



**KARMA MTK MODELLERİNİN FARKLI KOŞULLARDA
PARAMETRE KESTİRİMİNE VE SINIFLAMA DOĞRULUĞUNA
ETKİSİNİN SİMÜLATİF VE GERÇEK VERİ İLE İNCELENMESİ**

Fatıma Münevver Saatçioğlu

**DOKTORA TEZİ
EĞİTİM BİLİMLERİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

Mart, 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 : Fatıma Münevver

Soyadı : SAATÇİOĞLU

Bölümü : Eğitim Bilimleri

İmza :

Teslim Tarihi :

TEZİN

Türkçe Adı: Karma MTK Modellerinin Farklı Koşullarda Parametre Kestirimine ve Sınıflama Doğruluğuna Etkisinin Simülatif ve Gerçek Veri ile İncelenmesi

İngilizce Adı: Investigation of The Effect of Mixture IRT Models Under Different Conditions On Parameter Recovery and Classification Accuracy With Simulative and Real Data

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: F. Münevver SAATÇIOĞLU

İmza :

JÜRİ ONAY SAYFASI

F.Münevver SAATÇİOĞLU tarafından hazırlanan “Karma MTK Modellerinin Farklı Koşullarda Parametre Kestirimine ve Sınıflama Doğruluğuna Etkisinin Simülatif ve Gerçek Veri ile İncelenmesi” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı’nda Doktora tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Hakan Yavuz ATAR

Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi

Başkan: Prof. Dr. Duygu ANIL

Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Hacettepe Üniversitesi

Üye: Prof. Dr. Şeref TAN

Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi

Üye: Prof. Dr. İsmail KARAKAYA

Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi

Üye: Doç. Dr. Sedat ŞEN

Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Harran Üniversitesi

Tez Savunma Tarihi: 12/02/2021

Bu tezin Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı’nda Doktora tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Yücel GELİŞLİ

Eğitim Bilimleri Enstitüsü Müdürü

.....

TEŞEKKÜR

Doktora sürecimden itibaren değerli önerileriyle ufkumu açan ve beni geliştiren kıymetli hocam Prof. Dr. Hakan Yavuz ATAR'a,
Derin bilgisi, hoşgörüsü ve tevazusuyla desteklerini esirgemeyen saygıdeğer hocam Prof. Dr. Şeref TAN'a,
Pozitif enerjisi ve güler yüzüyle içimi ferahlatan, simülasyon konusunda verdiği fikirlerle tezime destek olan sevgili hocam Prof. Dr. Duygu ANIL'a,
Doktoraya başladığım ilk günden itibaren bana yol gösteren ve tezime katkılarını sunan kıymetli hocam Prof. Dr. İsmail KARAKAYA'ya,
Simülasyon konusundaki engin bilgisi, hoşgörülü tavırları ile her sorumu sabırla yanıtlayıp tezimi üst noktalara taşıyan değerli hocam Doç. Dr. Sedat ŞEN'e,
Bu tezi bitirebilmemde bana her daim destek olan, haklarını asla ödeyemeyeceğim canım annem ve babama,
Tezimi yazabilmem için tüm enerjisiyle bana destek olan kıymetli eşim Özkan'a,
Varlığını yüreğimde hissedip güç bulduğum oğlum Umut Selim'e,
Yanımda olan ve olamayan tüm arkadaşlarıma,
Tezimi bitirebileceğime gönülden inanan herkese,
Sonsuz teşekkürlerimi sunarım...

**KARMA MTK MODELLERİNİN FARKLI KOŞULLARDA
PARAMETRE KESTİRİMİNE VE SINIFLAMA DOĞRULUĞUNA
ETKİSİNİN SİMÜLATİF VE GERÇEK VERİ İLE İNCELENMESİ
(Doktora Tezi)**

**Fatıma Münevver Saatçioğlu
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
Mart, 2021**

ÖZ

Araştırmanın amacı Karma MTK modellerinin farklı koşullarında madde parametre kestirimine ve sınıflama doğruluğuna etkisinin simülatif ve gerçek veri ile incelenmesidir. Simülasyon çalışmasında Karma MTK modelleri (Rasch, 2PL, 3PL), örneklem büyüklüğü (600, 1000), madde sayısı (10, 30), örtük sınıf sayısı (2 ,3) kayıp veri türü (tam, rastgele kayıp veri ve rastgele olmayan kayıp veri) ve kayıp veri yüzdesi (%10, %20) değişkenleri manipüle edilen değişkenlerdir. Verilerin üretilmesinde R programı kullanılmıştır. Verilerin analizinde R programı ile Mplus programının otomasyonuna imkan sağlayan “Mplus Automation” paketi kullanılmıştır. Araştırma sonucunda madde güçlük, madde ayırt edicilik ve şans parametre kestirimleri için ortalama RMSE ve yanlılık değerleri elde edilmiştir. Madde parametrelerinin kestiriminde seçilen faktörlerin etkileşim etkilerini incelemek için çok yönlü varyans analizi gerçekleştirilmiştir. Seçilen faktörlerin etkileşim etkileri incelendiğinde örneklem büyüklüğü, madde sayısı, örtük sınıf sayısı, kayıp veri türü ve kayıp veri yüzdesi faktörlerinin Karma MTK modelleri için madde parametrelerinin kestiriminde önemli birer faktör oldukları söylenebilir. Karma Rasch model için elde edilen ortalama RMSE değerlerinin Karma 2PL ve Karma 3PL modelden daha düşük olduğu gözlenmiştir. Ayrıca ele alınan faktörlerin kombinasyonundan elde edilen veriler için sınıflama doğruluğu yüzdeleri hesaplanmıştır. En yüksek sınıflama doğruluğu yüzdesine Karma Rasch modelin 30 madde, 2 sınıf, 1000 örneklem büyüklüğü ve tam veri koşullarında ulaşıldığı görülmüştür. Son olarak çalışmada TIMSS 2015 Türkiye 8.sınıf fen alt testinin verisi kullanılarak simülasyon veri gerçek veri ile desteklenmiştir. Gerçek veriden elde edilen sonuçlar ile simülasyon veriden elde edilen sonuçların birbiriyle tutarlı olduğu tespit edilmiştir.

Anahtar Kelimeler : Karma MTK Modelleri, Örtük Sınıf, TIMSS 2015, Kayıp veri,
Sınıflama doğruluğu,
Sayfa Adedi : xviii + 124
Danışman : Prof. Dr. Hakan Yavuz ATAR

**INVESTIGATION OF THE EFFECT OF MIXTURE IRT MODELS
UNDER DIFFERENT CONDITIONS ON PARAMETER RECOVERY
AND CLASSIFICATION ACCURACY WITH SIMULATIVE AND
REAL DATA (Ph.D Thesis)**

Fatıma Münevver Saatçioğlu

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

March, 2021

ABSTRACT

The aim of the study is to examine the effect of mixture IRT models on item parameter estimation and classification accuracy for different conditions with simulative and real data. In the simulation study, Mixture IRT models (Rasch, 2PL, 3PL), sample size (600, 1000), number of items (10, 30), number of latent classes (2, 3) missing data type (complete, missing at random (MAR) and missing not at random (MNAR)) and percentage of missing data (10%, 20%) variables are manipulated variables. R program was used to generate the data. In the analysis of the data, the "Mplus Automation" package, which allows the automation of the R program and the Mplus program, was used. As a result of the research, mean RMSE and bias values were obtained for item difficulty, item discrimination and pseudo-chance parameter estimation. Multi-way ANOVA was performed to examine the interaction effects of selected factors in the estimation of item parameters. When the interaction effects of the selected factors are examined, it can be said that the factors of sample size, number of items, latent classes, missing data type, and percentage of missing data are important factors in estimating item parameters for Mixture IRT models. It was observed that the mean RMSE values obtained for the Mixture Rasch model were lower than the Mixture 2PL and Mixture 3PL models. In addition, classification accuracy percentages were calculated for the data obtained from the combination of the considered factors. It was observed that the highest classification accuracy percentage was reached in the Mixture Rasch model with 30 item, 2 class, 1000 sample size and complete data conditions. Finally, simulation data was supported using TIMSS 2015 8th science test data for Turkey. It has been determined that the results obtained from the real data and the results obtained from the simulation data are consistent with each other.

Key Words : Mixture IRT Models, Latent Class, TIMSS 2015, Missing Data,
Classification Accuracy

Page Number : xviii + 124

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.....	iv
ÖZ	v
ABSTRACT	vii
İÇİNDEKİLER.....	ix
TABLolar LİSTESİ.....	xiii
ŞEKİLLER LİSTESİ.....	xvi
BÖLÜM 1	1
GİRİŞ.....	1
Problem Durumu	1
Araştırmanın Amacı	5
Araştırmanın Önemi.....	6
Araştırmanın Sınırlılıkları	7
BÖLÜM 2.....	8
KAVRAMSAL ÇERÇEVE	8
Madde Tepki Kuramı Modelleri.....	8

Örtük Sınıf Analizi (ÖSA).....	10
Karma MTK Modelleri	13
Karma Rasch model.....	14
Karma 2PL ve 3PL MTK Modelleri	15
İlgili Araştırmalar	16
BÖLÜM 3.....	21
YÖNTEM.....	21
Araştırmanın Türü.....	21
Simülasyon Koşulları	21
Örtük Sınıf Sayısı.....	22
Örneklem Büyüklüğü	23
Madde sayısı	23
Madde ve Yetenek Parametrelerinin Dağılımı.....	23
Kayıp veri.....	26
Karma MTK Modelleri İçin Kestirim Yöntemleri	27
Etiket Değiştirme (Label Switching)	29
Simülasyon Veri Analizi	30
Parametre Doğrulanması	30
Ana Etkiler ve Etkileşim Etkileri İçin ANOVA	32
İstatistiksel Anlamlılık ve Etki Büyüklüğü	35
Sınıflama Doğruluğu	36
Gerçek Veri.....	37
Verilerin Elde Edilmesi.....	37
Gerçek Veri Analizi.....	38
Model Veri Uyumunun Belirlenmesi.....	38

BÖLÜM 4.....	40
BULGULAR	40
Madde Güçlük Parametrelerine Ait Bulgular.....	40
Madde Güçlük Parametrelerinden Elde Edilen Ortalama RMSE Değerleri İçin ANOVA Sonuçları.....	44
Madde Güçlük Parametrelerinden Elde Edilen Ortalama Yanlılık Değerleri İçin ANOVA Sonuçları.....	57
Madde Ayırt Edicilik Parametrelerine Ait Bulgular	62
Madde Ayırt Edicilik Parametrelerinden Elde Edilen Ortalama RMSE Değerleri İçin ANOVA Sonuçları.....	64
Madde Ayırt Edicilik Parametrelerinden Elde Edilen Ortalama Yanlılık Değerleri İçin ANOVA Sonuçları.....	77
Şans Parametrelerine Ait Bulgular	79
Şans Parametrelerinden Elde Edilen Ortalama RMSE Değerleri İçin ANOVA Sonuçları	81
Şans Parametrelerinden Elde Edilen Ortalama Yanlılık Değerleri İçin ANOVA Sonuçları	82
Sınıflama Doğruluğu.....	83
Gerçek Veri.....	86
BÖLÜM 5.....	93
TARTIŞMA, SONUÇ VE ÖNERİLER	93
Tartışma ve Sonuç.....	93
Öneriler	97
KAYNAKLAR.....	99
EKLER.....	111
EK-1. Veri Üretimi İçin Kullanılan R Kodu	112

EK-2. Birden Çok Input Dosyası Üretmek İçin Kullanılan Mplus Kodu.....	122
---	------------

