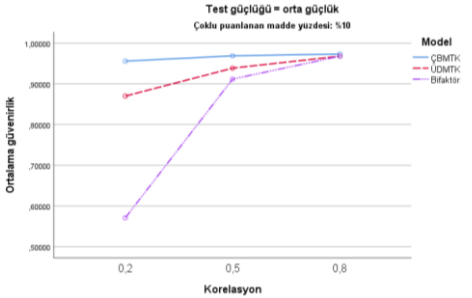

Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %5
Ortalama RMSE
Ortalama güvenirlik
Korelasyon
0,2
0,5
0,8
Model
ÇBMTK
ÜDMTK
Bifaktör

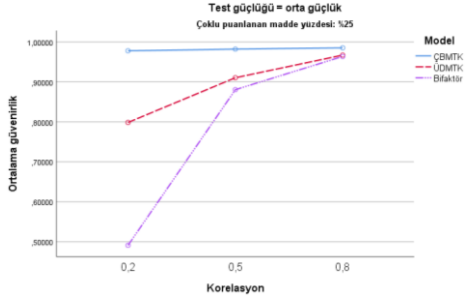

Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %10
Ortalama güvenirlik
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör

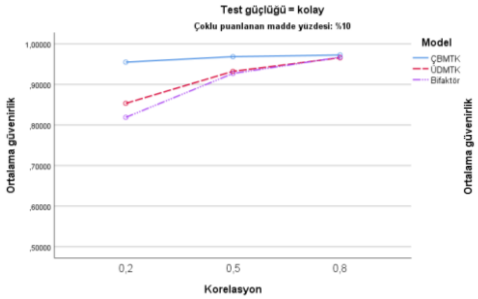
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %10
Ortalama güvenirlik
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör


Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %25
Ortalama güvenirlik
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör

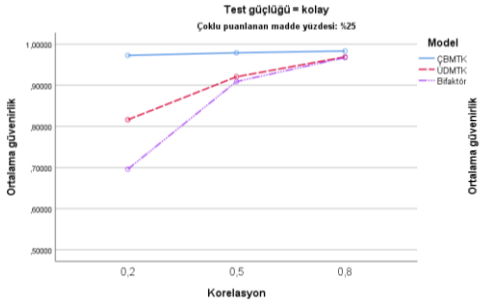

Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %25
Ortalama güvenirlik
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör

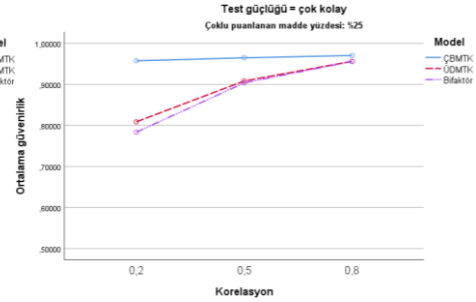

Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %10
Ortalama güvenirlik
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör

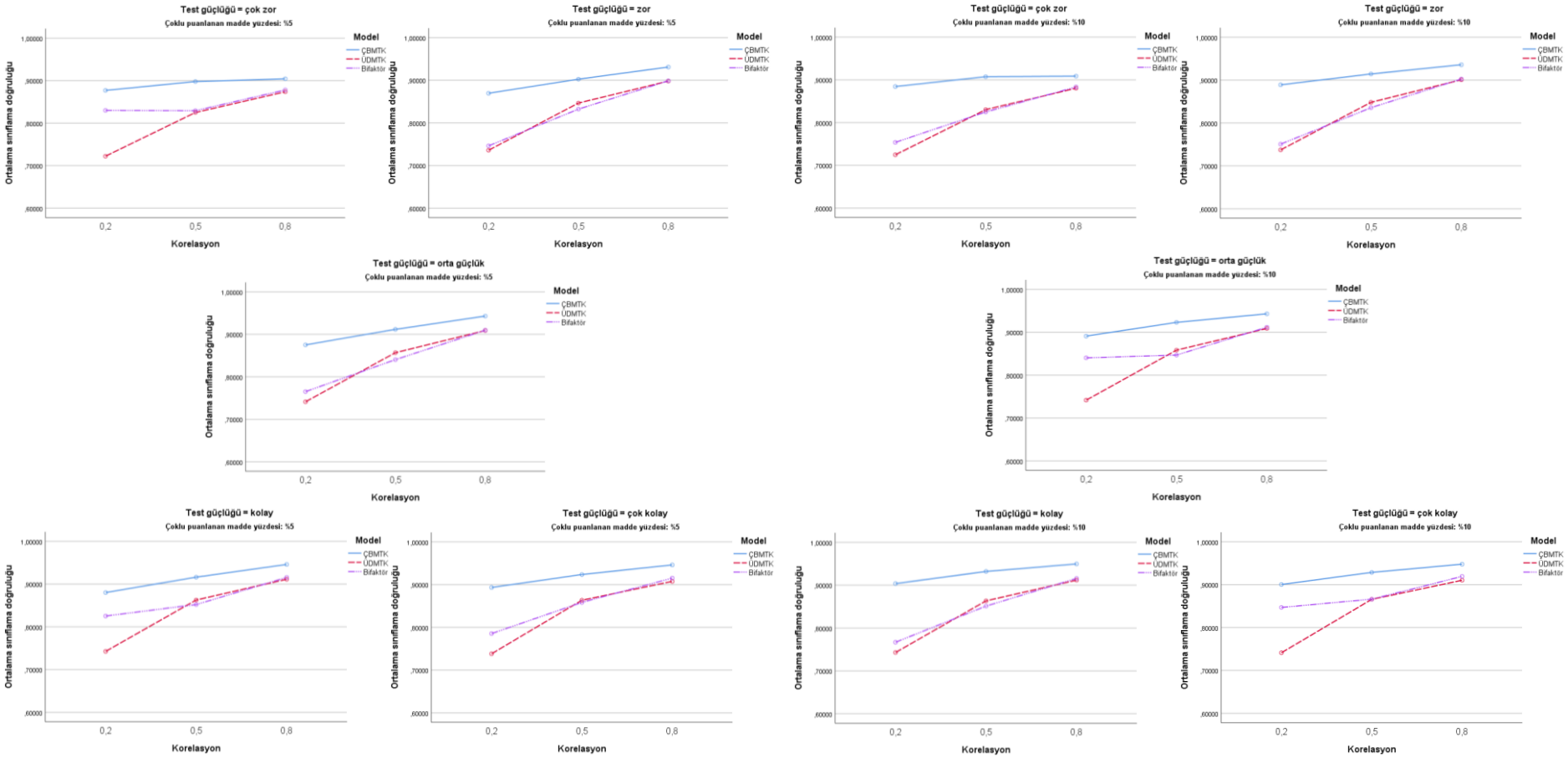

Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %25
Ortalama güvenirlik
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör

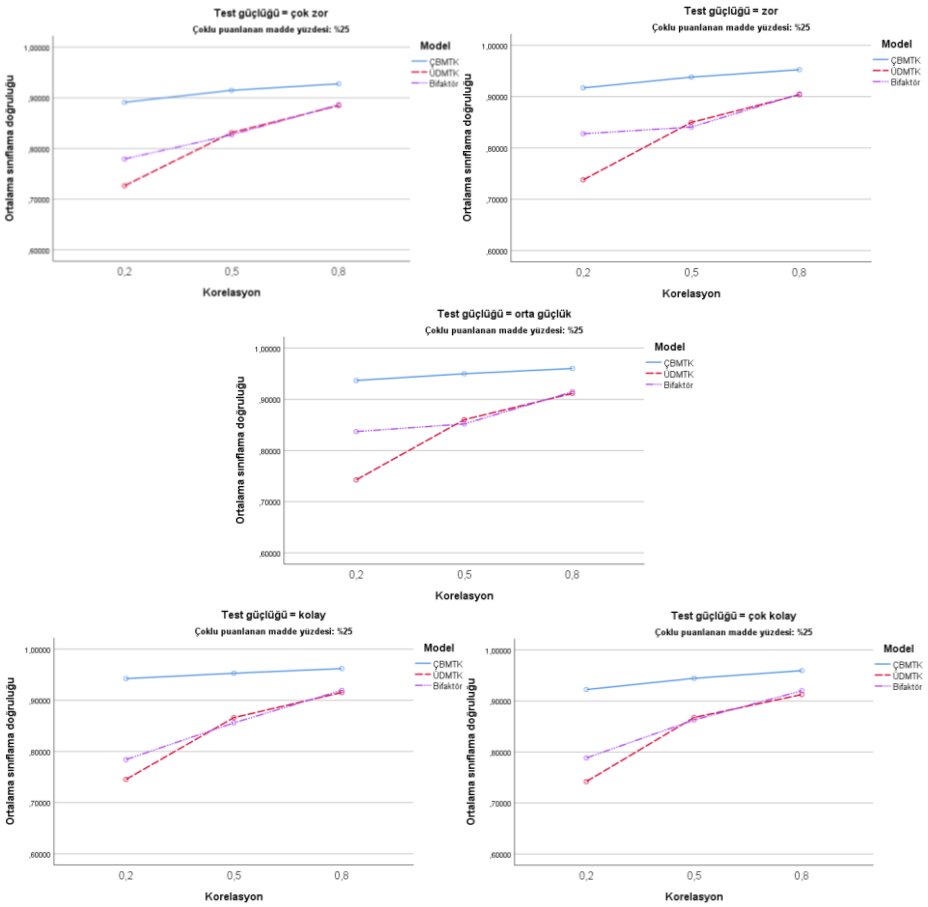

Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %10
Ortalama güvenirlik
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör

Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %10
Ortalama güvenirlik
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör


Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %25
Ortalama güvenirlik
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör


Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %25
Ortalama güvenirlik
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör


Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %5
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %10
Ortalama sınıflama doğruluğu
1,00000
,90000
,80000
,70000
,60000
0,2
0,5
0,8
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör


Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %25
Model
ÇBMTK
ÜDMTK
Bifaktör
Ortalama sınıflama doğruluğu
1,00000
,90000
,80000
,70000
,60000
0,2
0,5
0,8
Korelasyon

Ek 11. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,2)

Alt test için korelasyon 0,2 olduğunda her bir yöntem, çoklu puanlanan madde yüzdesi ve test güçlüğü düzeylerinde elde edilen ortalama RMSE değerleri

		Ortalama RMSE				Ortalama güvenirlik			
		Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
Model	Test güçlüğü	%5	%10	%25	%50	%5	%10	%25	%50
ÇBMTK	Çok zor	,525	,505	,492	,412	,849	,858	,864	,901
	Zor	,429	,409	,393	,344	,902	,911	,914	,932
	Orta	,381	,368	,348	,296	,924	,928	,933	,950
	Kolay	,388	,379	,348	,317	,921	,924	,935	,945
	Çok kolay	,444	,423	,416	,371	,894	,905	,907	,925
ÜDMTK	Çok zor	,528	,507	,494	,414	,847	,857	,862	,900
	Zor	,431	,410	,394	,345	,902	,910	,913	,931
	Orta	,382	,369	,349	,297	,923	,928	,933	,950
	Kolay	,389	,380	,349	,318	,921	,924	,934	,945
	Çok kolay	,446	,425	,418	,373	,893	,904	,906	,924
Bifaktör	Çok zor	,615	,576	,573	,514	,781	,812	,811	,846
	Zor	,513	,500	,518	,464	,857	,864	,842	,881
	Orta	,482	,504	,485	,433	,875	,853	,865	,897
	Kolay	,506	,476	,462	,439	,858	,879	,885	,897
	Çok kolay	,523	,528	,506	,478	,851	,845	,861	,876

Korelasyon 0,2 için çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25) her bir yöntem için test güçlüklerinin ikili karşılaştırmaları