TABLÖLAR LİSTESİ

Tablo 1. Karma MTK modellerinde 10 madde için üretilen madde parametre değerleri...	25
Tablo 2. Simülasyon Koşulları.....	27
Tablo 3. Karma Rasch Modeli İçin Madde Güçlük Parametre Değerlerinin Ortalama RMSE ve Yanlılık Değerleri	41
Tablo 4. Karma 2PL Modeli İçin Madde Güçlük Parametre Değerlerinin Ortalama RMSE ve Yanlılık Değerleri	42
Tablo 5. Karma 3PL Modeli İçin Madde Güçlük Parametre Değerlerinin Ortalama RMSE ve Yanlılık Değerleri	43
Tablo 6. Madde Güçlük Parametrelerinden Elde edilen $\log[RMSE]$ değerleri için ANOVA Sonuçları	45
Tablo 7. Sınıf Faktörünün Her Bir Koşulunda Kayıp Veri Türlerine Ait Basit Etki ANOVA Sonuçları	46
Tablo 8. Kayıp Veri Yüzdesi Faktörünün Her Bir Koşulunda Kayıp Veri Türlerinin Basit Etki ANOVA Sonuçları	47
Tablo 9. Madde Faktörünün Her Bir Koşulunda Sınıf Faktörünün Basit Etki ANOVA Sonuçları	49
Tablo 10. Madde Faktörünün Her Bir Koşulunda Modellerin Basit Etki ANOVA Sonuçları	50
Tablo 11. Sınıf Faktörünün Her Bir Koşulunda Modellerin Basit Etki ANOVA Sonuçları	51
Tablo 12. Kayıp Veri Yüzdesi Faktörünün Her Bir Koşulunda Modellerin Basit Etki ANOVA Sonuçları	52

Tablo 13. <i>Madde Güçlük Parametrelerinden Elde Edilen Ortalama Yanlılık değerleri için ANOVA Sonuçları</i>	58
Tablo 14. <i>Karma 2PL Model İçin Madde Ayırt Edicilik Parametre Değerlerinin Ortalama RMSE ve Yanlılık Değerleri</i>	63
Tablo 15. <i>Karma 3PL Model İçin Madde Ayırt Edicilik Parametre Değerlerinin Ortalama RMSE ve Yanlılık Değerleri</i>	64
Tablo 16. <i>Madde Ayırt Edicilik Parametrelerinden Elde edilen RMSE değerleri için ANOVA Sonuçları</i>	65
Tablo 17. <i>Madde Faktörünün Her Bir Koşulunda Kayıp Veri Türlerine Ait Basit Etki ANOVA Sonuçları</i>	66
Tablo 18. <i>Sınıf Faktörünün Her Bir Koşulunda Kayıp Veri Türlerine Ait Basit Etki ANOVA Sonuçları</i>	68
Tablo 19. <i>Kayıp Veri Türü Faktörünün Her Bir Koşulunda Modellerin Basit Etki ANOVA Sonuçları</i>	69
Tablo 20. <i>Madde Faktörünün Her Bir Koşulunda Sınıf Faktörünün Basit Etki ANOVA Sonuçları</i>	71
Tablo 21. <i>Madde Faktörünün Her Bir Koşulunda Model Faktörünün Basit Etki ANOVA Sonuçları</i>	72
Tablo 22. <i>Sınıf Faktörünün Her Bir Koşulunda Örneklem Faktörünün Basit Etki ANOVA Sonuçları</i>	74
Tablo 23. <i>Madde Ayırt Edicilik Parametrelerinden Elde Edilen Yanlılık Değerleri İçin ANOVA Sonuçları</i>	78
Tablo 24. <i>Şans Parametre Değerleri İçin Elde Edilen Ortalama RMSE ve Yanlılık Değerleri</i>	80
Tablo 25. <i>Şans Parametrelerinden Elde edilen RMSE değerleri için ANOVA Sonuçları</i>	81
Tablo 26. <i>Şans Parametrelerinden Elde edilen Yanlılık değerleri için ANOVA Sonuçları</i>	83

Tablo 27. Karma Rasch model için elde edilen sınıflama doğruluğu yüzdeleri	84
Tablo 28. Karma 2 parametrelili model için elde edilen sınıflama doğruluğu yüzdeleri	85
Tablo 29. Karma 3 parametrelili model için elde edilen sınıflama doğruluğu yüzdeleri	86
Tablo 30. TIMSS 8.sınıf fen alt testinin 4.ve 5. Kitapçığından elde edilen veriler için Karma MTK model veri uyum indeks değerleri.....	87
Tablo 31. İki Sınıflı Karma Rasch model için tam veriden elde edilen madde parametre değerleri.....	87
Tablo 32. İki Sınıflı Karma Rasch model %10 RKV için elde edilen madde parametre değerleri.....	88
Tablo 33. İki Sınıflı Karma Rasch model %20 RKV için elde edilen madde parametre değerleri.....	89
Tablo 34. İki Sınıflı Karma Rasch model %10 ROKV için elde edilen madde parametre değerleri.....	90
Tablo 35. İki Sınıflı Karma Rasch model %20 ROKV için elde edilen madde parametre değerleri.....	91

ŞEKİLLER LİSTESİ

<i>Şekil 1.</i> Rasch Modelin Örtük Süreklilik Ölçeği.....	9
<i>Şekil 2.</i> Kayıp veri türü ve sınıf faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik.....	47
<i>Şekil 3.</i> Kayıp veri türü ile kayıp veri yüzdesi faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik	48
<i>Şekil 4.</i> Madde ve sınıf faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik	49
<i>Şekil 5.</i> Madde ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik	50
<i>Şekil 6.</i> Sınıf ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik	52
<i>Şekil 7.</i> Model ile kayıp veri yüzdesi faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik	53
<i>Şekil 8.</i> Kayıp veri türü, sınıf ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafikler.....	54
<i>Şekil 9.</i> Madde, sınıf ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafikler	55
<i>Şekil 10.</i> Örnekleme, sınıf ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafikler	56
<i>Şekil 11.</i> Model ile kayıp veri türü faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama yanlılık değerlerine ait grafik	59
<i>Şekil 12.</i> Kayıp veri türü, madde ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama yanlılık değerlerine ait grafik.....	61

<i>Şekil 13.</i> Kayıp veri türü ve madde faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik	67
<i>Şekil 14.</i> Kayıp veri türü ve sınıf faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik	68
<i>Şekil 15.</i> Kayıp veri türü ve model faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik	70
<i>Şekil 16.</i> Madde ve sınıf faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik.....	71
<i>Şekil 17.</i> Madde ve model faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik.....	73
<i>Şekil 18.</i> Örneklem ve sınıf faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik.....	74
<i>Şekil 19.</i> Kayıp veri türü, örneklem ve sınıf faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafikler.....	75
<i>Şekil 20.</i> Örneklem, sınıf ve model faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafikler.....	76
<i>Şekil 21.</i> Madde ve sınıf faktörlerinin etkileşimi sonucu elde edilen şans parametreleri için kestirilen RMSE değerlerine ait grafikler	82

BÖLÜM 1

GİRİŞ

Bu bölümde problem durumu ortaya konulmaya çalışılmış, araştırmanın amacı ve önemi hakkında bilgi verilmiştir.

Problem Durumu

Testler eğitim, psikoloji, endüstri, sağlık gibi farklı bağlamlarda yaygın olarak kullanılmaktadır. Eğitimsel ve psikolojik alanlarda test sonuçları, bireylerin seçimi, gelişimlerinin takip edilmesi veya eğitim sistemlerinin verimliliğini değerlendirmesi gibi çeşitli amaçlar için tercih edilmektedir. Testin önemi ve etkisi konusunda artan farkındalık, daha iyi testlerin geliştirilmesine ve test puanlarının analizi için istatistiksel yöntemlerin geliştirilmesine yol açmıştır. Klasik Test Kuramı (KTK; Novick, 1966) geçen yüzyılın büyük bir kısmında ölçme alanına hizmet etmiştir. Bununla birlikte, KTK ile ortaya çıkan sorunlar (Linden & Hambleton, 1997) test geliştirme ve eşitleme (Skaggs & Lissitz, 1986), bilgisayar uyarlamalı test (Computer Adaptive Testing-CAT; Linden & Glas, 2000) ve değişen madde fonksiyonu (Differential Item Functioning-DIF; Holland & Wainer, 1993) gibi konularda temel bir araç haline gelen Madde Tepki Kuramının (MTK) gelişimine yol açmıştır (Lord, 1980).

MTK'nin KTK'ye göre birçok avantajı bulunmaktadır. KTK'nin örneklem ve madde bağımlılığı özelliklerinin aksine MTK'da bireylerin yetenek parametreleri, madde parametrelerine bağlı değildir ve madde parametreleri de yetenek parametrelerinden bağımsızdır (Hambleton & Swaminathan, 1985; Lord & Novick, 1968). Başka bir deyişle, MTK modellerinde madde ve yetenek parametre tahminlerinin örneklemden örnekleme değişmez olduğu bilinmektedir. Örneklemden bağımsız olarak, aynı MTK modeline sahip bir test için aynı madde parametre değerleri seti elde edilebilmektedir. Örneğin, madde

güçlük parametresi, bir maddenin yüksek veya düşük yetenekli bir gruba uygulanmasına bakılmaksızın aynı kestirim değerine sahip olacaktır. Benzer şekilde, bir kişinin yeteneği, bireylerin örneklemeyle ilgili değildir ve bu nedenle farklı birey kümelerinden elde edilen yetenek tahminleri aynı olacaktır. Ek olarak, MTK'da, madde ve kişi parametreleri aynı örtük sürekliliğe (continuum) yerleştirilir. Bu sayede madde ve birey parametrelerinin aynı ölçek üzerinde karşılaştırılması mümkün olmaktadır. MTK'nin KTK'ye göre avantajları nedeniyle, MTK modelleri birçok psikolojik ölçme aracının geliştirilmesi, analizi ve ölçeklendirilmesinde oldukça yaygın olarak kullanılan modellerdir (DeAyala, 2009; Hambleton & Jones, 1993).

Literatürde, puanlama türüne göre iki veya çok kategorili puanlanan, boyut sayısına göre tek veya çok boyutlu olmak üzere çok sayıda farklı MTK modelleri bulunmaktadır. Bu modeller, madde yanıt fonksiyonu (item response function-IRF) aracılığıyla belirli bir maddeyi yanıtlama olasılığını örtük değişkenle (latent variable) ilişkilendirmenin farklı yollarını önermektedir. (Embretson & Reise, 2013; Hambleton & Swaminathan, 1985; Lord & Novick, 1968, Reckase, 2008).

MTK modelleri birçok avantaja sahip olmasına rağmen tek boyutluluk, yerel bağımsızlık ve parametrelerin değişmezliği gibi katı varsayımlara sahiptir (Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991). Madde Tepki Kuramının avantajları, varsayımlarının karşılanması gerektiren modelin geçerliliğine bağlıdır. Modelin geçerliliğine ilişkin doğru kanıtların toplanabilmesi için analizde kullanılan modelin varsayımlarının sağlanması ve yanlışlık içeren maddelerin bulunmaması gerekmektedir (Kreiner & Christensen, 2007).

Tek boyutluluk varsayımı gözlenen değişkenlerin sadece tek bir örtük değişkenin bir fonksiyonu olduğunu göstermektedir (De Ayala, 2009). Başka bir ifade ile maddelere verilen tepkiler arasındaki ortak varyansı yeterli düzeyde açıklayan tek bir örtük özelliğin bulunması tek boyutluluğun varlığına işaret etmektedir (Embretson & Reise, 2000). Yerel bağımsızlık varsayımı, belirli bir yetenek koşulu altında maddelerin birbiriyle ilişkisiz olmasına dayanmaktadır. Diğer bir ifade ile belirli bir yetenek düzeyinde bir maddeye verilen yanıtların başka bir maddeye verilen yanıtlardan bağımsız olması yerel bağımsızlığın sağlanması anlamına gelmektedir. İlgili alan yazında genellikle tek boyutluluk varsayımı sağlandığı durumlar için yerel bağımsızlık varsayımının da sağlanacağına yönelik görüşler (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985; Lord, 1980) olmasına rağmen, DeMars (2010) tek boyutluluk ve yerel bağımsızlık varsayımlarının ayrı ayrı test

edilmesini önermiştir. Yen (1984) yerel bağımsızlığı test etmede, ayrı ayrı madde çiftleri arasındaki ilişkileri dikkate alan Q_3 istatistiğinin incelenmesini önermiştir. Bu istatistik parametre kestirimi yapıldıktan sonra her bir maddenin artıkları (residual) ile artık matrisler arasındaki korelasyonların incelenmesine dayanmaktadır.

Madde tepki kuramı modellerinin varsayımlarından bir diğeri ise maddeye verilen cevaplar ile örtük değişken arasındaki ilişkinin tüm bireyler için aynı madde tepki fonksiyonu (item response function) ile temsil edilmesidir (Lord & Novick, 1968). Bu varsayım MTK'nin parametre değişmezliği varsayımı olarak adlandırılmaktadır. Bu varsayımın sağlanması madde parametrelerinin tahmini değerlerinin farklı gruplarda değişmeyeceği anlamına gelmektedir (DeMars, 2010; Embretson & Reise, 2000; Hambleton vd. 1991). Örneğin madde güçlük parametresinin değişmezlik özelliği, maddelerin doğru yanıtlanma olasılıklarının farklı gruplardaki tüm bireyler için aynı bilgi, beceri ve stratejileri gerektirdiğini göstermektedir (Andrich, 1988; Linacre, 1996; Perline, Rasch, 1960; Wright & Wainer, 1979). Ancak bazı durumlarda, bireylerin maddeleri doğru yanıtlamak için kullandıkları yanıtlama stratejileri ve teknikleri nedeniyle farklı gruplar oluşabilmekte ve bu gruplar örtük sınıf olarak tanımlanmaktadır (Embretson, 2007; Glück & Spiel, 2007). Diğer bir anlatımla farklı problem çözme teknikleri, madde içeriğine aşinalık veya farklı eğitimsel geçmişe sahip olma durumları gibi birtakım özellikler açısından bireylerin birbirlerinden farklılaşması, farklı örtük sınıfların oluşmasına sebep olabilmektedir (Mislevy & Huang, 2007; Rijkes & Kelderman, 2006). Testlerden elde edilen verilerde birden çok sayıda örtük sınıfın bulunması ise, ölçülen psikolojik yapının farklı gruplar için değişiklik gösterdiği anlamına gelmekte ve bu durum testin geçerliğini tehdit etmektedir (Kreiner & Christensen, 2007; Messick 1994; Toker, 2016; von Davier & Yamamoto, 2007). Çünkü tek boyutluluk, yerel bağımsızlık, parametre değişmezliği ve DIF içeren maddelerin olmaması gibi varsayımların yapı geçerliğinin gereklilikleri olarak ele alınabileceği ifade edilmektedir (Kreiner & Christensen, 2007).

Uluslararası geniş ölçekli uygulamalarda öğrencilerin başarısını kesin ve doğru bir şekilde belirleyebilmek için de MTK modellerinden yararlanılmaktadır (Adams & Wu, 2007; Martin, Mullis & Hooper, 2016; Yamamoto & Kulick, 2000). Geniş ölçekli uluslararası uygulamalarda kullanılan modelin geçerli olması için tek boyutluluk, yerel bağımsızlık, madde parametrelerinin değişmezliği gibi MTK model varsayımlarının sağlanması gerekmektedir. Ancak örtük sınıfların varlığında standart MTK modelleri ile kestirim yapıldığında parametrelerin değişmezliği varsayımı ihlal edilmekte ve bireyler arasındaki

farklılıkların tam olarak belirlenememesi sebebiyle bireylerin yanıt örüntülerini doğru kestirmek mümkün olmamaktadır (Kelderman & Macready, 1990; Mislevy & Verhelst, 1990; Rost, 1990). Bu durumda MTK varsayımlarının sağlanamaması sonucu elde edilen parametre kestirimleri yanlış olabilmekte ve yanlış sonuçlar elde edilebilmektedir (DeMars & Lau, 2011). Ayrıca varsayımlarının sağlanamaması aynı yetenek düzeyine sahip bireylerin karşılaştırılmasını sınırlamaktadır (Oliveri & von Davier, 2011; Rutkowski & Rutkowski, 2018). Başka bir ifade ile varsayımların ihlali durumunda bireyleri ortak bir yetenek düzeyinde karşılaştırmak adil bir yaklaşım olarak görülmemektedir (Baghaei & Carstensen, 2013; Embretson, 2007). Sonuç olarak madde parametrelerinin değişmezlik özelliği, bireylerin testten aldıkları puanları karşılaştırmak için gerekli bir koşul olarak belirtilmektedir (Oliveri & von Davier, 2011).

Standart MTK modellerinin varsayımlarının sağlanamaması veya ihlal edilmesi durumlarında çok boyutlu MTK modelleri (Reckase 2009), çoklu grup MTK modelleri (Bock & Zimowski 1997) ve Karma MTK (Mixture IRT) modelleri (Rost 1990; Mislevy & Verhelst 1990) olmak üzere çeşitli modeller geliştirilmiştir. Karma MTK modelleri, MTK modellerinin aksine madde parametrelerinin gruplar arasında değişmez olduğu varsayımına ihtiyaç duymamakta ve örtük sınıflar arasında madde parametrelerinin değişimine izin vermektedir (De Ayala & Santiago, 2017; von Davier & Yamamoto, 2007). Dolayısıyla gruplar arasında madde parametrelerinin değişimine imkân veren Karma MTK modellerinin kullanılmasının daha geçerli ve güvenilir sonuçlar elde etmek adına gerekli olduğu düşünülmektedir. İlgili alan yazın incelendiğinde, genellikle Karma Rasch ve Karma iki parametrelili (2PL) modelin odakta olduğu çalışmaların çoğunlukta, Karma üç parametrelili (3PL) model ile ilgili yapılan çalışmaların ise azınlıkta olduğu tespit edilmiştir (Finch & French, 2012; Lee, 2012; Oliveri, Ercikan, Zumbo & Lawless, 2014; Oliveri & von Davier, 2011; Preinerstorfer & Formann, 2012; Toker, 2016; Zhang, Orrill & Campbell, 2015).

Modelin varsayımlarının incelenmesinin yanında testlerle ölçülmeye çalışılan örtük değişkenlere ulaşılabilmesi için elde edilen veri setinin kayıp veriye sahip olup olmaması bakımından da incelenmesi gerektiği önerilmektedir (Little & Rubin, 1987). Kayıp veri oluşmasına neden olan durumlar, hız testi uygulamalarında, bireylerin tüm maddeleri cevaplayabilmek için yeterli zaman kalmaması durumunda test formunun sonlarında bulunan maddelerin cevaplanamaması veya bireylerin bazı maddeleri kasıtlı bir biçimde boş bırakmalarına bağlı olarak ortaya çıkabilir (De Ayala, Plake & Impara, 2001). Cevap örüntülerinde yer alan kayıp veriler, parametre kestirimlerinde yanlışlığın ve standart

hataların artması ve testin gücünün düşmesi gibi olumsuz durumlara neden olmaktadır (De Ayala, Plake & Impara, 2001; Finch, 2008). Bu noktada verilerin analizine başlamadan önce kayıp veri oranı ve kayıp veri türüne ait mekanizmanın belirlenmesi oldukça önemli bir yer oluşturmaktadır. Bu sebeple, kayıp veri türü ve kayıp veri oranı tespit edilerek kullanılan analizlerle ilgili daha geçerli ve güvenilir sonuçlar elde edebilecektir.

Farklı kitapçıkların kullanıldığı uygulamalarda bireylerin almadıkları kitapçıklarda yer alan maddeler için cevap üretememesi de kayıp veri oluşmasının nedenlerinden biri olarak görülmektedir (De Ayala, Plake & Impara, 2001). Farklı kitapçıkların kullanıldığı uluslararası geniş ölçekli bir uygulama olan TIMSS veri setinde de kayıp veri oldukça önemli bir problem olarak karşımıza çıkmaktadır. Çalışmada kayıp veri türü ve oranının incelenmesi amaçlandığından dolayı gerçek veri uygulamasına dair incelemelerde TIMSS veri setinin kullanılmasına karar verilmiştir. Simülasyon çalışmalarında bütünden parçaya doğru bir yaklaşımla genel bir model belirlenip farklı veri setlerine uyarlanır ve bu veri setleri üzerinde farklı önermeler ve ilişkiler incelenir. Böylece karmaşık yapıyı doğru bir şekilde yansıtabilecek veri setleri ulaşılabilir hale gelir (Gilbert, 2004). Bu nedenle incelenecek olan modellerin farklı koşullar altındaki değişimini görmek için ve gerçek veride ulaşılamayacak durumlar için simülasyon veri setine başvurulmuştur.

Yapılan açıklamalar doğrultusunda farklı örneklem büyüklüğü, madde sayısı, örtük sınıf sayısı, kayıp veri türü ve kayıp veri yüzdesi gibi çeşitli değişkenler ele alınarak farklı simülasyon koşullarında Karma 3PL modellerin incelenmesine gerek olduğu görülmüştür. Bu sebeple Karma MTK modellerinden elde edilen parametre kestirimlerinin RMSE ve yanlılık değerlerinin incelenmesi, Karma MTK modellerinde kayıp veri, madde sayısı, sınıf sayısı, örneklem büyüklüğü, kayıp veri yüzdesi faktörlerinin minimum hataya sahip kestirimler elde etmek için hangi koşullarda alınmasının en uygun olacağı yönünde bilgiler elde edilmesine imkân sağlanmıştır.

Araştırmanın Amacı

Araştırmanın amacı, Karma MTK modellerinin (Karma Rasch, Karma 2PL, Karma 3PL) farklı örneklem büyüklüğü, madde sayısı, sınıf sayısı, kayıp veri türü ve kayıp veri yüzdesi koşullarında madde parametre kestirimlerine ilişkin hataların ortalama karekökünü (RMSE) ve yanlılık (BIAS) değerlerini belirlemek ve sınıflama doğruluğu hakkında bilgi elde edebilmektir. Elde edilen sonuçlara ilişkin olarak Karma MTK modelleri kullanılarak minimum RMSE ve yanlılık değerlerine sahip kestirimler yapabilmek için gerekli örneklem

büyükülüğü, madde sayısı, sınıf sayısı, kayıp veri türü ve kayıp veri oranı ile ilişkili önerilerin verilmesi amaçlanmıştır. Ele alınan faktörlerin etkileşim etkilerinin incelenmesi ve özellikle Karma 3PL model ile kayıp veri yüzdesi ve kayıp veri türü faktörleri ile ilgili literatüre katkı sağlayacak sonuçların önerilmesi hedeflenmiştir. Ayrıca TIMSS 2015 8.sınıf Türkiye verisinden simülasyon koşullarına benzer durumlar oluşturularak simülasyon verisi gerçek veri ile desteklenmiştir.

Bu amaçlar doğrultusunda, simülasyon ve gerçek verinin kullanıldığı çalışmada aşağıdaki sorulara yanıt aranmıştır:

- 1) Karma MTK modelleri ile yapılan parametre kestirimleri sonucu elde edilen ortalama RMSE ve yanlılık değerleri, örneklem büyüklüğü, madde sayısı, model, sınıf sayısı, kayıp veri oranı ve kayıp veri türü değişkenleri ele alındığında nasıl değişmektedir?
- 2) Karma MTK modelleri ile yapılan parametre kestirimleri sonucu elde edilen ortalama RMSE ve yanlılık değerlerine göre ele alınan değişkenlerin etkileşim etkisi nasıl değişmektedir?
- 3) Ele alınan değişkenlerin kombinasyonundan elde edilen sınıflama doğruluğu yüzdeleri nasıl değişmektedir?
- 4) Gerçek veri kısmında TIMSS 2015 Türkiye 8. Sınıf fen alt testinden rastgele çekilen tam, rastgele kayıp veri (RKV) ve rastgele olmayan kayıp veri (ROKV) koşullarında kestirilen madde parametre değerleri nasıl değişmektedir?

Araştırmanın Önemi

Karma MTK modellerinden elde edilen parametre kestirimlerinin RMSE ve yanlılık değerlerinin incelenmesi, Karma MTK modellerinin farklı koşullardaki performansının değerlendirilerek belirlenen koşullarda hangi modelin daha az hataya sahip olduğunun tespit edilmesi yönüyle önemlidir. Bu yönde elde edilen bulguların, uygulayıcıların yapacakları çalışmalarda kendi koşullarına göre hangi modeli seçmelerinin uygun olacağı yönünde bilgi sunması ve yol göstermesi bakımından önemli görülmektedir.