			RMSE				Güvenirlik			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50
Model	(I) test güçlüğü	(J) test güçlüğü	Ortalama farkı (I-J)				Ortalama farkı (I-J)			
ÇBMTK	Çok zor	Zor	,096*	,096*	,099*	,068*	-,053*	-,053*	-,050*	-,031*
		Orta güçlük	,144*	,137*	,144*	,116*	-,075*	-,070*	-,069*	-,049*
		Kolay	,137*	,126*	,144*	,095*	-,073*	-,066*	-,071*	-,044*
		Çok kolay	,081*	,081*	,076*	,041*	-,045*	-,047*	-,043*	-,024*
	Zor	Orta güçlük	,048*	,041*	,045*	,048*	-,021*	-,017*	-,019*	-,018*
		Kolay	,041*	,030*	,045*	,027*	-,019*	-,013*	-,020*	-,013*
		Çok kolay	-,015*	-,014*	-,023*	-,027*	,008*	,006*	,007*	,007*
	Orta güçlük	Kolay	-,007*	-,011*	0,00	-,021*	0,002	,004*	-0,001	,005*
		Çok kolay	-,063*	-,055*	-,068*	-,075*	,029*	,023*	,026*	,025*
	Kolay	Çok kolay	-,056*	-,044*	-,068*	-,054*	,027*	,019*	,028*	,020*
		Zor	-,041*	-,030*	-,045*	-,027*	,019*	,013*	,020*	,013*
		Orta güçlük	,007*	,011*	0,00	,021*	-0,002	-,004*	0,001	-,005*
ÜDMTK	Çok zor	Zor	,097*	,097*	,100*	,069*	-,054*	-,054*	-,051*	-,031*
		Orta güçlük	,146*	,139*	,145*	,117*	-,076*	-,071*	-,070*	-,050*
		Kolay	,139*	,127*	,145*	,096*	-,074*	-,067*	-,072*	-,045*
		Çok kolay	,082*	,082*	,076*	,041*	-,046*	-,048*	-,044*	-,025*
	Zor	Orta güçlük	,049*	,041*	,045*	,048*	-,022*	-,018*	-,019*	-,018*
		Kolay	,042*	,030*	,045*	,027*	-,019*	-,013*	-,021*	-,013*
		Çok kolay	-,015*	-,015*	-,024*	-,028*	,008*	,006*	,008*	,007*
	Orta güçlük	Kolay	-,007*	-,011*	0,00	-,021*	0,002	,004*	-0,001	,005*
		Çok kolay	-,064*	-,056*	-,069*	-,076*	,030*	,024*	,027*	,025*
	Kolay	Çok kolay	-,057*	-,045*	-,069*	-,055*	,028*	,019*	,028*	,020*
Bifaktör	Çok zor	Zor	,102*	,076*	,054*	,050*	-,077*	-,052*	-,031*	-,036*
		Orta güçlük	,133*	,072*	,088*	,081*	-,094*	-,041*	-,054*	-,052*
		Kolay	,109*	,100*	,111*	,076*	-,077*	-,066*	-,073*	-,051*
		Çok kolay	,092*	,048*	,066*	,036*	-,070*	-,032*	-,050*	-,030*
	Zor	Orta güçlük	,031*	-0,004	,034*	,031*	-,018*	,011*	-,023*	-,016*
		Kolay	0,007	,024*	,057*	,025*	0,000	-,015*	-,042*	-,016*
		Çok kolay	-0,010	-,028*	0,01	-,014*	0,007	,019*	-,018*	0,006
	Orta güçlük	Kolay	-,024*	,028*	,023*	-0,006	,018*	-,026*	-,019*	0,000
		Çok kolay	-,041*	-,023*	-,022*	-,045*	,025*	0,008	0,005	,022*
	Kolay	Çok kolay	-,017*	-,052*	-,045*	-,040*	0,007	,034*	,024*	,021*

*p<,001

Korelasyon 0,2 için çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25 ve %50) her bir test güçlük düzeyi için modellerin ikili karşılaştırmaları

			RMSE				Güvenirlik			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50
Test güçlüğü	(I)Model	(J)Model	Ortalama farkı (I-J)				Ortalama farkı (I-J)			
Çok zor	ÇBMTK	ÜDMTK	-,003*	-,003*	-,002*	-,002*	,002*	,002*	,001*	,001*
		Bifaktör	-,090*	-,071*	-,081*	-,102*	,068*	,046*	,053*	,055*
	ÜDMTK	Bifaktör	-,087*	-,068*	-,078*	-,100*	,066*	,044*	,051*	,054*
Zor	ÇBMTK	ÜDMTK	-,002*	-,001*	-,001*	-,001*	,001*	,001*	,001*	,001*
		Bifaktör	-,084*	-,091*	-,126*	-,120*	,045*	,047*	,072*	,050*
	ÜDMTK	Bifaktör	-,083*	-,090*	-,124*	-,119*	,044*	,046*	,071*	,050*
Orta güçlük	ÇBMTK	ÜDMTK	-,001*	-,001*	-,001*	-,001*	,000*	,000*	,000*	,000*
		Bifaktör	-,101*	-,136*	-,137*	-,137*	,049*	,075*	,068*	,053*
	ÜDMTK	Bifaktör	-,100*	-,135*	-,136*	-,136*	,048*	,075*	,068*	,052*
Kolay	ÇBMTK	ÜDMTK	-,001*	-,001*	-,001*	-,001*	,001*	,001*	,000*	,000*
		Bifaktör	-,118*	-,097*	-,114*	-,122*	,064*	,045*	,050*	,048*
	ÜDMTK	Bifaktör	-,117*	-,096*	-,113*	-,121*	,063*	,045*	,050*	,048*
Çok kolay	ÇBMTK	ÜDMTK	-,002*	-,002*	-,002*	-,002*	,001*	,001*	,001*	,001*
		Bifaktör	-,079*	-,104*	-,090*	-,107*	,044*	,060*	,046*	,049*
	ÜDMTK	Bifaktör	-,077*	-,103*	-,088*	-,106*	,043*	,059*	,045*	,049*

*p<,001

Ek 12. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,2; alt test farkları)

Alt test için korelasyon 0,2 olduğunda her bir yöntem, çoklu puanlanan madde yüzdesi ve alt testlerde elde edilen ortalama RMSE değerleri

		Ortalama RMSE				Ortalama güvenirlik			
		Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
Model	Test güçlüğü	%5	%10	%25	%50	%5	%10	%25	%50
ÇBMTK	Alt test 1	,454	,455	,458	,465	,893	,893	,891	,888
	Alt test 2	,461	,443	,451	,444	,886	,895	,890	,895
	Alt test 3	,453	,452	,441	,241	,886	,886	,893	,970
	Alt test 4	,365	,317	,247	,242	,928	,947	,967	,969
ÜDMTK	Alt test 1	,456	,457	,460	,467	,892	,892	,890	,887
	Alt test 2	,463	,444	,453	,445	,885	,894	,889	,894
	Alt test 3	,455	,455	,443	,242	,885	,885	,892	,970
	Alt test 4	,367	,317	,247	,243	,927	,946	,967	,969
Bifaktör	Alt test 1	,629	,640	,662	,578	,784	,781	,764	,821
	Alt test 2	,526	,510	,515	,528	,849	,858	,855	,849
	Alt test 3	,518	,518	,505	,381	,849	,849	,858	,923
	Alt test 4	,438	,399	,352	,376	,895	,914	,933	,925

Korelasyon 0,2 için çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25) her bir yöntem için alt testlerin ikili karşılaştırmaları

			RMSE				Güvenirlik			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50
Model	(I) Alt test	(J) Alt test	Ortalama farkı (I-J)				Ortalama farkı (I-J)			
ÇBMTK	Alt test 1	Alt test 2	-,006*	,013*	,007*	,021*	,008*	-0,001	0,001	-,007*
		Alt test 3	0,001	0,003	,017*	,224*	,008*	,007*	-0,002	-,082*
		Alt test 4	,089*	,139*	,211*	,223*	-,035*	-,053*	-,076*	-,081*
	Alt test 2	Alt test 3	,007*	-,010*	,010*	,202*	0,000	,008*	-0,002	-,075*
		Alt test 4	,095*	,126*	,204*	,201*	-,042*	-,052*	-,077*	-,074*
	Alt test 3	Alt test 4	,088*	,136*	,194*	-0,001	-,042*	-,060*	-,075*	0,000
ÜDMTK	Alt test 1	Alt test 2	-,006*	,013*	,007*	,021*	,008*	-0,001	0,001	-,007*
		Alt test 3	0,001	0,003	,017*	,225*	,008*	,007*	-0,002	-,083*
		Alt test 4	,090*	,140*	,213*	,224*	-,035*	-,054*	-,077*	-,082*
	Alt test 2	Alt test 3	,007*	-,010*	,010*	,204*	0,000	,009*	-0,002	-,075*
		Alt test 4	,096*	,127*	,206*	,203*	-,043*	-,052*	-,078*	-,075*
	Alt test 3	Alt test 4	,089*	,137*	,196*	-0,001	-,043*	-,061*	-,075*	0,000
Bifaktör	Alt test 1	Alt test 2	,103*	,130*	,147*	,050*	-,064*	-,077*	-,091*	-,028*
		Alt test 3	,111*	,122*	,157*	,197*	-,065*	-,068*	-,094*	-,103*
		Alt test 4	,191*	,241*	,310*	,202*	-,111*	-,133*	-,169*	-,104*
	Alt test 2	Alt test 3	0,008	-0,007	0,010	,147*	0,000	,009*	-0,002	-,075*
		Alt test 4	,088*	,111*	,163*	,151*	-,046*	-,056*	-,078*	-,076*
	Alt test 3	Alt test 4	,080*	,119*	,153*	0,004	-,046*	-,065*	-,076*	-0,001
		Alt test 2	-0,008	0,007	-0,010	-,147*	0,000	-,009*	0,002	,075*

*p<,001

Ek 13. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,5)

Alt testler için korelasyon 0,5 olduğunda her bir yöntem, çoklu puanlanan madde yüzdesi ve test güçlüğü düzeylerinde elde edilen ortalama RMSE değerleri

		Ortalama RMSE				Ortalama güvenirlik			
		Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
Model	Test güçlüğü	%5	%10	%25	%50	%5	%10	%25	%50
ÇBMTK	Çok zor	,504	,485	,472	,397	,861	,870	,875	,910
	Zor	,416	,397	,380	,333	,908	,916	,920	,936
	Orta	,370	,357	,337	,288	,928	,932	,937	,953
	Kolay	,374	,365	,336	,307	,927	,930	,939	,939
	Çok kolay	,424	,405	,398	,358	,905	,914	,915	,932
ÜDMTK	Çok zor	,507	,488	,474	,399	,860	,869	,874	,909
	Zor	,417	,398	,381	,334	,908	,915	,919	,936
	Orta	,371	,358	,338	,288	,928	,932	,937	,952
	Kolay	,375	,366	,337	,308	,927	,929	,939	,939
	Çok kolay	,426	,407	,400	,360	,904	,913	,915	,931
Bifaktör	Çok zor	,756	,744	,742	,717	,651	,663	,664	,690
	Zor	,713	,707	,708	,690	,701	,707	,704	,723
	Orta	,697	,697	,691	,673	,718	,718	,722	,740
	Kolay	,701	,694	,687	,677	,714	,721	,728	,730
	Çok kolay	,715	,712	,707	,697	,698	,701	,707	,717

Alt testler - Korelasyon 0,5 için çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25) her bir yöntem için test güçlüklerinin ikili karşılaştırmaları

			RMSE				Güvenirlik			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50
Model	(I) test güçlüğü	(J) test güçlüğü	Ortalama farkı (I-J)				Ortalama farkı (I-J)			
ÇBMTK	Çok zor	Zor	,089*	,088*	,092*	,064*	-,047*	-,046*	-,044*	-,027*
		Orta güçlük	,135*	,128*	,135*	,109*	-,067*	-,062*	-,062*	-,043*
		Kolay	,131*	,120*	,136*	,090*	-,066*	-,059*	-,064*	-,029*
		Çok kolay	,080*	,080*	,074*	,039*	-,043*	-,043*	-,040*	-,022*
	Zor	Orta güçlük	,046*	,040*	,043*	,045*	-,020*	-,016*	-,017*	-,016*
		Kolay	,042*	,031*	,044*	,026*	-,019*	-,013*	-,020*	-0,003
		Çok kolay	-,008*	-,008*	-,018*	-,025*	,004*	0,002	,004*	0,005
	Orta güçlük	Kolay	-0,004	-,008*	0,001	-,019*	0,001	,003*	-0,002	,013*
		Çok kolay	-,054*	-,048*	-,061*	-,070*	,024*	,019*	,022*	,021*
	Kolay	Çok kolay	-,050*	-,040*	-,062*	-,051*	,023*	,016*	,024*	0,007
ÜDMTK	Çok zor	Zor	,090*	,090*	,093*	,065*	-,048*	-,047*	-,045*	-,027*
		Orta güçlük	,137*	,129*	,136*	,111*	-,068*	-,063*	-,063*	-,043*
		Kolay	,132*	,121*	,137*	,091*	-,067*	-,061*	-,065*	-,030*
		Çok kolay	,081*	,081*	,074*	,039*	-,044*	-,044*	-,041*	-,022*
	Zor	Orta güçlük	,047*	,040*	,043*	,046*	-,020*	-,016*	-,018*	-,017*
		Kolay	,042*	,032*	,044*	,026*	-,019*	-,014*	-,020*	-0,003
		Çok kolay	-,009*	-,009*	-,019*	-,026*	,004*	0,003	,004*	0,005
	Orta güçlük	Kolay	-,004	-,008*	0,001	-,020*	0,001	,003*	-0,002	,013*
		Çok kolay	-,055*	-,049*	-,062*	-,072*	,024*	,019*	,022*	,021*
	Kolay	Çok kolay	-,051*	-,040*	-,063*	-,052*	,023*	,016*	,024*	0,008
Bifaktör	Çok zor	Zor	,043*	,037*	,034*	,027*	-,050*	-,044*	-,040*	-,032*
		Orta güçlük	,058*	,048*	,051*	,044*	-,067*	-,054*	-,058*	-,049*
		Kolay	,055*	,051*	,056*	,040*	-,063*	-,057*	-,063*	-,040*
		Çok kolay	,041*	,032*	,035*	,020*	-,047*	-,038*	-,042*	-,027*
	Zor	Orta güçlük	,015*	,011*	,016*	,017*	-,016*	-,011*	-,018*	-,017*
		Kolay	,012*	,014*	,021*	,013*	-,012*	-,014*	-,024*	-0,007
		Çok kolay	-,002	-,004	0,001	-,007*	0,003	,006*	-0,003	0,005
	Orta güçlük	Kolay	-0,004	0,003	,005*	-0,004	0,004	-0,003	-,006*	0,010
		Çok kolay	-,018*	-,015*	-,015*	-,024*	,020*	,016*	,015*	,023*
	Kolay	Çok kolay	-,014*	-,018*	-,020*	-,020*	,016*	,019*	,021*	,013*

*p<,001

Alt testler - Korelasyon 0,5 için çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25 ve %50) her bir test güçlük düzeyi için modellerin ikili karşılaştırmaları

			RMSE				Güvenirlik			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50
Test güçlüğü	(I)Model	(J)Model	Ortalama farkı (I-J)				Ortalama farkı (I-J)			
Çok zor	ÇBMTK	ÜDMTK	-,003*	-,003*	-,003*	-,003*	,002*	,002*	,001*	,001*
		Bifaktör	-,251*	-,260*	-,271*	-,321*	,210*	,207*	,211*	,219*
	ÜDMTK	Bifaktör	-,248*	-,257*	-,268*	-,318*	,208*	,205*	,210*	,219*
Zor	ÇBMTK	ÜDMTK	-,002*	-,001*	-,001*	-,001*	,001*	,001*	,001*	,000*
		Bifaktör	-,297*	-,311*	-,328*	-,357*	,207*	,209*	,216*	,214*
	ÜDMTK	Bifaktör	-,296*	-,309*	-,327*	-,356*	,206*	,208*	,215*	,213*
Orta güçlük	ÇBMTK	ÜDMTK	-,001*	-,001*	-,001*	-,001*	,000*	,000*	,000*	,000*
		Bifaktör	-,328*	-,340*	-,354*	-,386*	,210*	,215*	,215*	,213*
	ÜDMTK	Bifaktör	-,327*	-,339*	-,353*	-,385*	,210*	,214*	,215*	,213*
Kolay	ÇBMTK	ÜDMTK	-,001*	-,001*	-,001*	-,001*	,000*	,000*	,000*	,000*
		Bifaktör	-,327*	-,329*	-,351*	-,370*	,213*	,209*	,211*	,209*
	ÜDMTK	Bifaktör	-,326*	-,327*	-,350*	-,369*	,213*	,208*	,211*	,209*
Çok kolay	ÇBMTK	ÜDMTK	-,002*	-,002*	-,002*	-,002*	,001*	,001*	,001*	,001*
		Bifaktör	-,291*	-,307*	-,309*	-,340*	,206*	,212*	,209*	,215*
	ÜDMTK	Bifaktör	-,289*	-,305*	-,306*	-,337*	,205*	,211*	,208*	,214*

*p<,001

Ek 14. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,5; alt test farkları)

Alt testler için korelasyon 0,5 olduğunda her bir yöntem, çoklu puanlanan madde yüzdesi ve alt testlerde elde edilen ortalama RMSE değerleri

		Ortalama RMSE				Ortalama güvenirlik			
		Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
Model	Test güçlüğü	%5	%10	%25	%50	%5	%10	%25	%50
ÇBMTK	Alt test 1	,432	,432	,434	,438	,900	,900	,899	,896
	Alt test 2	,443	,427	,434	,426	,895	,903	,900	,903
	Alt test 3	,433	,433	,422	,237	,896	,896	,902	,969
	Alt test 4	,362	,315	,248	,244	,932	,949	,969	,969
ÜDMTK	Alt test 1	,434	,434	,436	,440	,899	,899	,898	,895
	Alt test 2	,445	,428	,436	,428	,894	,902	,899	,902
	Alt test 3	,435	,435	,424	,238	,895	,895	,901	,969
	Alt test 4	,363	,316	,249	,245	,932	,949	,969	,969
Bifaktör	Alt test 1	,745	,751	,760	,754	,666	,662	,651	,657
	Alt test 2	,728	,720	,727	,734	,689	,697	,691	,683
	Alt test 3	,714	,716	,711	,629	,692	,690	,696	,773
	Alt test 4	,680	,657	,630	,647	,739	,759	,781	,767

Alt testler - Korelasyon 0,5 için çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25) her bir yöntem için alt testlerin ikili karşılaştırmaları

			RMSE				Güvenirlik			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50
Model	(I) Alt test	(J) Alt test	Ortalama farkı (I-J)				Ortalama farkı (I-J)			
ÇBMTK	Alt test 1	Alt test 2	-,011*	,006*	0,00	,012*	,005*	-,003*	-0,001	-0,007
		Alt test 3	-0,001	0,000	,012*	,201*	,004*	,004*	-,003*	-,074*
		Alt test 4	,070*	,117*	,186*	,194*	-,033*	-,049*	-,070*	-,073*
	Alt test 2	Alt test 3	,010*	-,006*	,012*	,189*	-0,001	,007*	-,003*	-,067*
		Alt test 4	,081*	,112*	,186*	,182*	-,037*	-,047*	-,069*	-,066*
	Alt test 3	Alt test 4	,071*	,118*	,174*	-,007	-,037*	-,053*	-,067*	0,001
ÜDMTK	Alt test 1	Alt test 2	-,011*	,006*	0,00	,012*	,005*	-,003*	-0,001	-0,007
		Alt test 3	-0,001	-0,001	,012*	,202*	,004*	,004*	-,003*	-,075*
		Alt test 4	,071*	,118*	,187*	,195*	-,033*	-,050*	-,071*	-,074*
	Alt test 2	Alt test 3	,010*	-,006*	,012*	,190*	-0,001	,007*	-,003*	-,067*
		Alt test 4	,082*	,112*	,187*	,183*	-,038*	-,047*	-,070*	-,067*
	Alt test 3	Alt test 4	,072*	,119*	,175*	-,007	-,037*	-,054*	-,067*	0,001
Bifaktör	Alt test 1	Alt test 2	,017*	,031*	,033*	,020*	-,023*	-,035*	-,040*	-,026*
		Alt test 3	,031*	,035*	,049*	,125*	-,025*	-,028*	-,045*	-,116*
		Alt test 4	,065*	,094*	,130*	,107*	-,072*	-,097*	-,130*	-,110*
	Alt test 2	Alt test 3	,014*	,004*	,016*	,105*	-0,003	,007*	-,005*	-,090*
		Alt test 4	,048*	,063*	,097*	,087*	-,050*	-,062*	-,091*	-,084*
	Alt test 3	Alt test 4	,034*	,059*	,081*	-,018	-,047*	-,069*	-,085*	0,006

*p<,001

Ek 15. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,8)

Alt testler için korelasyon 0,8 olduğunda her bir yöntem, çoklu puanlanan madde yüzdesi ve test güçlüğü düzeylerinde elde edilen ortalama RMSE değerleri

		Ortalama RMSE				Ortalama güvenirlik			
		Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
Model	Test güçlüğü	%5	%10	%25	%50	%5	%10	%25	%50
ÇBMTK	Çok zor	,446	,434	,412	,346	,895	,900	,910	,938
	Zor	,369	,359	,335	,295	,929	,933	,940	,953
	Orta	,331	,327	,301	,259	,943	,944	,951	,963
	Kolay	,336	,335	,302	,279	,942	,942	,952	,959
	Çok kolay	,376	,368	,356	,324	,926	,930	,935	,947
ÜDMTK	Çok zor	,449	,437	,415	,349	,893	,899	,908	,937
	Zor	,370	,360	,336	,296	,929	,932	,939	,953
	Orta	,332	,328	,302	,260	,943	,944	,951	,963
	Kolay	,337	,336	,303	,280	,942	,942	,951	,959
	Çok kolay	,378	,370	,358	,326	,926	,930	,934	,947
Bifaktör	Çok zor	,925	,922	,923	,925	,410	,415	,416	,431
	Zor	,920	,921	,923	,923	,436	,440	,438	,450
	Orta	,921	,923	,922	,922	,446	,446	,449	,463
	Kolay	,921	,920	,921	,922	,442	,446	,452	,459
	Çok kolay	,919	,921	,920	,924	,431	,434	,436	,444

Alt testler - Korelasyon 0,8 için çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25) her bir yöntem için test güçlüklerinin ikili karşılaştırmaları

			RMSE				Güvenirlik			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50
Model	(I) test güçlüğü	(J) test güçlüğü	Ortalama farkı (I-J)				Ortalama farkı (I-J)			
ÇBMTK	Çok zor	Zor	,077*	,075*	,077*	,051*	-,034*	-,033*	-,030*	-,015*
		Orta güçlük	,114*	,107*	,111*	,087*	-,048*	-,044*	-,041*	-,025*
		Kolay	,110*	,099*	,109*	,067*	-,047*	-,042*	-,042*	-,021*
		Çok kolay	,069*	,066*	,056*	,023*	-,032*	-,030*	-,025*	-,009*
	Zor	Orta güçlük	,038*	,031*	,033*	,035*	-,014*	-,011*	-,011*	-,010*
		Kolay	,033*	,023*	,032*	,016*	-,013*	-,009*	-,012*	-,006*
		Çok kolay	-,007*	-,010*	-,021*	-,029*	,003*	,003*	,005*	,006*
	Orta güçlük	Kolay	-,004*	-,008*	-0,001	-,020*	0,001	,002*	-0,001	,004*
		Çok kolay	-,045*	-,041*	-,054*	-,064*	,017*	,014*	,016*	,015*
	Kolay	Çok kolay	-,041*	-,033*	-,053*	-,045*	,016*	,012*	,017*	,012*
ÜDMTK	Çok zor	Zor	,079*	,077*	,079*	,054*	-,035*	-,034*	-,031*	-,016*
		Orta güçlük	,117*	,109*	,113*	,090*	-,050*	-,045*	-,042*	-,026*
		Kolay	,112*	,101*	,111*	,069*	-,048*	-,043*	-,043*	-,022*
		Çok kolay	,071*	,067*	,057*	,023*	-,033*	-,031*	-,026*	-,010*
	Zor	Orta güçlük	,038*	,032*	,034*	,036*	-,014*	-,012*	-,012*	-,010*
		Kolay	,034*	,024*	,033*	,016*	-,013*	-,009*	-,012*	-,006*
		Çok kolay	-,008*	-,010*	-,022*	-,030*	,003*	,003*	,005*	,006*
	Orta güçlük	Kolay	-,005*	-,008*	-0,001	-,021*	0,001	,002*	-0,001	,004*
		Çok kolay	-,046*	-,042*	-,056*	-,066*	,017*	,014*	,017*	,016*
	Kolay	Çok kolay	-,041*	-,034*	-,055*	-,046*	,016*	,012*	,017*	,012*
Bifaktör	Çok zor	Zor	,005*	0,001	0,000	0,002	-,027*	-,025*	-,022*	-,020*
		Orta güçlük	,004*	-0,001	0,001	,003*	-,036*	-,030*	-,033*	-,032*
		Kolay	,003*	0,002	0,002	,003*	-,032*	-,030*	-,036*	-,028*
		Çok kolay	,006*	0,001	,003*	0,001	-,021*	-,018*	-,020*	-,013*
	Zor	Orta güçlük	-0,001	-0,001	0,001	0,001	-,009*	-,005*	-,011*	-,013*
		Kolay	-0,002	0,001	0,002	0,001	-,005*	-,005*	-,014*	-,008*
		Çok kolay	0,000	0,000	,003*	-0,001	,006*	,006*	0,002	,007*
	Orta güçlük	Kolay	-0,001	,003*	0,001	0,000	0,004	0,000	-0,003	,004*
		Çok kolay	0,001	0,001	0,002	-0,002	,015*	,012*	,013*	,019*
	Kolay	Çok kolay	0,002	-0,001	0,001	-0,002	,011*	,012*	,016*	,015*