Literatürde Karma Rasch ve Karma 2PL model ile madde sayısı, sınıf sayısı, örneklem büyüklüğü, kayıp veri yüzdesi ve kayıp veri türleri gibi faktörlerin birlikte ele alındığı simülasyon çalışmaları olmasına rağmen yapılan alan yazın taraması sonucunda Karma 3PL model ile kayıp veri türleri ve kayıp veri yüzdesi faktörlerinin ele alındığı bir çalışmaya

rastlanmamıştır. Bu çalışmadan elde edilecek bulgular bağlamında Karma 3PL modeli ile kayıp veri türü ve kayıp veri yüzdesi faktörlerinin birlikte ele alınması sonucu elde edilen hata değerleri ve sınıflama doğruluğu ile ilgili geniş ve ayrıntılı yorumlar yapılması yönüyle araştırmanın önemli olduğu düşünülmektedir.

Araştırmanın bir diğer önemi ise farklı koşullarda yapılan incelemenin hem simülasyon hem de gerçek veri bağlamında ele alınmasıdır. Yapılan simülasyon çalışması TIMSS 2015 Türkiye 8.sınıf fen alt testinden elde edilen veriler ile desteklenerek Karma MTK modellerinin hem gerçek hem de simülasyon veri ile ele alındığı bir çalışma olarak literatüre katkı sağlaması yönüyle önemli görülmektedir. Bu sayede yapılan simülasyon çalışmasından elde edilen bulguların gerçek veriden elde edilen bulgularla karşılaştırılarak değerlendirilmesi yapılan simülasyon çalışmasının geçerliğine katkı sunmuştur. Ayrıca kayıp veri oranı ve türü ile ilgili yapılan incelemelerin uluslararası geniş ölçekli bir uygulama olan TIMSS veri seti üzerinden de incelenmesi ulusal ve uluslararası düzeyde kayıp veriye sahip veri setlerinden yapılacak analizlerde politika yapıcılara fikir sunması yönüyle de önemlidir.

Araştırmanın Sınırlılıkları

1. Araştırmada dikkate alınan simülasyon koşulları örneklem büyüklüğü (600, 1000), madde sayısı (10, 30), örtük sınıf sayısı (2, 3), Karma MTK modelleri (Rasch, 2PL, 3PL), kayıp veri türü (tam, RKV, ROKV) ve kayıp veri yüzdesi (%10, %20) ile sınırlıdır.
2. Veri üretimi her bir koşul için 100 replikasyon ile sınırlıdır.
3. Gerçek veride Türkiye verisi kullanıldığı için iki kitapçık birleştirildiğinde en fazla 880 kişiye ulaşılması sebebiyle Türkiye verisinde 1000 kişilik koşul için analizler gerçekleştirilememiştir. İki kitapçıkta bulunan iki kategorili puanlanan ortak madde sayısının 14 olması sebebiyle madde sayısı sabit koşul olarak alınmıştır.

BÖLÜM 2

KAVRAMSAL ÇERÇEVE

Bu bölümde MTK modelleri, Örtük Sınıf Analizi ve Karma MTK modelleri ile ilgili genel bilgiler verilerek ayrıca alan yazında yapılmış olan ilgili araştırmalar sunulmuştur.

Madde Tepki Kuramı Modelleri

Madde Tepki Kuramında kullanılan modeller boyutluluk özelliğine göre, tek bir örtük özelliği ölçmek için kullanılan tek boyutlu MTK modelleri ve birden fazla örtük özelliği ölçmek için kullanılan çok boyutlu MTK modelleri olmak üzere ikiye ayrılmaktadır. MTK modelleri veri yapısına göre ise iki kategorili puanlanan ya da çok kategorili puanlanan modeller olarak adlandırılmaktadır (Embretson & Reise, 2000; Hambleton & Swaminathan, 1989). İkili puanlanan MTK modelleri için tek boyutlu modeller 1PL, 2PL ve 3PL formlarına göre normal ogive ve lojistik modelleri içermektedir. Psikoloji ve eğitim alanındaki ölçmelerde en çok kullanılan modeller 1PL, 2PL ve 3PL lojistik modellerdir. 3PL MTK modeli için bir maddeyi doğru yanıtlama olasılığı şu şekilde tanımlanmaktadır:

$$P(x_{ij} = 1|\theta_j) = P_{ij} = Y_i + (1 - Y_i) \frac{\exp[\alpha_i(\theta_j - \beta_i)]}{1 + \exp[\alpha_i(\theta_j - \beta_i)]} \quad (2)$$

Üç parametrelili lojistik modele ait bu eşitlikte madde güçlük, madde ayırt edicilik ve şans parametreleri bulunmaktadır. Madde güçlüğü (β_i) bireyin bir maddeyi doğru cevaplama olasılığı hakkında bilgi sağlarken, ayırt edicilik (α_i) maddenin ölçülen yapının farklı seviyelerine sahip bireyleri ne kadar iyi ayırt ettiğini göstermektedir. Şans parametresi en düşük yetenek düzeyindeki bireyin maddeyi doğru yanıtlama olasılığı olarak tanımlanmaktadır (De Ayala, 2009). Bu üç parametre kullanılarak bireylerin yetenekleri kestirilmektedir (Hambleton vd.,1991). Üç parametrelili lojistik modelden şans

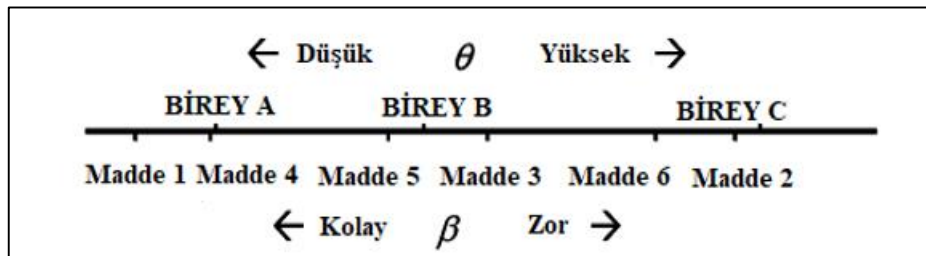
parametresinin (Y_i) çıkarılması ile iki parametrelili lojistik model elde edilmektedir (Birnbaum,1968). 2PL MTK modeli için bir maddeyi doğru yanıtlama olasılığı şu şekilde tanımlanmaktadır:

$$P(x_{ij} = 1|\theta_j) = P_{ij} = \frac{\exp[\alpha_i(\theta_j - \beta_i)]}{1 + \exp[\alpha_i(\theta_j - \beta_i)]} \quad (3)$$

Bir parametrelili model ise iki ve üç parametrelili lojistik modellerin özel bir halidir (Hambleton & Swaminathan, 1985). Bir parametrelili lojistik modelde tüm maddeler eşit derecede ayırt edicilik gücüne (maddelerin tümü için kestirilen ortalama bir α parametresi değerine) sahiptir ve şans parametresi sıfır olarak kabul edilmektedir. Bu modelde sadece madde güçlük parametresi dikkate alınır ve bu parametreye dayalı olarak yetenek kestirimi yapılmaktadır. Ayırt edicilik parametresinin tüm maddeler için eşit ve 1 olarak kabul edildiği model ise bir parametrelili lojistik modelin özel bir hali olan Rasch modelidir (Embretson & Reise, 2000; Hambleton vd.,1991). George Rasch tarafından 1960 yılında, iki kategorili puanlanan maddeler için geliştirilen Rasch modelinde bireyin bir maddeye doğru yanıt verme olasılığı, bireyin yetenek (θ_j) düzeyi ile madde güçlüğü (β_i) arasındaki farkın lojistik fonksiyonuna dayanmaktadır. Bu modelde madde güçlük parametresinin (β_i) yanında bireyin yetenek (θ_j) düzeyi de kestirilmektedir (Rost, 1990). Rasch modelin matematiksel gösterimi aşağıdaki eşitlik ile ifade edilmektedir:

$$P(x_{ij} = 1|\theta_j) = P_{ij} = \left(\frac{\exp[(\theta_j - \beta_i)]}{1 + \exp[(\theta_j - \beta_i)]} \right) \quad (4)$$

Eşitlikte; j bireyi, i maddeyi, β_i , i maddesinin güçlüğü temsil etmektedir. Bu olasılık, madde güçlüğü ve birey yeteneğinin bir fonksiyonudur. Birey yeteneği, madde güçlüğüne eşit olduğunda, bireyin maddeyi doğru cevaplama olasılığı 0,5'dir. Her bir yetenek için, madde güçlük değeri azaldığında veya yetenek arttığında performans artmaktadır. Yani, maddeler ve bireyler monoton olarak sıralanmıştır ve madde güçlüğü ile bireyin yeteneği aynı ölçek üzerinde yer almaktadır. Şekil 1'de Rasch modelin örtük sürekliliğinin (latent continuum) grafiksel gösterimi yer almaktadır.



Şekil 1. Rasch Modelin Örtük Süreklilik Ölçeği

Şekil 1'deki ölçekte sağ tarafa gittikçe yani $+\infty$ 'a yaklaştıkça bireylerin yeteneği artmaktadır. Yani C bireyi en yüksek yeteneğe sahip birey olması sebebiyle maddeleri doğru yanıtlama olasılığı en yüksek olan bireydir. A bireyi ise maddeleri doğru yanıtlama olasılığı en düşük olan bireydir. Ölçekte yer alan bireyler için en kolay olan madde, madde 1; en zor olan madde ise madde 2 dir. Bu ölçek $(-\infty, +\infty)$ aralığında yer alan madde güçlük ve yetenek parametrelerinin yorumlanmasını kolaylaştırmaktadır.

Rasch Model oldukça basit olmasına rağmen tek boyutluluk ve yerel bağımsızlık gibi varsayımlara sahiptir. Tek boyutluluk, testi oluşturan maddelerin doğru cevaplanma olasılığının sadece bir yetenek türüne bağlı olduğunu ifade etmektedir (Crocker & Algina, 1986; Embretson & Reise, 2000). Yerel bağımsızlık varsayımı ise herhangi bir bireyin bir maddeye verdiği tepkinin, yalnızca bireyin yetenek düzeyine ve madde cevap fonksiyonunu tanımlayan madde parametrelerine bağlı olduğunu bir kanıttır (Reckase, 2009). Ancak bu varsayımlar pratikte ihmal edilebilmektedir. Ayrıca araştırmaya konu olan evren farklı örtük sınıflardan meydana gelebilir ve bu örtük sınıflarda bireylerin maddelere yanıt verme eğilimleri farklılaşabilir (Rost, 1990; von Davier & Rost, 2007). Bu örtük sınıfları belirleyebilmek için Örtük Sınıf Analizi yaklaşımı kullanılmaktadır.

Örtük Sınıf Analizi (ÖSA)

Örtük Sınıf Analizi (ÖSA) ilk kez Lazarsfeld tarafından kategorik olarak gözlenen değişkenlere dayalı olarak kümeleme için kullanılarak tanıtılmıştır (Lazarsfeld,1950). Goodman (1974) modelin pratikte uygulanabilmesi amacıyla model parametrelerinin maksimum olasılık tahminlerini elde etmek için bir algoritma geliştirmiştir. Daha sonra çoklu gözlenen değişkenler (polytomous manifest variables) ve çoklu örtük değişkenler (multiple latent variables) üzerinde çalışmıştır. Ayrıca, model belirleme konusunda önemli bir çalışmayı tamamlamıştır. Frekans tabloları için örtük sınıf modelleri ve loglinear modelleri arasındaki ilişkiler ile sürekli değişkenler, sıralı değişkenler, örtük değişken ve tekrarlı ölçümler içeren modellerin geliştirilmesi gibi diğer bazı önemli çalışmalar da gerçekleştirilmiştir (Haberman, 1979; Hagenaars, 1990)

Collins ve Lanza'ya (2010) göre, ÖSA bireylerin örtük bir yapıya dayanarak alt gruplara veya örtük sınıflara ayrılıp ayrılmadığını belirlemek için sosyal, davranışsal ve sağlık bilimlerinde kullanılan bir modeldir. ÖSA ve doğrulayıcı faktör analizi benzer fikirlere sahiptir. Ancak doğrulayıcı faktör analizinde değişken (yani faktör) sürekli ve sürekli olarak ele alınan göstergelerle normal bir dağılıma sahiptir. ÖSA'da örtük değişken (yani

örtük sınıf değişkeni) kategoriktir ve kategorik olarak ele alınan göstergelerle çok terimli (multinomial) dağılıma sahiptir (Collins & Lanza, 2010). Örtük sınıf analizi kavramsal açıdan MTK modellerinin genelleştirilmesi olarak bir parametrelili MTK'na benzerdir (Samuelsen & Dayton, 2010). Ancak, örtük değişken ÖSA'da kategorik iken, MTK'da sürekli (Collins & Lanza, 2010).