*p<,001

Alt testler - Korelasyon 0,8 için çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25 ve %50) her bir test güçlük düzeyi için modellerin ikili karşılaştırmaları

			RMSE				Güvenirlik			
			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50
Test güçlüğü	(I)Model	(J)Model	Ortalama farkı (I-J)				Ortalama farkı (I-J)			
Çok zor	ÇBMTK	ÜDMTK	-,003*	-,003*	-,003*	-,003*	,002*	,002*	,001*	,001*
		Bifaktör	-,479*	-,488*	-,511*	-,579*	,485*	,485*	,494*	,507*
	ÜDMTK	Bifaktör	-,476*	-,485*	-,508*	-,576*	,483*	,483*	,492*	,506*
Zor	ÇBMTK	ÜDMTK	-,001*	-,001*	-,001*	-,001*	,001*	,001*	,000*	,000*
		Bifaktör	-,551*	-,563*	-,589*	-,628*	,493*	,493*	,502*	,503*
	ÜDMTK	Bifaktör	-,549*	-,561*	-,587*	-,627*	,492*	,492*	,502*	,502*
Orta güçlük	ÇBMTK	ÜDMTK	-,001*	-,001*	-,001*	0,00	,000*	,000*	,000*	,000
		Bifaktör	-,589*	-,596*	-,621*	-,663*	,497*	,499*	,502*	,499*
	ÜDMTK	Bifaktör	-,589*	-,595*	-,620*	-,662*	,497*	,499*	,502*	,499*
Kolay	ÇBMTK	ÜDMTK	-,001*	-,001*	-,001*	-,001*	,000*	,000*	,000*	,000*
		Bifaktör	-,586*	-,585*	-,619*	-,643*	,500*	,496*	,500*	,500*
	ÜDMTK	Bifaktör	-,585*	-,584*	-,618*	-,642*	,500*	,496*	,500*	,500*
Çok kolay	ÇBMTK	ÜDMTK	-,002*	-,002*	-,003*	-,002*	,001*	,001*	,001*	,000*
		Bifaktör	-,543*	-,553*	-,565*	-,600*	,496*	,497*	,499*	,503*
	ÜDMTK	Bifaktör	-,541*	-,551*	-,562*	-,598*	,495*	,496*	,498*	,503*

*p<,001

Ek 16. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,8; alt test farkları)

Alt testler için korelasyon 0,8 olduğunda her bir yöntem, çoklu puanlanan madde yüzdesi ve alt testlerde elde edilen ortalama RMSE değerleri

		Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
		%5	%10	%25	%50	%5	%10	%25	%50
Model	Test güçlüğü	Ortalama RMSE				Ortalama güvenirlik			
ÇBMTK	Alt test 1	,381	,380	,377	,375	,923	,924	,925	,928
	Alt test 2	,389	,377	,377	,368	,920	,925	,925	,931
	Alt test 3	,387	,386	,376	,230	,921	,922	,927	,975
	Alt test 4	,329	,315	,234	,230	,944	,949	,972	,975
ÜDMTK	Alt test 1	,383	,382	,379	,377	,922	,923	,924	,927
	Alt test 2	,391	,379	,380	,369	,919	,924	,924	,930
	Alt test 3	,389	,388	,377	,231	,921	,921	,926	,975
	Alt test 4	,330	,316	,235	,231	,944	,949	,972	,974
Bifaktör	Alt test 1	,929	,935	,941	,951	,412	,406	,396	,384
	Alt test 2	,925	,925	,933	,942	,421	,426	,415	,406
	Alt test 3	,932	,937	,938	,901	,419	,413	,414	,505
	Alt test 4	,898	,889	,876	,899	,479	,500	,527	,502

Alt testler - Korelasyon 0,8 için çoklu puanlanan madde yüzdesi düzeylerinde (%5, %10, %25) her bir yöntem için alt testlerin ikili karşılaştırmaları

			Çoklu puanlanan madde yüzdesi				Çoklu puanlanan madde yüzdesi			
			%5	%10	%25	%50	%5	%10	%25	%50
Model	(I) Alt test	(J) Alt test	Ortalama farkı (I-J)				Ortalama farkı (I-J)			
ÇBMTK	Alt test 1	Alt test 2	-,007*	,004*	0,000	,008*	,003*	-0,001	0,000	-,003*
		Alt test 3	-,006*	-,006*	0,002	,145*	,002*	,002*	-,002*	-,047*
		Alt test 4	,053*	,066*	,143*	,145*	-,021*	-,025*	-,047*	-,047*
	Alt test 2	Alt test 3	0,001	-,009*	0,002	,138*	-,002*	,003*	-,002*	-,044*
		Alt test 4	,060*	,062*	,143*	,137*	-,024*	-,024*	-,047*	-,044*
	Alt test 3	Alt test 4	,058*	,071*	,141*	0,000	-,023*	-,027*	-,045*	0,000
ÜDMTK	Alt test 1	Alt test 2	-,007*	,004*	0,000	,008*	,003*	-0,001	0,000	-,003*
		Alt test 3	-,006*	-,006*	0,002	,146*	,002*	,002*	-,002*	-,048*
		Alt test 4	,053*	,066*	,144*	,146*	-,021*	-,026*	-,047*	-,047*
	Alt test 2	Alt test 3	0,001	-,009*	0,002	,138*	-,002*	,003*	-,002*	-,045*
		Alt test 4	,060*	,063*	,144*	,138*	-,024*	-,024*	-,048*	-,044*
	Alt test 3	Alt test 4	,059*	,072*	,142*	0,000	-,023*	-,027*	-,046*	0,000
Bifaktör	Alt test 1	Alt test 2	,004*	,010*	,008*	,009*	-,009*	-,020*	-,020*	-,021*
		Alt test 3	-,003*	-,002*	,003*	,050*	-,007*	-,007*	-,018*	-,121*
		Alt test 4	,031*	,046*	,066*	,051*	-,067*	-,094*	-,132*	-,118*
	Alt test 2	Alt test 3	-,007*	-,012*	-,006*	,041*	0,002	,013*	0,002	-,099*
		Alt test 4	,027*	,036*	,057*	,043*	-,058*	-,074*	-,112*	-,097*
	Alt test 3	Alt test 4	,034*	,049*	,063*	0,002	-,060*	-,087*	-,113*	0,002

*p<,001

Ek 17. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,2; ÇPMY basit ana etkisi)

Alt testler için korelasyon 0,2 olduğunda test güçlüğü düzeylerinde her bir yöntem için çoklu puanlanan madde yüzdelerinin karşılaştırılması (RMSE)

		RMSE		Güvenirlik	
Test güçlüğü	Model	F	η_p^2	F	η_p^2
Çok zor	ÇBMTK	340,869	,391	309,826	,368
	ÜDMTK	340,324	,391	309,657	,368
	Bifaktör	289,121	,353	245,436	,316
Zor	ÇBMTK	200,072	,274	169,319	,242
	ÜDMTK	199,942	,274	169,595	,242
	Bifaktör	72,442	,120	80,790	,132
Orta güçlük	ÇBMTK	331,227	,384	320,383	,376
	ÜDMTK	330,517	,384	319,288	,376
	Bifaktör	180,087	,253	166,323	,239
Kolay	ÇBMTK	348,034	,396	336,903	,388
	ÜDMTK	347,337	,395	336,578	,388
	Bifaktör	163,148	,235	149,806	,220
Çok kolay	ÇBMTK	354,889	,401	335,742	,387
	ÜDMTK	352,746	,399	334,196	,386
	Bifaktör	101,109	,160	91,751	,147

*p<,001

Alt testler için korelasyon 0,2 olduğunda her bir yöntem için çoklu puanlanan madde yüzdelerinde hesaplanan ortalama RMSE değerleri

		Test güçlüğü					Test güçlüğü				
		Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
Model	Çoklu puanlanan madde yüzdesi	Ortalama RMSE					Ortalama Güvenirlik				
ÇBMTK	%5	,525	,429	,381	,388	,444	,849	,902	,924	,921	,894
	%10	,505	,409	,368	,379	,423	,858	,911	,928	,924	,905
	%25	,492	,393	,348	,348	,416	,864	,914	,933	,935	,907
	%50	,412	,344	,296	,317	,371	,901	,932	,950	,945	,925
ÜDMTK	%5	,528	,431	,382	,389	,446	,847	,902	,923	,921	,893
	%10	,507	,410	,369	,380	,425	,857	,910	,928	,924	,904
	%25	,494	,394	,349	,349	,418	,862	,913	,933	,934	,906
	%50	,414	,345	,297	,318	,373	,900	,931	,950	,945	,924
Bifaktör	%5	,615	,513	,482	,506	,523	,781	,857	,875	,858	,851
	%10	,576	,500	,504	,476	,528	,812	,864	,853	,879	,845
	%25	,573	,518	,485	,462	,506	,811	,842	,865	,885	,861
	%50	,514	,464	,433	,439	,478	,846	,881	,897	,897	,876

Toplam test için korelasyon 0,2 olduğunda her bir yöntem için çoklu puanlanan madde yüzdelerinin ikili karşılaştırmaları

			Test güçlüğü					Test güçlüğü				
			Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
Model	(I) Çoklu puanlanan madde yüzdesi	(J) Çoklu puanlanan madde yüzdesi	Ortalama fark (I-J)					Ortalama fark (I-J)				
ÇBMTK	%5	%10	,020*	,020*	,013*	0,009	,021*	-,009*	-,008*	-,005*	-0,003	-,011*
		%25	,033*	,036*	,033*	,040*	,028*	-,015*	-,012*	-,010*	-,013*	-,012*
		%50	,113*	,085*	,085*	,071*	,073*	-,052*	-,029*	-,026*	-,024*	-,031*
	%10	%25	0,013	,016*	,020*	,031*	0,007	-0,005	-0,003	-,005*	-,011*	-0,002
		%50	,093*	,065*	,072*	,062*	,052*	-,043*	-,021*	-,022*	-,021*	-,020*
	%25	%50	,080*	,049*	,052*	,031*	,045*	-,037*	-,018*	-,017*	-,010*	-,018*
ÜDMTK	%5	%10	,020*	,020*	,013*	0,009	,021*	-,009*	-,009*	-,005*	-0,003	-,011*
		%25	,034*	,037*	,033*	,040*	,028*	-,015*	-,012*	-,010*	-,013*	-,013*
		%50	,114*	,086*	,086*	,072*	,073*	-,053*	-,030*	-,027*	-,024*	-,031*
	%10	%25	0,013	,016*	,020*	,031*	0,007	-0,006	-0,003	-,005*	-,011*	-0,002
		%50	,093*	,065*	,072*	,063*	,052*	-,043*	-,021*	-,022*	-,021*	-,020*
	%25	%50	,080*	,049*	,052*	,031*	,045*	-,038*	-,018*	-,017*	-,010*	-,019*
Bifaktör	%5	%10	,039*	0,013	-,022*	,030*	-0,004	-,032*	-0,007	,022*	-,021*	0,006
		%25	,042*	-0,005	-0,003	,044*	,017*	-,030*	,015*	,010*	-,027*	-,010*
		%50	,101*	,049*	,049*	,067*	,045*	-,065*	-,024*	-,022*	-,039*	-,025*
	%10	%25	0,003	-,018*	,020*	,014*	,021*	0,001	,022*	-,012*	-0,006	-,016*
		%50	,062*	,036*	,071*	,037*	,049*	-,033*	-,017*	-,045*	-,018*	-,031*
	%25	%50	,059*	,054*	,051*	,023*	,028*	-,035*	-,039*	-,032*	-,012*	-,015*

*p<,001

Ek 18. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,5; ÇPMY basit ana etkisi)

Alt testler için korelasyon 0,5 olduğunda test güçlüğü düzeylerinde her bir yöntem için çoklu puanlanan madde yüzdelerinin karşılaştırılması (RMSE)

		RMSE		Güvenirlik	
Test güçlüğü	Model	F	η_p^2	F	η_p^2
Çok zor	ÇBMTK	384,025	,420	372,622	,412
	ÜDMTK	380,393	,417	372,690	,412
	Bifaktör	121,299	,186	115,084	,178
Zor	ÇBMTK	219,162	,292	194,619	,268
	ÜDMTK	218,355	,291	194,908	,269
	Bifaktör	57,096	,097	57,162	,097
Orta güçlük	ÇBMTK	345,782	,394	335,246	,387
	ÜDMTK	345,102	,394	334,243	,386
	Bifaktör	155,293	,226	163,449	,235
Kolay	ÇBMTK	349,689	,397	6,771	,013
	ÜDMTK	346,265	,395	6,953	,013
	Bifaktör	169,284	,242	13,484	,025
Çok kolay	ÇBMTK	362,635	,406	361,977	,405
	ÜDMTK	354,294	,400	361,271	,405
	Bifaktör	80,047	,131	85,674	,139

*p<,001

Alt testler için korelasyon 0,5 olduğunda her bir yöntem için çoklu puanlanan madde yüzdelerinde hesaplanan ortalama RMSE değerleri

		Test güçlüğü					Test güçlüğü				
		Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
Model	Çoklu puanlanan madde yüzdesi	Ortalama RMSE					Ortalama güvenirlik				
ÇBMTK	%5	,504	,416	,370	,374	,424	,861	,908	,928	,927	,905
	%10	,485	,397	,357	,365	,405	,870	,916	,932	,930	,914
	%25	,472	,380	,337	,336	,398	,875	,920	,937	,939	,915
	%50	,397	,333	,288	,307	,358	,910	,936	,953	,939	,932
ÜDMTK	%5	,507	,417	,371	,375	,426	,860	,908	,928	,927	,904
	%10	,488	,398	,358	,366	,407	,869	,915	,932	,929	,913
	%25	,474	,381	,338	,337	,400	,874	,919	,937	,939	,915
	%50	,399	,334	,288	,308	,360	,909	,936	,952	,939	,931
Bifaktör	%5	,756	,713	,697	,701	,715	,651	,701	,718	,714	,698
	%10	,744	,707	,697	,694	,712	,663	,707	,718	,721	,701
	%25	,742	,708	,691	,687	,707	,664	,704	,722	,728	,707
	%50	,717	,690	,673	,677	,697	,690	,723	,740	,730	,717

Alt testler için korelasyon 0,5 olduğunda her bir yöntem için çoklu puanlanan madde yüzdelerinin ikili karşılaştırmaları

			Test güçlüğü					Test güçlüğü				
			Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
Model	(I) Çoklu puanlanan madde yüzdesi	(J) Çoklu puanlanan madde yüzdesi	Ortalama fark (I-J)					Ortalama fark (I-J)				
ÇBMTK	%5	%10	,020*	,019*	,013*	,009*	,019*	-,009*	-,008*	-,004*	-0,002	-,009*
		%25	,033*	,036*	,033*	,038*	,026*	-,014*	-,011*	-,009*	-0,012	-,011*
		%50	,108*	,083*	,082*	,067*	,066*	-,049*	-,028*	-,025*	-0,012	-,027*
	%10	%25	,013*	,017*	,020*	,029*	0,007	-0,005	-0,004	-,005*	-0,010	-0,002
		%50	,088*	,064*	,069*	,058*	,047*	-,040*	-,020*	-,020*	-0,010	-,018*
	%25	%50	,075*	,047*	,049*	,029*	,040*	-,034*	-,017*	-,016*	0,000	-,016*
ÜDMTK	%5	%10	,020*	,019*	,013*	,009*	,019*	-,009*	-,008*	-,004*	-0,002	-,009*
		%25	,033*	,036*	,033*	,038*	,026*	-,015*	-,011*	-,009*	-0,012	-,011*
		%50	,108*	,083*	,082*	,067*	,066*	-,049*	-,028*	-,025*	-0,012	-,028*
	%10	%25	,013*	,017*	,020*	,030*	0,006	-0,006	-0,004	-,005*	-0,010	-0,002
		%50	,088*	,064*	,070*	,058*	,047*	-,040*	-,020*	-,021*	-0,010	-,018*
	%25	%50	,075*	,047*	,050*	,029*	,040*	-,035*	-,017*	-,016*	0,000	-,017*
Bifaktör	%5	%10	,011*	0,005	0,001	,007*	0,003	-,012*	-0,006	0,000	-0,007	-0,003
		%25	,014*	0,005	,006*	,014*	,008*	-,013*	-0,002	-0,004	-,014*	-,008*
		%50	,038*	,023*	,024*	,024*	,018*	-,039*	-,021*	-,022*	-,016*	-,019*
	%10	%25	0,002	0,000	,005*	,007*	,005*	-0,001	0,003	-,004*	-0,007	-,005*
		%50	,027*	,017*	,024*	,017*	,015*	-,027*	-,016*	-,022*	-0,009	-,016*
	%25	%50	,025*	,018*	,018*	,009*	,010*	-,026*	-,019*	-,018*	-0,002	-,011*

*p<,001

Ek 19. Alt Test - Çoklu Puanlanan Madde Yüzdesi x Model x Test Güçlüğü Etkileşimi (r=0,8; ÇPMY basit ana etkisi)

Alt testler için korelasyon 0,8 olduğunda test güçlüğü düzeylerinde her bir yöntem için çoklu puanlanan madde yüzdelerinin karşılaştırılması (RMSE)

		RMSE		Güvenirlik	
Test güçlüğü	Model	F	η_p^2	F	η_p^2
Çok zor	ÇBMTK	725,144*	,577	911,194	,632
	ÜDMTK	720,143*	,576	922,569	,635
	Bifaktör	2,935	-	30,344	,054
Zor	ÇBMTK	315,451*	,373	342,772	,392
	ÜDMTK	316,465*	,373	344,342	,393
	Bifaktör	6,356*	,012	21,177	,038
Orta güçlük	ÇBMTK	417,760*	,440	439,358	,453
	ÜDMTK	425,059*	,445	448,833	,458
	Bifaktör	3,180	-	71,076	,118
Kolay	ÇBMTK	421,562	,443	478,490	,474
	ÜDMTK	422,564	,443	484,719	,477
	Bifaktör	6,233	,012	75,187	,124
Çok kolay	ÇBMTK	418,412	,441	532,376	,501
	ÜDMTK	410,310	,436	537,850	,503
	Bifaktör	22,970	,041	36,520	,064

*p<,001

Alt testler için korelasyon 0,8 olduğunda her bir yöntem için çoklu puanlanan madde yüzdelerinde hesaplanan ortalama RMSE değerleri

		Test güçlüğü					Test güçlüğü				
		Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
Model	Çoklu puanlanan madde yüzdesi	Ortalama RMSE					Ortalama güvenirlik				
ÇBMTK	%5	,446	,369	,331	,336	,376	,895	,929	,943	,942	,926
	%10	,434	,359	,327	,335	,368	,900	,933	,944	,942	,930
	%25	,412	,335	,301	,302	,356	,910	,940	,951	,952	,935
	%50	,346	,295	,259	,279	,324	,938	,953	,963	,959	,947
ÜDMTK	%5	,449	,370	,332	,337	,378	,893	,929	,943	,942	,926
	%10	,437	,360	,328	,336	,370	,899	,932	,944	,942	,930
	%25	,415	,336	,302	,303	,358	,908	,939	,951	,951	,934
	%50	,349	,296	,260	,280	,326	,937	,953	,963	,959	,947
Bifaktör	%5	,925	,920	,921	,921	,919	,410	,436	,446	,442	,431
	%10	,922	,921	,923	,920	,921	,415	,440	,446	,446	,434
	%25	,923	,923	,922	,921	,920	,416	,438	,449	,452	,436
	%50	,925	,923	,922	,922	,924	,431	,450	,463	,459	,444

Alt testler için korelasyon 0,8 olduğunda her bir yöntem için çoklu puanlanan madde yüzdelerinin ikili karşılaştırmaları

			Test güçlüğü					Test güçlüğü				
			Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
Model	(I) Çoklu puanlanan madde yüzdesi	(J) Çoklu puanlanan madde yüzdesi	Ortalama fark (I-J)					Ortalama fark (I-J)				
ÇBMTK	%5	%10	,012*	,010*	0,004	0,001	,008*	-,005*	-,004*	-0,001	0,000	-,004*
		%25	,034*	,034*	,030*	,033*	,021*	-,015*	-,010*	-,008*	-,010*	-,008*
		%50	,099*	,074*	,072*	,057*	,053*	-,043*	-,024*	-,020*	-,017*	-,021*
	%10	%25	,022*	,024*	,026*	,033*	,012*	-,009*	-,007*	-,006*	-,010*	-,004*
		%50	,088*	,064*	,068*	,056*	,044*	-,037*	-,020*	-,018*	-,017*	-,017*
	%25	%50	,066*	,040*	,042*	,023*	,032*	-,028*	-,013*	-,012*	-,007*	-,012*
ÜDMTK	%5	%10	,012*	,010*	0,004	0,001	,008*	-,005*	-,004*	-0,001	0,000	-,004*
		%25	,034*	,034*	,030*	,033*	,020*	-,015*	-,011*	-,008*	-,010*	-,008*
		%50	,100*	,075*	,072*	,057*	,052*	-,044*	-,024*	-,020*	-,017*	-,021*
	%10	%25	,022*	,024*	,026*	,033*	,012*	-,010*	-,007*	-,007*	-,010*	-,004*
		%50	,088*	,064*	,069*	,056*	,044*	-,038*	-,020*	-,018*	-,017*	-,017*
	%25	%50	,065*	,040*	,043*	,023*	,032*	-,029*	-,014*	-,012*	-,007*	-,013*
Bifaktör	%5	%10	0,003	0,00	-0,002	0,001	-0,002	-0,006	-0,004	0,000	-0,004	-0,003
		%25	0,002	0,00	-0,002	0,000	-0,001	-0,006	-0,001	-0,003	-,010*	-,005*
		%50	0,000	0,00	-0,001	-0,001	-,005*	-,021*	-,014*	-,017*	-,017*	-,013*
	%10	%25	-0,001	0,00	0,000	-0,001	0,001	-0,001	0,003	-0,003	-,006*	-0,002
		%50	-0,003	0,00	0,001	-,002*	-,003*	-,015*	-,010*	-,018*	-,013*	-,010*
	%25	%50	-0,002	0,00	0,000	-0,001	-,004*	-,015*	-,013*	-,015*	-,007*	-,008*

*p<,001

Ek 20. Alt Test - Test Güçlüğü x Model x Korelasyon Etkileşimi (Korelasyon basit ana etkisi)

Alt testler için çoklu puanlanan madde yüzdesinin ve test güçlüğünün her bir düzeyi için model ve korelasyon değişkenleri ile iki yönlü karma ANOVA (RMSE)