Örtük sınıf analizinin amacı kategorik olarak gözlenen değişkenler kullanılarak sınıf üyeliğini belirlemektir. Örtük değişkenler doğrudan gözlenen değişkenler değildir, ancak doğrudan ölçülen gözlenen değişkenler tarafından belirtilmiştir. Örtük bir sınıf, homojen yanıt örüntüleri olan öğrencilerin istatistiksel olarak belirlenmiş bir grubunu gösterir. Başka bir deyişle, farklı örtük sınıflar maddelere verilen cevap örüntülerini yansıtır. Sınıflar, doğrudan gözlenebilir olmadıkları için örtük olarak etiketlenmektedir (Bolt, Cohen, & Wollack, 2001). Aynı örtük sınıftaki bireylerin aynı yeteneğe sahip bireyler olduğu söylenebilir.

$Y_{ij} \in \{0,1\}$, $j \in \{1,2,\dots,N\}$ bireylerinin $i \in \{1,2,\dots,T\}$ maddelerine verdiği cevabı gösteren rastgele bir değişken, $g \in \{1,2,\dots,G\}$ ise bireyin örtük sınıf üyeliğini gösterebilir. Yerel bağımsızlık varsayımı altında doğru cevaplama marjinal olasılığı $P(Y_{ij} = 1)$ aşağıdaki gibi yazılabilir:

$$P(Y_{ij} = 1) = \sum_{g=1}^G \pi_g P(Y_{ij} = 1 | G = g) \quad (5)$$

Burada π_g örtük sınıf olasılığını, $P(Y_{ij} = 1 | G = g)$ koşullu olasılığı ise g sınıfındaki j bireyinin i maddesini doğru yanıtlama olasılığını göstermektedir.

Denklem 5'te gösterildiği gibi, Y_{ij} cevabının elde edilme olasılığı, sınıfa özgü (class-specific) olasılıkların ağırlıklı ortalamasıdır. Örtük sınıf olasılığı π_g ve koşullu olasılık $P(Y_{ij}=1|G=g)$ örtük sınıf modellerinin iki ana bileşenidir. Örtük sınıf modelinin; örtük sınıf olasılıkları ve koşullu olasılıklar olmak üzere tahmin edilmesi gereken iki tür parametresi vardır. Örtük sınıf olasılığı, örtük sınıf dağılımlarını, sınıf sayısını ve bu sınıfların göreceli büyüklüklerini tanımlar. Bundan dolayı koşullu olasılık $P(Y_{ij}=1 | G = g)$, g sınıfında j maddesine doğru yanıt veren bireylerin olasılığını belirtmektedir. Koşullu olasılığın bu şekilde kullanımı, faktör analizinde faktör yüküne benzer olarak düşünülebilir (McCutcheon, 1987). Standart faktör analizinde, bilindiği gibi faktör yükleri bir değişkenin örtük faktörle olan ilişkisini göstermektedir. Örtük sınıf modelinde yer alan koşullu olasılıklar ise örtük değişkenin g . sınıfı ile gösterge değişkenin i . kategorisi arasındaki

ilişkinin gücünün bir ölçüsüdür. Her bir örtük sınıfta, koşullu olasılıklar g sınıfında yer alan gözlemlerin her bir gösterge değişken ile benzer özelliklere sahip olup olmadığını göstermektedir. Örtük değişkenin her bir g sınıfındaki gösterge değişkenlerin her biri için koşullu olasılıklar seti bulunmaktadır. Örneğin; örtük sınıfları tanımlamada üç gösterge değişken kullanıldıysa her bir sınıf için üç koşullu olasılık seti elde edilecektir. Sonuç olarak, her bir örtük sınıfta, gözlemler gösterge değişkenin belirli bir düzeyinde bulunmasına ait belirli bir olasılık değerine sahiptir. Her bir örtük sınıfta, her bir gösterge değişkene ilişkin koşullu olasılıkların toplamı 1 olduğundan, gösterge değişkenin kategori sayısının bir eksiği kadar koşullu olasılık hesaplamak yeterli olacaktır (Collins & Lanza, 2010; Magidson & Vermunt, 2004). Bu durumda hem örtük sınıf olasılığı hem de koşullu olasılık, olasılık özelliklerini karşılamalıdır:

- 1) Her ikisi de negatif olmamalıdır.
- 2) Tüm örtük sınıflar üzerindeki örtük sınıf olasılıklarının toplamı bire eşit olmalıdır.
- 3) Her sınıfta, her madde toplamı için tüm kategorilerdeki koşullu olasılıkların toplamı bire eşit olmalıdır (Dayton & Macready, 2007). Örneğin iki kategorili puanlanan maddeler için;

$$P(Y_{ij} = 1 \mid G = g) + P(Y_{ij} = 0 \mid G = g) = 1.$$

Örtük Sınıf analizi modelini tahmin etmek için dört adım tanımlanmaktadır (J. Wang & Wang, 2012):

1. Optimal sınıf sayısını bulmak,
2. Örtük sınıf üyeliğinin sınıflandırma kalitesini değerlendirmek,
3. Örtük sınıfları tanımlamak
4. Örtük sınıf üyeliğini tahmin etmek

Sınıfların optimal sayısını belirlemek için, k-sınıfı modeli, k-1 sınıfı modellerle iteratif olarak karşılaştırılır. Örtük sınıf modellerinde model uyumu karşılaştırmaları için AIC (Akaike, 1974), tutarlı AIC (CAIC; Bozdoğan, 1987), BIC (Schwarz, 1978) gibi kullanılan farklı uyum indeksleri vardır. Bu indekslerin daha küçük değerleri daha iyi model uyumu göstermektedir. Bu indekslerin performansları, tahmini parametrelerin sayısı, bireylerin sayısı, tahmin teknikleri, model tipi ve karmaşıklık gibi çeşitli faktörlere bağlıdır.

Bir bireyin cevap örüntülerine dayanarak, örtük sınıf üyeliği olasılığı, sonsal sınıf üyeliği olasılığı ile ölçülmektedir. (J. Wang & Wang, 2012). Bayes kuralı kullanılarak birey üyeliğinin en yüksek sonsal olasılığı olan muhtemel sınıfa yerleştirilir:

$$P(G = g|Y = y) = \frac{P(G=g)P(Y = y|G = g)}{P(Y=y)} \quad (6)$$

Bire yakın bir olasılık, o kişinin yanlış sınıflandırma şansının çok düşük olduğunu göstermektedir. Örneğin bir bireyin dört sınıflı bir örtük sınıf modelinde Sınıf 1, Sınıf 2, Sınıf 3 ve Sınıf 4 için kestirilen sonsal sınıf üyeliği olasılıkları sırasıyla 0,06; 0,10; 0,12 ve 0,72 olsun. Bu durumda birey dördüncü sınıfa atanacaktır. Bu bireyin doğru sınıflama olasılığı 0,72; yanlış cevaplama olasılığı ise 0,28 olacaktır. Pratikte bire yakın bir sonsal olasılık puanı elde etmek oldukça zordur. Nagin (2005) 0,70 ve üzeri olasılık değerinin bir sınıfa atanmada uygun değer olduğunu belirtmektedir.

Örtük sınıfların tanımlaması adımında ise her bir örtük sınıfın anlamlı ve yorumlanabilir olması oldukça önemlidir. Örtük sınıf analizinin esas amacı veri setindeki heterojenliği açıklamaktır. Heterojen bir evreni oluşturan farklı örtük gruplar, cevap örüntülerine dayalı olarak istatistiksel analizlerle belirlenmektedir. Ayrıca, model bir matematiksel analizin tüm gereksinimlerini karşılarsa bile, teorik olarak yorumlanabilir bir örtük sınıf sağlayamıyorsa, tahmini model yararlı olmayacaktır (J. Wang & Wang, 2012). Bu noktada, sonuçları yorumlamak için araştırmacının amacına göre cinsiyet ve ülke gibi demografik veriler kullanılabilir. Örtük sınıf analizinin son adımı, sınıf üyeliği tahminidir. Bu amaçla, ortak değişkenleri analize kolayca dahil etmek mümkündür (Muthén, 2004). Öte yandan örtük sınıf analizi modellemesinde bilinen bir problem, global maksimumlara dayalı yanlış parametre tahminleri sağlamasıdır (J. Wang & Wang, 2012). Farklı rastgele başlangıç değerlerini alarak modeli tahmin etmek bu sorun için pratik bir çözüm sağlayabilir.

Karma MTK Modelleri

Madde tepki kuramı ile örtük sınıf analizinin birlikte kullanılması “Karma Madde Tepki Kuramını” (Karma MTK) ortaya çıkarmaktadır (Mislevy & Verhelst, 1990; Rost, 1990). Karma MTK modelleri evrenin sınırlı sayıda örtük sınıftan oluştuğunu varsayar ve bu sınıflar madde cevap örüntülerine göre ayırt edilebilir (von Davier & Rost, 2007). Karma MTK modellerinin amacı homojen alt grupları tespit etmektir (Mislevy & Verhelst, 1990). Karma MTK modelleri, madde parametrelerinin gruplar arasında değişmez olması varsayımına ihtiyaç duymamakta ve bu durum madde parametrelerinin örtük gruplar

arasında değişmesine izin vermektedir. Bu model sadece örtük sınıflar arasında heterojenliği değil, aynı zamanda bireylerin örtük sınıf içindeki yetenekleri arasında da değişkenlik sağlar. Sınıf üyeliği, örtük sınıflar boyunca cevap örüntülerindeki niteliksel farklılıkları yansıtırken, yetenek düzeyindeki niceliksel farklılıkları da yansıtmaktadır. Böylece Karma MTK modelleri, araştırmacıların standart MTK modeliyle tespit edilemeyen, her bir örtük sınıftaki yanıt süreçleri hakkında bilgi sağlamaya yardımcı olmaktadır. Örtük sınıflar, örneğin öğretim geçmişi (Muthén, 2004), çözüm stratejileri (Mislevy & Verhelst, 1990; Rost, 1990) veya tepki örüntüleri (Cho, Cohen, Kim & Bottge, 2010) gibi heterojen nitel özelliklerden oluşabilir. Karma MTK modellerinin, bireyleri yeteneğine göre örtük gruplara ayırması, madde ve grup karakteristikleri hakkında daha geçerli ve güvenilir sonuçlar elde etmeyi sağlamaktadır. Karma MTK modelleri, değişmezlik varsayımının ihlal edildiği durumlarda, maddelerle örtük yapı arasındaki ilişki için yararlı bir açıklama sağlayabilir (von Davier, Rost & Carstensen, 2007). Ayrıca Karma MTK yaklaşımı sürekli ve kategorik verinin eş zamanlı modellenmesine imkân tanımakta ve daha fazla bilgi edilmesini sağlamaktadır (De Ayala & Santiago, 2017).

Örtük sınıf analizi ile MTK modelleri yaklaşımlarını birleştiren modeller ise Karma MTK modelleridir. Karma MTK modelleri, ÖSA ve MTK modellerinin sınırlılıklarının üstesinden gelen ve örtük sınıfların varlığında MTK parametrelerinin kestirimine olanak sağlayan modellerdir. Sırasıyla Karma Rasch, Karma 2PL ve Karma 3PL modelleri tanıtılacaktır:

Karma Rasch model

Karma Rasch Modeli, ilk kez iki kategorili puanlanan maddeler için Rost tarafından tanımlanmıştır (Rost, 1990). Karma Rasch model madde cevaplarına dayanarak çoklu örtük sınıfları belirleyen Rasch modelin teorik gücü ile ve örtük sınıf analizinin deneysel gücünü birleştiren bir modeldir (Kelderman & Macready, 1990; Mislevy & Verhelst, 1990; Rost, 1990; von Davier & Rost, 2006). Ayrıca Karma Rasch modelinin, Rasch modelinin ve ÖSA'nin varsayımlarını esnettiği söylenebilir. Çünkü ÖSA'de her bir örtük sınıf için madde güçlük parametrelerinin aynı olması koşulu varken Karma Rasch modelde ise farklı örtük sınıflarda madde güçlüklerinin farklılaşmasına olanak sağlamaktadır. Başka bir ifade ile Rasch model ve ÖSA'nin eksikliklerinin üstesinden gelmek ve pozitif özelliklerini korumak için Karma Rasch modeli kullanılmaktadır.

Rasch modelinde, sürekli örtük değişken madde performansının temelini oluştururken örtük sınıf modelinde örtük sınıf üyeliği, madde performansının temelini oluşturmaktadır. Rasch

model ve örtük sınıf modelinin kombinasyonu olan Karma Rasch model ise, örtük yeteneğin ve aynı zamanda bireylerin maddeyi doğru cevaplama olasılığındaki farklılıklarına sebep olan örtük grup üyeliğinin eş zamanlı olarak tahmin edilmesini sağlamaktadır (Baghaei & Carstensen). Bu modelleme çerçevesinde, bireyler hem bir konum parametresi (θ_{jg}) hem de bir örtük sınıf üyelik parametresi (π_g) ile karakterize edilir. Karma Rasch model altında, g örtük sınıfında j bireyinin bir i maddesini doğru cevaplama olasılığı aşağıdaki eşitlik ile hesaplanabilir (De Ayala & Santiago, 2017):

$$P(x_{ij} = 1|\theta_j) = P_{ij} = \sum_{g=1}^G \pi_g \left(\frac{\exp[(\theta_{jg} - \beta_{ig})]}{1 + \exp[(\theta_{jg} - \beta_{ig})]} \right) \quad (7)$$

Bu eşitlikte $g = (1, 2, \dots, G)$ örtük sınıf üyeliğini, (β_{ig}) i maddesi için sınıf içi (intraclass) madde güçlük parametresini, (θ_{jg}) g sınıfındaki j bireyi için yetenek parametresini, π_g ise bir sınıftaki bireylerin oranını (mixing proportion) temsil etmektedir. Her bir bireyin örtük sınıflardan birine ait olma olasılığı ve her bir sınıftaki bireylerin oranı (π_g), $\sum_{i=1}^g \pi_g = 1$ ve $0 \leq \pi_g \leq 1$ sınırlaması ile kestirilmektedir (Rost, 1990).

Denklemdaki birey ve madde konum parametreleri üzerindeki alt indis g , her bir örtük sınıf için bu parametre kümesinin var olduğunu yansıtır. MTK'da olduğu gibi, bu parametreler her sınıfta sürekli. Örtük sınıf oranlarının toplamı bire eşittir. Sadece bir örtük sınıf olduğunda, Karma Rasch modeli tek boyutlu Rasch modeline eşdeğerdir. Karma Rasch modelinde madde ve yetenek parametreleri her bir örtük sınıf için ayrı ayrı tahmin edilir, bu nedenle tahmin edilen parametreler örtük sınıf g 'ye koşulludur (De Ayala & Santiago, 2017).

Karma 2PL ve 3PL MTK Modelleri

Her bir sınıf için madde parametrelerinin ve şans parametresinin yer aldığı üç parametrelilik Karma MTK modeli ise şu şekilde gösterilmektedir (Choi, Alexeev & Cohen, 2015):

$$P(x_{ij} = 1|Q_j) = P_{ij} = \sum_{g=1}^G \pi_g \left(Y_{ig} + (1 - Y_{ig}) \frac{\exp[\alpha_{ig}(\theta_{jg} - \beta_{ig})]}{1 + \exp[\alpha_{ig}(\theta_{jg} - \beta_{ig})]} \right) \quad (8)$$

Bu eşitlikte üç parametrelilik Karma MTK modeli için $g = (1, 2, \dots, G)$ örtük sınıf üyeliğini, (β_{ig}) i maddesi için sınıf içi (intraclass) madde güçlüğü parametresini, (α_{ig}) i maddesi için sınıf içi madde ayırt ediciliği parametresini, (Y_{ig}) i maddesi için düşük asimptotu (lower-asymptote) yani şans parametresini, (θ_{jg}) g sınıfındaki j bireyi için yetenek parametresini, π_g ise bir sınıftaki bireylerin oranını (mixing proportion) temsil etmektedir.

Her bir bireyin örtük sınıflardan birine ait olma olasılığı ve her bir sınıftaki bireylerin oranı (π_g) , $\sum_{g=1}^G \pi_g = 1$ ve $0 \leq \pi_g \leq 1$ sınırlaması ile kestirilmektedir (Rost, 1990). Ayrıca standart MTK kapsamında ele alındığı gibi; düşük-asimptot parametresinin sıfıra eşit olması yani şansın ortadan kalkması ile Karma 2PL model formuna dönüşmektedir İki parametrelili Karma MTK modeli şu şekilde gösterilmektedir (Finch & French, 2012):

$$P(x_{ij} = 1|Q_j) = P_{ij} = \sum_{g=1}^G \pi_g \left(\frac{\exp[\alpha_{ig}(\theta_{jg} - \beta_{ig})]}{1 + \exp[\alpha_{ig}(\theta_{jg} - \beta_{ig})]} \right) \quad (9)$$

Karma MTK modellerinde güçlük, ayırt edicilik ve şans parametresi genel MTK çatısındaki parametrelerle aynı anlama sahiptir. Bu nedenle, madde güçlüğü bireyin bir maddeyi doğru cevaplama olasılığı hakkında bilgi sağlar, ayırt edicilik maddenin ölçülen yapının farklı seviyelerine sahip bireyleri ne kadar iyi ayırt ettiğini gösterirken şans parametresi ise bireyin sadece şansa bağlı olarak maddeyi doğru bir şekilde yanıt verme olasılığının bir ölçüsüdür (De Ayala, 2009). Güçlük ve ayırt edicilik parametrelerinin sınıflar için farklılaşması, belirli bir madde için örtük sınıf üyeliğinin farklı performans gösterdiği anlamına gelmektedir (Cohen & Bolt, 2005). Örneğin analiz sonuçlarına göre evrende iki örtük sınıf bulunduğunu varsayalım. Bu durumda madde güçlük parametresinin (β_{ig}) sınıf 1 için sınıf 2'ye göre daha yüksek değere sahip olması, o maddenin sınıf 1'deki bireyler için daha zor olduğu anlamına gelmektedir. Benzer şekilde madde ayırt edicilik parametresinin (α_{ig}) sınıf 2 için daha yüksek değere sahip olması, o maddenin örtük yeteneğin farklı düzeylerine sahip bireyleri daha iyi ayırt edebildiği ifade edilebilmektedir (Finch & French, 2012). Karma MTK modelleri doğru/yanlış gibi iki kategorili puanlanan maddeler, Likert tipi dereceli ölçeklenen maddeler ve kısmi puanlanan maddeler için kullanılabilir (De Ayala & Santiago 2017).

İlgili Araştırmalar

İlgili literatür incelendiğinde simülasyon ve gerçek verilerin kullanıldığı Karma MTK modellerini ele alan çalışmaların olduğu görülmektedir (Choi, Alexeev & Cohen, 2015; Oliveri, Ercikan, Zumbo & Lawless, 2014; Oliveri & von Davier, 2011; Park, Lee & Xing, 2016; Sen, Cohen & Kim, 2016; Toker, 2016; Zhang, Orrill & Campbell, 2015). Alanyazında Karma MTK modelleri kullanılarak yapılan simülasyon çalışmalarında; madde sayısı, örneklem büyüklüğü, örtük sınıf sayısı gibi farklı koşulları ele alan çalışmalar bulunmaktadır (Cho, Cohen & Kim, 2013; Li, Cohen & Kim, 2009). Bu koşullara ek olarak

MTK ve Karma MTK modellerinde kayıp veri türü ve kayıp veri yüzdesi faktörlerinin manipüle edilerek parametre kestirimlerinin incelendiği araştırmalar da literatürde yer almaktadır (Edwards & Finch, 2017; Finch, 2008; Finch & French, 2012; Huang, 2020; Lee,2012).

Finch (2008)'de MTK modelleri için kayıp veri türü (tam, RKV, ROKV), kayıp veri oranı (%5, %15, %30), örneklem büyüklüğü (500, 1000), madde sayısı (20) ve MTK modeli (3PL) faktörlerini ele aldıkları simülasyon çalışmasında madde parametre kestirimlerini incelemişlerdir. Kayıp veri türlerinin RKV ve ROKV koşullarında ve kayıp veri yüzdesi arttıkça madde güçlük ve ayırt edicilik parametre kestirimlerinde daha yüksek RMSE değerleri elde edilirken şans parametresinin kestiriminde tam, RKV ve ROKV koşullarının ortalama RMSE değerleri düşük ve birbirine oldukça yakın çıkmıştır.

Li vd. (2009) çalışmalarında iki kategorili puanlanan MTK modelleri (Rasch, 2PL, 3PL), madde sayısı (6, 15, 30), örneklem büyüklüğü (600, 1200), sınıf sayısı (1, 2, 3, 4) olarak belirlenmiştir. Her bir örtük sınıf için yetenek parametresinin normal dağılım sergilediği kabul edilmiştir. Karma MTK modelleri için her bir örtük sınıfın madde güçlük parametreleri, ayırt edicilik ve şans parametreleri manipüle edilerek simülasyon çalışması yapılmıştır. İki ayırt edicilik parametre değeri kullanılarak veri üretilmiştir. Ortalama ve düşük performans gösteren grup için ayırt edicilik değeri 1; yüksek performans gösteren grup için ise 2 olarak belirlenmiştir. Madde güçlükleri ise -2 ve +2 aralığında değişecek şekilde üretilmiştir. Şans parametreleri ise zor, orta zorlukta ve kolay maddeler için sırasıyla 0,25; 0,2 ve 0,1 olarak üretilmiştir. Örtük sınıflar ise düşük, orta ve yüksek performans gösteren sınıflar olarak sınıflandırılmıştır. Madde sayısının 5 olduğu koşul için ilk 5 madde parametre seti kullanılarak, 30 madde koşulu için ise 15 madde seti iki kez kullanılarak simülasyon veri üretilmiştir. Madde sayısının 6 ve sınıf sayısının 2 ile ikiden daha fazla olduğu koşullar için 50000 iterasyon yapılmasına rağmen sonuç elde edilememiştir. Dolayısıyla madde sayısının altı olduğu koşul çıkarılarak analize devam edilmiştir. Madde sayısı arttıkça örtük sınıf üyeliğinin doğru belirlenme yüzdesi artmakta ancak örneklem büyüklüğü arttıkça örtük sınıf üyeliğinin belirlenmesi etkilenmemektedir. Simülasyon çalışmasından elde edilen sonuçlara göre tüm iki kategorili puanlanan Karma MTK modelleri için belirlenen koşullar altında BIC değeri en doğru modeli seçmektedir.

Finch ve French (2012) araştırmalarında Karma MTK modelleri için parametre kestirimi ve sınıflama doğruluğunu inceledikleri araştırmalarında madde sayısı (5, 15, 30), örtük sınıf sayısı (2, 3, 4), örneklem büyüklüğü (400, 1000, 2000) ve MTK modelleri (1PL, 2PL)

faktörlerini ele almışlardır. Li vd. (2009) çalışmasında ele alınan yönteme benzer olarak beş madde koşulu için ilk beş madde parametre seti kullanılmıştır, 30 madde koşulu için ise 15 madde seti iki kez kullanılmıştır. Madde sayısı ve örneklem büyüklüğü arttıkça madde parametre kestirimleri için elde edilen yanlılık değerlerinin azaldığı, sınıf sayısı arttıkça ise yanlılık değerlerinin arttığı tespit edilmiştir. Bundan sonraki araştırmalarda, Karma 3PL model araştırmaya dahil edilmediği için Karma 3PL modelin kullanılmasını önermişlerdir.

Lee (2012) çalışmasında Karma MTK modelleri (Rasch, 2PL), madde sayısı (10, 30), sınıf sayısı (1, 2), örneklem büyüklüğü (600, 1200), kayıp veri türü (tam, RKV, ROKV), kayıp veri oranı (%10, %20) ele aldıkları çalışmalarında sahte örtük sınıfların oluşumunda kayıp verinin etkisini incelemişlerdir. Madde sayısı ve örneklem büyüklüğü arttıkça madde parametre kestirimleri için daha düşük RMSE değerleri elde edilirken, kayıp veri türleri RKV, ROKV koşullarında ve kayıp veri oranı arttıkça daha yüksek RMSE değerleri elde edilmiştir.