			RMSE								Güvenirlik					
			Çoklu puanlanan madde yüzdesi								Çoklu puanlanan madde yüzdesi					
Test güçlüğü			%5		%10		%25		%50		%5		%10		%25	
	Kaynak		F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	Gruplar içi	M	68183,935	,983	776138,049	,998	181895,308	,993	252197,720	,995	127465,714	,991	271387,174	,996	231349,034	,995
		M x AT	358,984	,474	708,301	,640	472,773	,543	820,999	,674	1231,542	,756	791,471	,665	1292,154	,765
		M *x* K	11713,174	,952	152866,492	,996	34383,296	,983	43132,694	,986	29694,872	,980	74246,274	,992	60669,199	,990
	Gruplar arası	AT	2291,339	,852	4327,837	,916	5691,873	,935	4545,503	,919	5010,074	,926	17798,673	,978	21334,061	,982
		K	617,218	,508	1238,211	,675	840,579	,585	934,154	,610	8711,578	,936	28173,588	,979	21461,194	,973
Zor	Gruplar içi	M	2608166,677	1,000	854232,342	,999	57154,028	,980	1083763,901	,999	1685479,257	,999	898673,907	,999	107218,668	,989
		M x AT	921,334	,698	527,064	,570	386,480	,493	5289,174	,930	557,004	,583	1070,121	,729	1073,834	,730
		M *x* K	493863,272	,999	153912,664	,996	8533,462	,935	172283,309	,997	470364,498	,999	245649,503	,998	24791,507	,976
	Gruplar arası	AT	4566,800	,920	3788,348	,905	4038,387	,910	7442,885	,949	16412,174	,976	19229,537	,980	6895,070	,945
		K	13026,947	,956	6187,391	,912	1395,927	,700	3460,431	,853	240650,589	,998	151697,692	,996	14923,276	,962
Orta	Gruplar içi	M	279299,328	,996	46742,757	,975	64390,309	,982	542400,717	,998	834848,988	,999	80105,818	,985	123804,338	,990
		M x AT	337,627	,459	373,586	,484	413,171	,509	1833,990	,822	1045,887	,724	804,739	,669	891,406	,691
		M *x* K	48492,612	,988	6484,745	,916	9209,761	,939	80248,401	,993	226742,369	,997	18026,680	,968	29402,969	,980
	Gruplar arası	AT	2908,127	,880	1323,052	,769	4049,973	,911	8006,662	,953	9354,170	,959	2054,144	,838	6140,587	,939
		K	9495,191	,941	1959,645	,766	2537,861	,810	6377,972	,914	165872,857	,996	13509,670	,958	24651,893	,976
Kolay	Gruplar içi	M	64905,190	,982	1916154,593	,999	366082,680	,997	1572488,071	,999	150584,015	,992	1002462,825	,999	827399,883	,999
		M x AT	288,583	,420	1175,317	,747	923,593	,699	5431,598	,932	857,332	,683	471,540	,542	1181,922	,748
		M *x* K	10096,018	,944	336969,941	,998	59868,499	,990	249506,893	,998	36832,516	,984	278815,605	,998	222751,178	,997
	Gruplar arası	AT	813,525	,671	3863,880	,907	6898,883	,945	9700,993	,961	1555,783	,796	18659,228	,979	33176,933	,988
		K	3155,024	,841	10621,097	,947	7971,073	,930	11416,537	,950	25029,798	,977	263816,392	,998	244505,562	,998
Çok kolay	Gruplar içi	M	1671479,755	,999	79888,434	,985	1297680,449	,999	1013315,731	,999	1205859,618	,999	220686,625	,995	678230,018	,998
		M x AT	1116,141	,737	250,509	,386	2505,941	,863	2684,193	,871	565,562	,587	1242,170	,757	480,275	,547
		M *x* K	325896,441	,998	13088,787	,956	237648,887	,997	169329,941	,996	342364,898	,998	55098,052	,989	188828,383	,997
	Gruplar arası	AT	5291,318	,930	1546,125	,795	7404,053	,949	6441,165	,942	13608,769	,972	3639,531	,901	37884,336	,990
		K	7155,444	,923	2174,958	,785	6259,483	,913	5322,087	,899	95692,927	,994	24002,685	,976	189666,168	,997

*p<,001

Alt testler için çoklu puanlanan madde yüzdesinin ve test güçlüğünün bütün düzeylerinde her bir yöntem için korelasyon düzeylerinin karşılaştırılması (RMSE, güvenirlik)

		RMSE								Güvenirlik					
		Çoklu puanlanan madde yüzdesi								Çoklu puanlanan madde yüzdesi					
		%5		%10		%25		%50		%5		%10		%25	
Test güçlüğü	Model	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2
Çok zor	ÇBMTK	4974,441	,893	1466,451	,711	2328,436	,796	746,779	,556	4279,600	,878	1563,020	,724	2256,129	,791
	ÜDMTK	4918,323	,892	1425,338	,705	2303,083	,794	716,169	,545	4222,897	,876	1525,641	,719	2236,678	,789
	Bifaktör	5127,424	,896	23959,460	,976	9675,047	,942	8731,044	,936	23606,133	,975	317119,318	,998	98860,223	,994
Zor	ÇBMTK	7543,690	,927	1981,651	,768	2195,538	,786	845,396	,586	6315,471	,914	1954,941	,766	2039,740	,774
	ÜDMTK	7413,392	,925	1955,728	,766	2183,614	,785	857,870	,590	6193,864	,912	1933,201	,764	2044,630	,774
	Bifaktör	123061,758	,995	47697,137	,988	4902,116	,891	28437,935	,979	799265,687	,999	577756,693	,999	23133,208	,975
Orta güçlük	ÇBMTK	6922,659	,921	1106,752	,650	2418,105	,802	983,925	,622	5620,560	,904	1248,757	,677	1971,664	,768
	ÜDMTK	6893,346	,920	1101,329	,648	2395,236	,800	989,279	,624	5565,311	,903	1246,440	,676	1961,138	,767
	Bifaktör	29450,528	,980	4636,092	,886	6255,466	,913	28899,893	,980	244866,425	,998	17185,866	,966	29429,276	,980
Kolay	ÇBMTK	12159,235	,953	1559,220	,723	2722,284	,820	1439,558	,707	13647,838	,958	1720,278	,742	2229,856	,789
	ÜDMTK	12008,831	,953	1537,863	,720	2672,365	,817	1421,803	,704	13568,011	,958	1705,342	,741	2219,314	,788
	Bifaktör	7132,523	,923	107071,920	,994	30627,930	,981	67573,728	,991	32925,603	,982	643263,081	,999	344237,541	,998
Çok kolay	ÇBMTK	6647,599	,918	2592,799	,813	3560,461	,856	1389,796	,700	3871,667	,866	2827,063	,826	3109,003	,839
	ÜDMTK	6420,764	,915	2531,441	,809	3353,449	,849	1313,895	,688	3785,128	,864	2796,867	,824	3049,069	,836
	Bifaktör	73064,182	,992	7704,983	,928	56566,002	,990	38061,386	,985	621711,053	,999	46435,594	,987	410325,749	,999

*p<,001

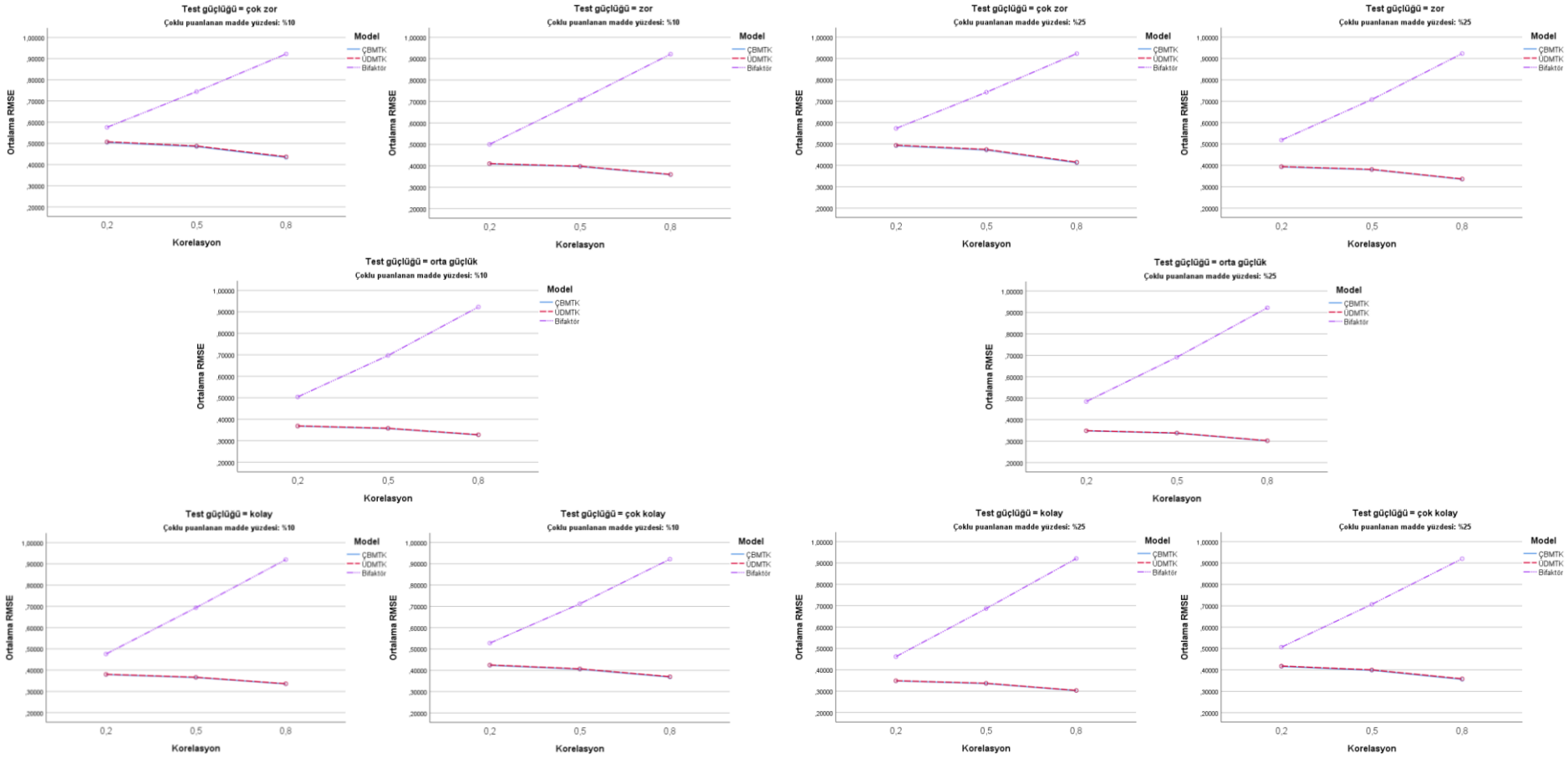
Toplam test için çoklu puanlanan madde yüzdesi, test güçlüğü, model ve korelasyon düzeylerinde elde edilen ortalama RMSE ve güvenirlik değerleri

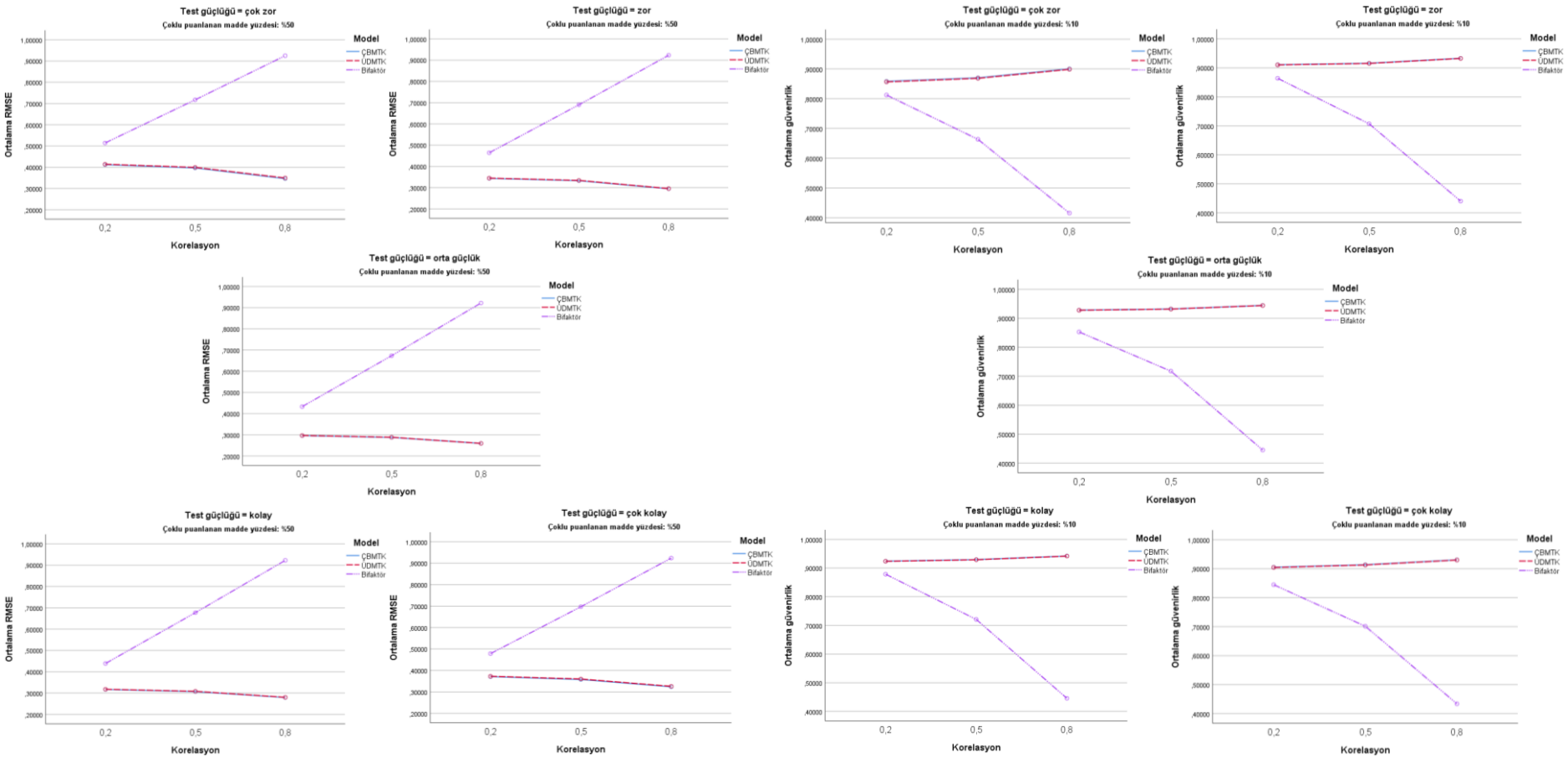
			Test güçlüğü					Test güçlüğü				
			Çok zor	Zor	Orta	Kolay	Çok kolay	Çok zor	Zor	Orta	Kolay	Çok kolay
ÇPMY	Model	Korelasyon	Ortalama RMSE					Ortalama güvenirlik				
%5	ÇBMTK	0,2	,525	,429	,381	,388	,444	,849	,902	,924	,921	,894
		0,5	,504	,416	,370	,374	,424	,861	,908	,928	,927	,905
		0,8	,446	,369	,331	,336	,376	,895	,929	,943	,942	,926
	ÜDMTK	0,2	,528	,431	,382	,389	,446	,847	,902	,923	,921	,893
		0,5	,507	,417	,371	,375	,426	,860	,908	,928	,927	,904
		0,8	,449	,370	,332	,337	,378	,893	,929	,943	,942	,926
	Bifaktör	0,2	,615	,513	,482	,506	,523	,781	,857	,875	,858	,851
		0,5	,756	,713	,697	,701	,715	,651	,701	,718	,714	,698
		0,8	,925	,920	,921	,921	,919	,410	,436	,446	,442	,431
%10	ÇBMTK	0,2	,505	,409	,368	,379	,423	,858	,911	,928	,924	,905
		0,5	,485	,397	,357	,365	,405	,870	,916	,932	,930	,914
		0,8	,434	,359	,327	,335	,368	,900	,933	,944	,942	,930
	ÜDMTK	0,2	,507	,410	,369	,380	,425	,857	,910	,928	,924	,904
		0,5	,488	,398	,358	,366	,407	,869	,915	,932	,929	,913
		0,8	,437	,360	,328	,336	,370	,899	,932	,944	,942	,930
	Bifaktör	0,2	,576	,500	,504	,476	,528	,812	,864	,853	,879	,845
		0,5	,744	,707	,697	,694	,712	,663	,707	,718	,721	,701
		0,8	,922	,921	,923	,920	,921	,415	,440	,446	,446	,434
%25	ÇBMTK	0,2	,492	,393	,348	,348	,416	,864	,914	,933	,935	,907
		0,5	,472	,380	,337	,336	,398	,875	,920	,937	,939	,915
		0,8	,412	,335	,301	,302	,356	,910	,940	,951	,952	,935
	ÜDMTK	0,2	,494	,394	,349	,349	,418	,862	,913	,933	,934	,906
		0,5	,474	,381	,338	,337	,400	,874	,919	,937	,939	,915
		0,8	,415	,336	,302	,303	,358	,908	,939	,951	,951	,934
	Bifaktör	0,2	,573	,518	,485	,462	,506	,811	,842	,865	,885	,861
		0,5	,742	,708	,691	,687	,707	,664	,704	,722	,728	,707
		0,8	,923	,923	,922	,921	,920	,416	,438	,449	,452	,436
%50	ÇBMTK	0,2	,412	,344	,296	,317	,371	,901	,932	,950	,945	,925
		0,5	,397	,333	,288	,307	,358	,910	,936	,953	,939	,932
		0,8	,346	,295	,259	,279	,324	,938	,953	,963	,959	,947
	ÜDMTK	0,2	,414	,345	,297	,318	,373	,900	,931	,950	,945	,924
		0,5	,399	,334	,288	,308	,360	,909	,936	,952	,939	,931
		0,8	,349	,296	,260	,280	,326	,937	,953	,963	,959	,947
	Bifaktör	0,2	,514	,464	,433	,439	,478	,846	,881	,897	,897	,876
		0,5	,717	,690	,673	,677	,697	,690	,723	,740	,730	,717
		0,8	,925	,923	,922	,922	,924	,431	,450	,463	,459	,444

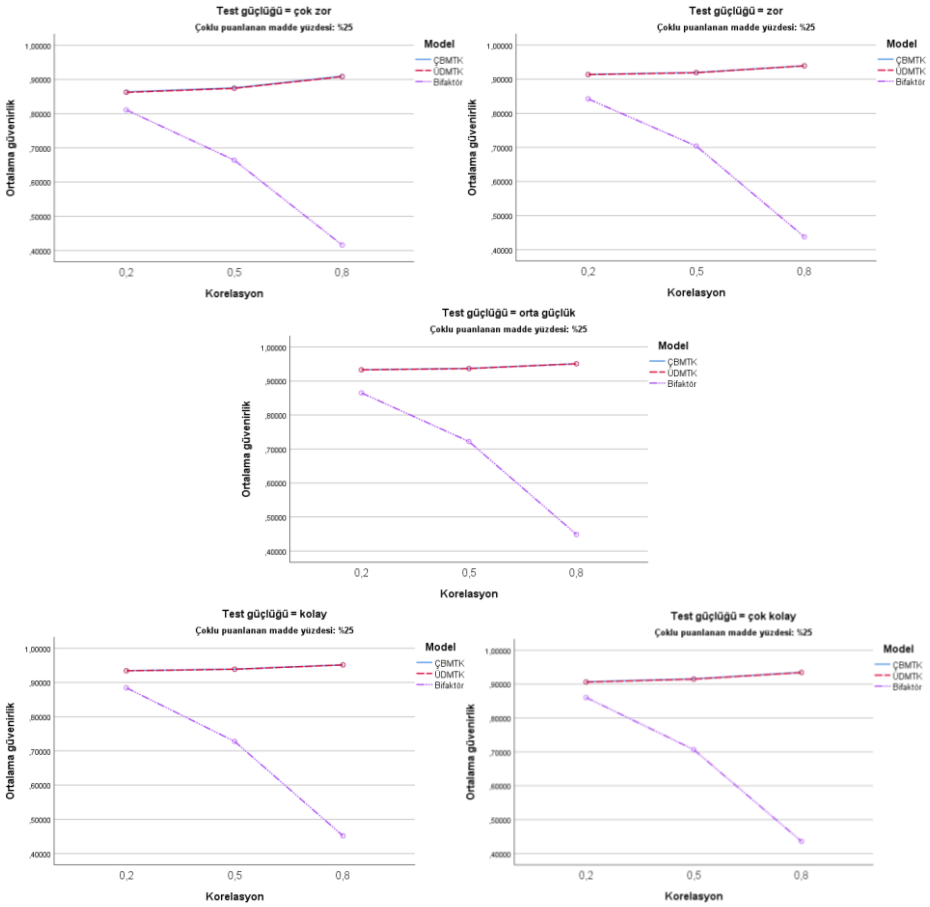
Toplam test için çoklu puanlanan madde yüzdesinin ve test güçlüğünün bütün düzeylerinde her bir yöntem için korelasyon düzeylerinin ikili karşılaştırmaları (RMSE, güvenirlik)

				RMSE					Güvenirlik				
				Test güçlüğü					Test güçlüğü				
				Çok zor	Zor	Orta güçlük	Kolay	Çok kolay	Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
ÇPMY	Model	(I) Korelasyon	(J) Korelasyon	Ortalama fark (I-J)					Ortalama fark (I-J)				
%5	ÇBMTK	0,2	0,5	,021*	,013*	,012*	,014*	,020*	-,012*	-,006*	-,004*	-,006*	-,010*
			0,8	,079*	,060*	,050*	,052*	,068*	-,046*	-,027*	-,019*	-,021*	-,032*
		0,5	0,8	,059*	,047*	,038*	,038*	,048*	-,034*	-,021*	-,015*	-,015*	-,022*
	ÜDMTK	0,2	0,5	,021*	,013*	,012*	,014*	,020*	-,012*	-,006*	-,004*	-,006*	-,010*
			0,8	,079*	,060*	,050*	,052*	,068*	-,046*	-,027*	-,020*	-,021*	-,032*
		0,5	0,8	,058*	,047*	,039*	,038*	,048*	-,034*	-,021*	-,015*	-,015*	-,022*
	Bifaktör	0,2	0,5	-,141*	-,199*	-,215*	-,195*	-,192*	,129*	,156*	,157*	,144*	,152*
			0,8	-,310*	-,406*	-,439*	-,415*	-,396*	,371*	,421*	,429*	,416*	,420*
		0,5	0,8	-,169*	-,207*	-,223*	-,220*	-,204*	,241*	,265*	,272*	,272*	,268*
%10	ÇBMTK	0,2	0,5	,020*	,012*	,011*	,014*	,018*	-,012*	-,005*	-,004*	-,005*	-,009*
			0,8	,071*	,050*	,041*	,044*	,055*	-,042*	-,022*	-,016*	-,018*	-,025*
		0,5	0,8	,051*	,038*	,030*	,030*	,037*	-,030*	-,017*	-,012*	-,013*	-,017*
	ÜDMTK	0,2	0,5	,020*	,012*	,011*	,014*	,018*	-,012*	-,005*	-,004*	-,005*	-,009*
			0,8	,070*	,050*	,041*	,044*	,055*	-,042*	-,022*	-,016*	-,018*	-,026*
		0,5	0,8	,050*	,038*	,030*	,030*	,037*	-,030*	-,017*	-,012*	-,013*	-,017*
	Bifaktör	0,2	0,5	-,169*	-,207*	-,193*	-,218*	-,184*	,149*	,157*	,135*	,158*	,143*
			0,8	-,346*	-,421*	-,418*	-,444*	-,394*	,397*	,424*	,407*	,433*	,411*
		0,5	0,8	-,178*	-,214*	-,226*	-,226*	-,209*	,248*	,267*	,272*	,275*	,268*
%25	ÇBMTK	0,2	0,5	,020*	,013*	,011*	,012*	,018*	-,012*	-,005*	-,004*	-,005*	-,009*
			0,8	,080*	,058*	,047*	,045*	,060*	-,046*	-,025*	-,018*	-,017*	-,028*
		0,5	0,8	,060*	,045*	,036*	,033*	,042*	-,034*	-,020*	-,014*	-,013*	-,019*
	ÜDMTK	0,2	0,5	,020*	,013*	,011*	,012*	,018*	-,012*	-,006*	-,004*	-,005*	-,009*
			0,8	,080*	,058*	,047*	,046*	,060*	-,046*	-,026*	-,018*	-,017*	-,028*
		0,5	0,8	,060*	,045*	,036*	,033*	,042*	-,034*	-,020*	-,014*	-,013*	-,019*
	Bifaktör	0,2	0,5	-,169*	-,190*	-,207*	-,225*	-,200*	,147*	,138*	,143*	,157*	,154*
			0,8	-,350*	-,405*	-,438*	-,459*	-,414*	,395*	,405*	,416*	,433*	,425*
		0,5	0,8	-,181*	-,215*	-,231*	-,234*	-,214*	,248*	,266*	,273*	,276*	,271*
%50	ÇBMTK	0,2	0,5	,015*	,011*	,008*	,010*	,013*					
			0,8	,066*	,049*	,037*	,038*	,047*					
		0,5	0,8	,050*	,038*	,028*	,028*	,034*					
	ÜDMTK	0,2	0,5	,015*	,011*	,008*	,009*	,013*					
			0,8	,065*	,049*	,037*	,038*	,047*					
		0,5	0,8	,050*	,038*	,029*	,028*	,034*					
	Bifaktör	0,2	0,5	-,203*	-,226*	-,240*	-,238*	-,219*					
			0,8	-,411*	-,459*	-,489*	-,484*	-,446*					
		0,5	0,8	-,208*	-,233*	-,249*	-,245*	-,227*					