Cho, Cohen ve Kim (2013) araştırmalarında madde sayısı (10, 30, 50), örneklem büyüklüğü (360, 1080, 3240) ve örtük sınıf sayısının (2, 3, 4) parametre kestirimi üzerindeki etkisini görmek için WinBUGS programı ile Karma Rasch modeline dayalı bir simülasyon çalışması gerçekleştirmişlerdir. Madde ve yetenek parametrelerinin normal dağılım sergilediği varsayılmıştır. Madde parametreleri Rost (1990)'un çalışmasında kullanılan madde parametreleri baz alınarak, yetenek parametreleri ise (-2,5 ile 2,5) aralığında, 0,5 kuadratik noktaları değişecek şekilde alınarak analiz gerçekleştirilmiştir. Madde güçlükleri için madde sayısı ve örneklem büyüklüğü arttıkça daha doğru kestirimler yapıldığı sonucuna ulaşılmıştır. Sonrasında gerçek veri için Karma Rasch modelini farklı programlarda (Mplus, Latentgold, Winmira, Winbugs) karşılaştırılarak model veri uyum indeksleri incelenmiştir. Tüm programlarda verinin en iyi iki sınıflı modele uyum sağladığı sonucuna ulaşılmışlardır.

Edwards ve Finch (2017), MTK modelleri (2PL, 3PL), madde sayısı (20, 30), örneklem büyüklüğü (1000, 2000), kayıp veri türü (RKV, ROKV), kayıp veri oranı (%5, %10 %15, %30) faktörleri ele alınarak madde parametre kestirimlerini inceledikleri simülasyon çalışmalarında madde güçlük ve ayırt edicilik parametre kestirimlerinde 3PL model için daha yüksek standart hata değerleri elde etmişlerdir. Ayrıca hem 2PL hem de 3PL modelin ROKV koşullarında madde parametre kestirimleri için daha yüksek hata değerleri bulunmuştur.

Huang (2020), örneklem büyüklüğü (1000, 2000), madde sayısı (20, 40), ROKV, %10 boş bırakılan ve %3 ulaşılamayan kayıp veri yüzdesi, 3PL MTK model faktörlerini ele alarak

gerçekleştirdiği simülasyon çalışmasında, ROKV türüne RKV gibi davranılmasının madde parametre kestiriminde hatayı artırdığı sonucuna ulaşmıştır. Madde sayısı ve örneklem büyüklüğü arttıkça madde ve yetenek parametre kestirimlerinin RMSE ve yanlılık değerlerinin azaldığı tespit edilmiştir.

Gerçek veri ile ilgili literatürde yer alan çalışmalara ise sırasıyla verilmiştir: Oliveri ve von Davier (2011), PISA 2006 uygulamasına katılan 30 OECD üyesi ülkenin matematik, okuma ve fen verisini kullanmıştır. Toplamda 103 fen, 48 matematik, 28 okuma alanından olmak üzere 179 madde analiz edilmiştir. Çoktan seçmeli maddeler ve iki kategorili puanlanan açık uçlu maddeler göz önünde bulundurulmuştur. Model veri uyumunu incelemek için tek boyutlu MTK modelleri (Rasch, 2PL) ve Karma MTK Modelleri (Karma Rasch, Karma 2PL) ele alınmıştır. Sonuç olarak bilgi kriter indeksleri incelendiğinde modelin Karma 2PL MTK modeline, tek boyutlu MTK modellerinden daha iyi uyum sağladığı görülmüştür.

Zhang vd. (2015) yaptıkları araştırmada matematik ve fen alanlarında öğrencilerin nasıl sınıflandırıldığına dair bilgi sağlamak amacıyla PISA 2009 uygulamasının yayımlanan maddelerinden oluşturulan bir ölçme aracı kullanmışlardır. Çoktan seçmeli ve açık uçlu maddelerin bir arada kullanıldığı ölçme aracında 15 matematik ile 16 fen alanına ait olmak üzere 31 madde bulunmaktadır. Fen dersi için özellikle uygun olan matematik maddelerinin seçimine dikkat edilmiştir. Örtük grupları belirlemek için matematik alt testi, fen alt testi ve matematik ve fen alt testleri birlikte ele alınarak üç ayrı analiz yürütülmüştür. PISA 2012 ölçme ve değerlendirme çerçevesinde dikkate alınan fen ve matematik yeterlikleri açısından sınıflar tanımlanmıştır. Sonuç olarak her bir alt test ve bu alt testlerin birlikte ele alınması durumlarında Karma Rasch modeli ile iki örtük sınıf bulunmuştur.

Sen vd. (2016), TIMSS 2011 sekizinci sınıf matematik verisi için en yüksek performans gösteren ülkelerden biri olan Güney Kore seçilerek Karma MTK modelleri ile veri analiz edilmiştir. Karma MTK modelleri için 18 çoktan seçmeli maddeden oluşan ve beşinci kitapçığı alan 369 birey analize dahil edilmiştir. Simülasyon çalışmasından elde edilen sonuçlara benzer olarak verinin iki sınıflı Karma Rasch modele daha iyi uyum sağladığı sonucuna ulaşılmıştır. Karma 2PL ve 3PL modellerin ise tek sınıflı modellere daha iyi uyum sergilediği görülmüştür.

Toker (2016), yaptığı araştırmada TIMSS 2011 sekizinci sınıf matematik verisi için örtük sınıf analizi ve Karma Rasch modeli kullanarak ülkeler arası bir karşılaştırma yapmıştır. Farklı eğitim sistemlerine sahip dört ülke (Türkiye, Amerika, Finlandiya, Singapur) seçilmiştir. Çoktan seçmeli ve açık uçlu (iki kategorili puanlanan) maddeler kullanılmıştır.

Araştırmada öncelikle örtük sınıf analizi kullanılarak üç örtük sınıf tespit edilmiştir. Örtük sınıflar, öğrencilerin cevap örüntülerine dayalı olarak araştırmacı tarafından yüksek beceri (highly skilled), orta beceri (moderately skilled) ve düşük beceri (somewhat skilled) düzeylerine göre gruplara ayrılmıştır. Ayrıca Karma Rasch modeli kullanılarak yapılan analizler sonucunda, örtük sınıf analizindeki sonuçlara benzer şekilde üç örtük sınıf bulunmuştur. Her iki analizden elde edilen parametrelerin standart hata değerleri birbirine yakın çıkmıştır. Fakat bu araştırma sadece Karma Rasch modeli kullanılarak sınırlandırılmıştır. Ayrıca bu araştırmada gerçek veri kullanıldığı için örneklem büyüklüğü, madde sayısı, grup sayısı gibi farklı koşulların ele alındığı durumlar incelenmemiştir. Toker (2016), farklı koşulların ele alındığı simülasyon çalışmasının yapılmasını önermiştir.

Park vd. (2016), tarafından yapılan madde parametre kaymasının incelendiği bir araştırmada ise TIMSS 1999, 2003 ve 2007 sekizinci sınıf matematik verilerinin 21 ankor (ortak) maddesi kullanılmıştır. Ankor maddeler farklı dönemlerde yapılan testler arasında yetenek kestirimlerinin ölçeklenmesinde ve eşitleme çalışmalarında kullanılmaktadır. Ankor maddelerin değişmez olduğu varsayılarak, tek boyutlu MTK modellerine göre parametre kestirimi yapılmıştır. Bu araştırmadan elde edilen sonuçlar, üç TIMSS uygulamasında da örtük sınıfların oluştuğunu ve madde parametrelerinin değişmezlik özelliğinin sağlanmadığını göstermiştir. Ayrıca TIMSS verisinin iki sınıflı Karma 3PL modele, tek boyutlu 3PL MTK modelinden daha iyi uyum sağladığı sonucuna ulaşılmıştır. Bu araştırmanın ikinci aşamasında ise simülasyon çalışması yapılmıştır. TIMSS 1999 verisine dayalı olarak üretilen simülasyon veride, üç parametrelili Karma MTK modeli kullanılarak madde ve yetenek parametrelerinin farklı koşullardaki kestirimleri incelenmiştir. Sonuç olarak hem gerçek hem simülasyon veri kullanılarak yapılan bu araştırmada, farklı simülasyon koşullarının incelemesi önerisinde bulunmuşlardır.

BÖLÜM 3

YÖNTEM

Bu bölümde araştırmanın modeli, araştırma deseni, simülasyon koşulları, Karma MTK modellerinde parametre kestirimde kullanılan yöntemler, etiket değiştirme, simülasyon veri analizi ve gerçek veri analizine ilişkin bilgilere yer verilmiştir.

Araştırmanın Türü

Araştırmada ele alınan değişkenler (faktörler) için farklı koşullar altında modellerin Karma MTK modelleri ile analiz edilmesi sonucu madde parametreleri için elde edilen ortalama RMSE ve yanlılık değerleri ile ele alınan faktörlerin etkileşim etkileri incelenmiş ve her bir modelin farklı koşulların kombinasyonlarında sınıflama doğruluğu yüzdeleri karşılaştırılmıştır. Ayrıca gerçek veri örneği için TIMSS 2015 8. sınıf Türkiye fen verisi simülasyon koşulları ile benzerlik gösterecek şekilde Karma MTK modelleri ile analiz edilmiştir. Bu kapsamda, bu çalışma simülasyon ve gerçek veriye dayalı temel bir araştırmadır.

Simülasyon Koşulları

Simülasyon çalışmaları, belirli parametreler kullanılarak rastgele verilerin üretimini sağlayan bilgisayar destekli araştırmalar olarak tanımlanmaktadır (Rubinstein & Kroese, 2017). Son yıllarda gerçek uygulamalardan elde edilen veriler kullanılarak ortaya konulan bilgilerin sınırlı kalması, gerçek veri toplanmasının maliyetli olması ve fazla zaman gerektirmesi gibi nedenlerden dolayı simülasyon çalışmalarına yer verilmektedir (Bulut & Sünbül, 2017).

İlgili alanyazındaki Karma MTK modelleri kullanılarak yapılan simülasyon çalışmalarında, MTK modelleri, madde sayısı, örneklem büyüklüğü, yetenek dağılımları, örtük sınıf sayısı

gibi farklı koşulların ele alındığı görülmektedir. Sen ve Cohen (2019) tarafından yapılan derleme çalışmasına göre bu araştırmalarda madde sayısının 4 ile 470 arasında (Cho, Lee, & Kingston, 2012; Jilke, Meuleman & Walle, 2015), örneklem büyüklüğünün 99 ile 251,278 arasında (Glück, Machat, Jirasko, & Rollett, 2002; Oliveri & von Davier, 2011), örtük sınıf sayısının 1 ile 10 arasında (Gollini & Murphy, 2014; Jilke vd., 2015; Liu, Liu & Li, 2018) olduğu farklı simülasyon koşulları dikkate alınmıştır. Madde parametreleri için ise, hem sabit hem de manipüle edilen koşulların ele alındığı çalışmalar yapılmıştır. Ayrıca iki kategorili puanlanan tek boyutlu (Rasch, 1PLM, 2PLM, 3PLM) ele alınarak farklı modeller için model veri uyumu incelenmiştir (Alexeev vd., 2011; Cho vd., 2013; Finch & French, 2012; Li vd., 2009; Park vd., 2016; Sen vd., 2016). Kayıp veri türleri olarak ise tam, RKV ve ROKV türleri ile kayıp veri yüzdesi olarak %5, %10, %15, %20 değerlerinin alındığı MTK ve Karma MTK modellerinin ele alındığı çalışmalar bulunmaktadır (Edward & Finch, 2017; French, 2008; Huang, 2020; Lee, 2012). Bu araştırma kapsamında ise alan yazındaki diğer çalışmalarda dikkate alınmayan Karma MTK modellerin 3 PL formu için kayıp veri türü ve kayıp veri oranları dikkate alınarak simülasyon çalışması gerçekleştirilmiştir. Bu simülasyon çalışmasının koşulları, üç farklı Karma MTK modeli (Rasch, 2PL ve 3PL), iki farklı sınıf (2 ve 3), iki madde sayısı (10 ve 30), iki örneklem büyüklüğü (600 ve 1000), üç kayıp veri mekanizması (tam veri, rastgele kayıp veri (RKV), rastgele olmayan kayıp veri (ROKV) ve iki kayıp veri yüzdesi (%10 ve %20) olmak üzere 144 koşuldan oluşmaktadır. Her bir koşul için yüz replikasyon simüle edilmiştir.