*p<,001


Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %10
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %25
Ortalama RMSE
Korelasyon
0,2
0,5
0,8
Model
ÇBMTK
ÜDMTK
Bifaktör


Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %50
Ortalama RMSE
Korelasyon
Model
ÇBMTK
ÜDMTK
Bifaktör
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %50
Ortalama RMSE
Korelasyon
Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %10
Ortalama güvenirlik
Korelasyon
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %10
Ortalama güvenirlik
Korelasyon
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %50
Ortalama RMSE
Korelasyon
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %10
Ortalama güvenirlik
Korelasyon
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %50
Ortalama RMSE
Korelasyon
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %50
Ortalama RMSE
Korelasyon
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %10
Ortalama güvenirlik
Korelasyon
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %10
Ortalama güvenirlik
Korelasyon


Test güçlüğü = çok zor
Çoklu puanlanan madde yüzdesi: %25
Model
ÇBMTK
ÜDMTK
Bifaktör
Ortalama güvenirlik
Korelasyon
Test güçlüğü = zor
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = orta güçlük
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = kolay
Çoklu puanlanan madde yüzdesi: %25
Test güçlüğü = çok kolay
Çoklu puanlanan madde yüzdesi: %25

Ek 21. Alt Test - Çoklu Puanlanan Madde Yüzdesi %50 için Model x Korelasyon Etkileşimi -Güvenirlik

Alt test model *x* korelasyon etkileşim etkisi için basit ana etkiler (güvenirlik)

		F	Kısmi η^2
	Model		
Korelasyon basit ana etkisi	ÇBMTK	349,049	,105
	ÜDMTK	354,227	,106
	Bifaktör	127056,885	,977
	Korelasyon		
Model basit ana etkisi	0,2	11225,409	,790
	0,5	176081,884	,983
	0,8	977353,019	,997

*p<,001

Çoklu puanlanan madde yüzdesi %50 iken model ve korelasyon düzeylerine göre elde edilen ortalama güvenirlik değerleri

Model	Korelasyon	Ortalama güvenirlik
ÇBMTK	0,2	,931
	0,5	,934
	0,8	,952
ÜDMTK	0,2	,930
	0,5	,933
	0,8	,952
Bifaktör	0,2	,879
	0,5	,720
	0,8	,449

Çoklu puanlanan madde yüzdesi %50 iken Korelasyon basit ana etkisi - her model için korelasyon düzeylerinin ikili karşılaştırmaları

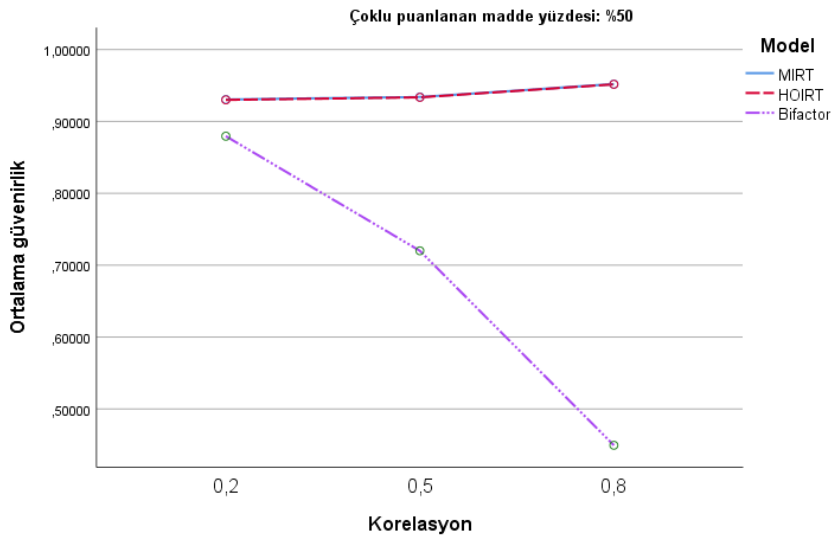
		ÇBMTK	ÜDMTK	Bifaktör
(I) korelasyon	(J) korelasyon	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)
0,2	0,5	-,003*	-,003*	,159*
	0,8	-,021*	-,022*	,430*
0,5	0,8	-,018*	-,018*	,271*

*p<,001

Çoklu puanlanan madde yüzdesi %50 iken Model basit ana etkisi - her korelasyon düzeyi için modellerin ikili karşılaştırmaları

		Korelasyon		
		0,2	0,5	0,8
(I) Model	(J) Model	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)
ÇBMTK	ÜDMTK	,001*	,001*	,000*
	Bifaktör	,051*	,214*	,502*
ÜDMTK	Bifaktör	,051*	,213*	,502*
Bifaktör	ÜDMTK	-,051*	-,213*	-,502*

*p<,001



Ek 21. Alt Test – Model x Korelasyon Etkileşimi (sınıflama doğruluğu)

Model ve korelasyon düzeylerine göre elde edilen ortalama sınıflama doğruluğu değerleri

Model	Korelasyon	Ortalama sınıflama doğruluğu
ÇBMTK	0,2	,895
	0,5	,898
	0,8	,911
ÜDMTK	0,2	,895
	0,5	,898
	0,8	,911
Bifaktör	0,2	,844
	0,5	,804
	0,8	,756

Korelasyon basit ana etkisi - her model için korelasyon düzeylerinin ikili karşılaştırmaları

		ÇBMTK	ÜDMTK	Bifaktör
(I) korelasyon	(J) korelasyon	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)
0,2	0,5	-,003*	-,003*	,040*
	0,8	-,016*	-,016*	,088*
0,5	0,8	-,014*	-,013*	,048*

*p<,001

Model basit ana etkisi - her korelasyon düzeyi için modellerin ikili karşılaştırmaları

		Korelasyon		
		0,2	0,5	0,8
(I) Model	(J) Model	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)
ÇBMTK	ÜDMTK	-,001*	,000*	0,00
	Bifaktör	,051*	,094*	,156*
ÜDMTK	Bifaktör	,051*	,094*	,156*

*p<,001

Ek 22. Alt Test – Model x Test Güçlüğü Etkileşimi (sınıflama doğruluğu)

Model ve test güçlüğü düzeylerine göre elde edilen ortalama sınıflama doğruluğu değerleri

		ÇBMTK	ÜDMTK	Bifaktör
Model	Test güçlüğü	Ortalama sınıflama doğruluğu	Ortalama sınıflama doğruluğu	Ortalama sınıflama doğruluğu
	Çok zor	,856	,857	,765
	Zor	,894	,894	,796
	Orta güçlük	,916	,916	,814
	Kolay	,923	,923	,819
	Çok kolay	,917	,917	,812

Test güçlüğü ana etkisi - her model için test güçlüğü düzeylerinin ikili karşılaştırmaları

		Modeller		
		ÇBMTK	ÜDMTK	Bifaktör
(I) test güçlüğü	(J) test güçlüğü	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)
Çok zor	Zor	-,038*	-,038*	-,031*
	Orta güçlük	-,060*	-,059*	-,049*
	Kolay	-,067*	-,067*	-,054*
	Çok kolay	-,061*	-,061*	-,047*
Zor	Orta güçlük	-,022*	-,022*	-,018*
	Kolay	-,029*	-,029*	-,023*
	Çok kolay	-,023*	-,023*	-,016*
Orta güçlük	Kolay	-,007*	-,007*	-,005*
	Çok kolay	-,001*	-,002*	,002*
Kolay	Çok kolay	,006*	,006*	,007*

*p<,001

Model ana etkisi - her test güçlüğü düzeyi için modellerin ikili karşılaştırmaları

		Çok zor	Zor	Orta güçlük	Kolay	Çok kolay
(I) Model	(J) Model	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)
ÇBMTK	ÜDMTK	-,001*	,000*	,000*	,000*	,000*
	Bifaktör	,091*	,098*	,102*	,104*	,105*
ÜDMTK	Bifaktör	,092*	,098*	,102*	,105*	,106*

*p<,001

Ek 23. Alt Test – Model x Çoklu Puanlanan Madde Yüzdesi Etkileşimi (sınıflama doğruluğu)

Model ve çoklu puanlanan madde yüzdesi düzeylerine göre elde edilen ortalama sınıflama doğruluğu değerleri

	ÇBMTK	ÜDMTK	Bifaktör
Test güçlüğü	Ortalama sınıflama doğruluğu	Ortalama sınıflama doğruluğu	Ortalama sınıflama doğruluğu
%5	,890	,890	,792
%10	,895	,896	,797
%25	,902	,902	,801
%50	,918	,918	,814

Çoklu puanlanan madde yüzdesi ana etkisi - her model için çoklu puanlanan madde yüzdesi düzeylerinin ikili karşılaştırmaları

		Modeller		
		ÇBMTK	ÜDMTK	Bifaktör
(I) çoklu puanlanan madde yüzdesi	(J) çoklu puanlanan madde yüzdesi	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)
%5	%10	-,006*	-,006*	-,005*
	%25	-,012*	-,012*	-,009*
	%50	-,028*	-,028*	-,022*
%10	%25	-,006*	-,006*	-,004*
	%50	-,023*	-,022*	-,017*
%25	%50	-,016*	-,016*	-,013*

*p<,001

Model ana etkisi - her çoklu puanlanan madde yüzdesi düzeyi için modellerin ikili karşılaştırmaları

		%5	%10	%25	%50
(I) Model	(J) Model	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)
ÇBMTK	ÜDMTK	,000*	,000*	,000*	,000*
	Bifaktör	,098*	,098*	,100*	,104*
ÜDMTK	Bifaktör	,098*	,099*	,101*	,104*

*p<,001

Ek 24. Alt Test – Model x Alt Test Etkileşimi (sınıflama doğruluğu)

Model ve alt testlere göre elde edilen ortalama sınıflama doğruluğu değerleri

	ÇBMTK	ÜDMTK	Bifaktör
Test güçlüğü	Ortalama sınıflama doğruluğu	Ortalama sınıflama doğruluğu	Ortalama sınıflama doğruluğu
Alt test 1	,886	,886	,770
Alt test 2	,886	,887	,797
Alt test 3	,899	,899	,806
Alt test 4	,934	,934	,831

Alt test ana etkisi - her model için alt testlerin ikili karşılaştırmaları

		Modeller		
		ÇBMTK	ÜDMTK	Bifaktör
(I) alt test	(J) alt test	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)
Alt test 1	Alt test 2	0,000	0,000	-,027*
	Alt test 3	-,013*	-,013*	-,036*
	Alt test 4	-,048*	-,048*	-,061*
Alt test 2	Alt test 3	-,013*	-,013*	-,009*
	Alt test 4	-,048*	-,048*	-,034*
Alt test 3	Alt test 4	-,035*	-,035*	-,025*
*p<,001				

Model ana etkisi – alt testler için modellerin ikili karşılaştırmaları

		Alt test 1	Alt test 2	Alt test 3	Alt test 4
(I) Model	(J) Model	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)	Ortalama farkı (I-J)
ÇBMTK	ÜDMTK	,000*	,000*	,000*	,000*
	Bifaktör	,116*	,089*	,093*	,103*
ÜDMTK	Bifaktör	,116*	,089*	,093*	,103*
*p<,001					

GAZİLİ OLMAK AYRICALIKTIR