Örtük Sınıf Sayısı

Örtük sınıfların belirlenmesi amacıyla yapılan farklı simülasyon çalışmalarında manipüle edilen değişkenlerden biri homojen alt sınıflardır. Heterojen bir evrende bulunan farklı sayıdaki homojen alt sınıflar birer koşul olarak alınmıştır. Sen ve Cohen (2019) tarafından yapılan derleme çalışmasına göre çalışmalarda genellikle örtük sınıf sayısının bir ile on arasında ele alındığı belirtilebilmektedir (Cho, Cohen & Kim, 2013; Finch & French, 2012; Gollini & Murphy, 2014; Jilke vd., 2015; Li vd., 2009; Liu vd., 2018; Sen vd., 2016). Ancak model veri uyumu çalışmalarının sonuçlarına göre verilerin genellikle iki veya üç sınıflı Karma MTK modeline uyum sağladığı ifade edilmektedir (Finch & French, 2012; Park vd., 2016). Ayrıca TIMSS 2015 Türkiye 8. Sınıf fen alt testinden elde edilen verilerin Karma MTK modelleri ile analiz edilmesi sonucunda da 4 ve 5 sınıflı modellerin daha yüksek BIC değerlerine sahip olduğu görülmüştür. Bu simülasyon çalışmasında örneklemdeki örtük sınıf

sayısının 2 ve 3 olduđu iki farklı koşul dikkate alınmıřtır. Sınıf profilleri Li vd. (2009)'da ele alınan çalışmaya benzer olarak yapılmıřtır. Farklı madde parametrelerinin deęerlerini temsil eden iki sınıfın profili, bir sınıftaki bireylerin maddeleri doęru cevaplamada daha iyi performans sergilediđini, dięer sınıftaki bireylerin ise maddeleri doęru cevaplamada daha düşük performans sergilediđini göstermektedir. Üç sınıfın profili ise bireylerin maddeleri doęru cevaplamada; birinci sınıftaki bireylerin daha iyi performans, ikinci sınıftaki bireylerin orta düzey bir performans, üçüncü sınıftaki bireylerin ise daha düşük performans sergilediđini göstermektedir.

Örneklem Büyüklüğü

Li vd. (2009), Karma Rasch modelleri için madde sayısının 15 ile 30 arasında olduđu durumlarda örneklem büyüklüğünün 600 olmasının uygun olacađını bildirmişlerdir. Ayrıca Li vd. 15 maddelik bir test için hem Karma 2PL hem de Karma 3PL modeller için 1 ile 4 sınıflı bir model için örneklem büyüklüğünün 600 alınmasının yeterli olacađını önermişlerdir. Cho vd. ise (2013), Karma Rasch model için 360'tan büyük bir örneklem büyüklüğü kullanılabileceđini önermektedir. Cohen ve Bolt (2005) 1000 örneklem büyüklüğü ile Karma 3PL modelini başarıyla uygulamışlardır. Bu çalışmada ise örneklem büyüklüğü 600 ve 1000 olarak alınmıřtır.

Madde sayısı

İlgili alanyazında yapılan çeşitli simülasyon çalışmalarında madde sayısı manipüle edilen deęişkenlerden biri olmuřtur. Sen ve Cohen (2019) tarafından yapılan derleme çalışmasına göre madde sayısının 4 ile 470 arasında (Cho vd., 2012; Jilke vd., 2015) deęiřtiđi görülmektedir. Ancak madde sayısının 10 ve 10'dan az olduđu koşullar için iterasyon sayısı artırılmasına rağmen sonuç elde edilememiřtir. TIMSS 2015 8.sınıf uygulamasında ise kitapçıklardaki çoktan seçmeli madde sayısı 11 ile 17 arasında deęişmektedir. Bu arařtırmada madde sayısı için koşullar 10 ve 30 olarak belirlenmiřtir.

Madde ve Yetenek Parametrelerinin Dağılımı

İlgili alanyazında yapılan simülasyon çalışmalarında örtük sınıflara ait madde ve yetenek parametrelerine ilişkin farklı dağılımların dikkate alındıđı görülmektedir (Alexeev vd., 2011; Cho vd., 2013; Li vd., 2009; Rost, 1990). Rost (1990) yaptıđı simülasyon

araştırmasında 1800 kişi ve 10 madde için Karma Rasch modeli ile veri üretimi sağlamıştır. Madde güçlük parametrelerinin dağılımı -2,7 ile 2,7 aralığında olacak şekilde belirlemiştir. Li vd. (2009), Karma MTK modelleri için her bir örtük sınıfın madde güçlük parametreleri, ayırt edicilik ve şans parametreleri manipüle edilerek simülasyon çalışması yapmıştır. İki ayırt edicilik parametre değeri kullanılarak veri üretilmiştir. Ortalama ve düşük performans gösteren grup için ayırt edicilik değeri 1; yüksek performans gösteren grup için ise 2 olarak belirlenmiştir. Madde güçlükleri ise (-2,7;+2,7) aralığında farklılık göstermektedir. Şans parametreleri ise zor, orta zorlukta ve kolay maddeler için sırasıyla 0,25; 0,2 ve 0,1 olarak belirlenmiştir. Alexeev vd. (2011) ise yeteneğin standart normal dağılımdan rastgele oluşturulduğu varsayılmıştır. Madde parametrelerinin ise ayırt edicilik ve güçlük için iki farklı dağılımdan rastgele çekildiği varsayılmıştır. Cho vd. (2013) ise yaptıkları simülasyon çalışmasında madde ve yetenek parametrelerinin normal dağılımdan çekildiğini varsaymaktadır.

Madde ve yetenek parametrelerinin önsel bir normal dağılım kullanılarak kestirilmesi daha doğru kestirimler yapılabilmesine olanak sağlamaktadır (Johnson& Albert,1998). Dolayısıyla bu çalışmada Li vd. (2009)'un çalışmasındaki madde parametrelerine benzer şekilde veri üretilmiştir. Madde güçlük parametre değerleri 10 madde koşulu için -2,7 ile +2,7 aralığında; 30 madde içinse -3 ile +3 aralığında tekdüze (uniform) dağılıma dayalı olarak rastgele bir şekilde üretilmiştir. Şans (guess) parametreleri ise zor,orta zorlukta ve kolay maddeler için sırasıyla 0,25; 0,2 ve 0,1 alınarak rastgele üretilmiştir (Li vd.,2009). Madde güçlük parametre değerleri Mplus input dosyasına birinci threshold, şans parametre değerleri ise ikinci threshold olarak yazılmıştır (Muthén & Muthén, 2017). Madde ayırt edicilik parametreleri ise düşük ve orta performans gösteren sınıf için 1, yüksek performans gösteren sınıf içinse 2 olarak Li vd. (2009) çalışmalarına benzer olarak belirlenmiştir. Yetenek parametreleri standart normal dağılımdan ($N(0,1)$) çekilerek *runif* fonksiyonu ile rastgele üretilmiştir. Tablo 1'de Karma MTK modellerinde 10 madde için üretilen madde parametre değerleri verilmektedir:

Tablo 1

Karma MTK modellerinde 10 madde için üretilen madde parametre değerleri

Madde	Sınıf1			Sınıf2			Sınıf1			Sınıf2			Sınıf3		
	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	2	-2,7	0,10	1	2,7	0,25	2	-2,7	0,10	1	-0,5	0,20	1	2,7	0,25
2	2	-2,1	0,10	1	2,1	0,25	2	-2,1	0,10	1	-0,4	0,20	1	2,1	0,25
3	2	-1,5	0,10	1	1,5	0,25	2	-1,5	0,10	1	-0,3	0,20	1	1,5	0,25
4	2	-0,9	0,10	1	0,9	0,25	2	-0,9	0,10	1	-0,2	0,20	1	0,9	0,25
5	2	-0,3	0,20	1	0,3	0,20	2	-0,3	0,20	1	-0,1	0,20	1	0,3	0,20
6	1	0,3	0,20	2	-0,3	0,20	1	0,3	0,20	1	0,1	0,20	2	-0,3	0,20
7	1	0,9	0,25	2	-0,9	0,10	1	0,9	0,25	1	0,2	0,20	2	-0,9	0,10
8	1	1,5	0,25	2	-1,5	0,10	1	1,5	0,25	1	0,3	0,20	2	-1,5	0,10
9	1	2,1	0,25	2	-2,1	0,10	1	2,1	0,25	1	0,4	0,20	2	-2,1	0,10
10	1	2,7	0,25	2	-2,7	0,10	1	2,7	0,25	1	0,5	0,20	2	-2,7	0,10

Tablo 1’de örtük sınıf sayısının 2 ve 3 olduğu durumlar için tüm sınıf sayılarına göre üretilen madde parametre değerleri sunulmuştur. İki sınıflı durum için; Sınıf 1’de madde güçlük parametrelerinin kolaydan zora doğru sıralanması Sınıf 1’deki bireylerin maddeleri doğru cevaplama da daha düşük performans gösterdiği anlamına gelmekte iken Sınıf 2’de madde güçlük parametrelerinin zordan kolayca doğru sıralanması Sınıf 2’deki bireylerin maddeleri doğru cevaplama da daha yüksek performans gösterdiği anlamına gelmektedir. Her iki sınıfta da madde ayırt edicilik ve şans parametrelerinin madde güçlük değerleri ile uyumlu olduğu görülmektedir.

Üç sınıflı durum için; Sınıf 1’de madde güçlük parametrelerinin kolaydan zora doğru sıralanması Sınıf 1’deki bireylerin maddeleri doğru cevaplama da daha düşük performans gösterdiği, Sınıf 2’de madde güçlük parametrelerinin orta güçlükte olması Sınıf 2’deki bireylerin maddeleri doğru cevaplama da orta düzey performans gösterdiği, Sınıf 3’de madde güçlük parametrelerinin zordan kolayca doğru sıralanması Sınıf 3’deki bireylerin maddeleri doğru cevaplama da daha yüksek performans gösterdiği anlamına gelmektedir. Her üç sınıfta da madde ayırt edicilik ve şans parametrelerinin madde güçlük değerleri ile uyumlu olduğu görülmektedir. Karma Rasch modelde sadece b (threshold) parametre değerleri, Karma 2PL modelde a (slope) ve b (threshold) parametre değerleri, Karma 3PL modelde ise a (slope), b (threshold) ve c (guessing) parametreleri koda yazılarak veriler üretilmiştir.

Kayıp veri

Rubin (1976) kayıp veri mekanizmaları için Tamamıyla Rastgele Kayıp Veri (TRKV), Rastgele Kayıp Veri (RKV) ve Rastgele Olmayan Kayıp Veri (ROKV) ve olmak üzere üç tür sınıflandırma yapmıştır. TRKV sistematik bir nedeni olmadığı düşünülebilir; yani, eksik veriler, gözlenen verilerin basit rastgele bir örneğidir (Schafer, 1997). Veriler RKV olduğunda, bir değerin kayıp olma olasılığı, bireyin ölçülebilir bazı diğer özelliklerine bağlıdır ancak bulunduğu değişkendeki gözlenen değişkenlerle ilişkisizdir. ROKV ise, bir değişken değerinin eksik olma olasılığı, modeldeki diğer değişkenlerle ve kendi değişkeniyle ilişkili olduğu ifade edilebilir. Gerçek test verileri bağlamında, araştırmacılar, bireyin soru cevaplarını bilmediğine inanması (De Ayala, Plake & Impara, 2001) veya yanlışlıkla maddeyi atlayarak (Huisman & Molenaar, 2001) geçmesi gibi madde yanıtlarının eksik olmasının çeşitli nedenlerini belirlemişlerdir. İlk durum, bir RKV (cevapsızlık temelde yatan örtük özellik ile ilişkiliyse, ancak verilmiş olan gerçek madde cevabı değilse) veya ROKV (cevapsızlık, verilmiş olan gerçek madde cevabı ile doğrudan ilişkiliyse) mekanizmasını, ikinci durum ise yanıt vermeme ile hiçbir sistematik mekanizmanın ilişkilendirilmediği TRKV mekanizmasını temsil eder.

RKV ve ROKV üretiminde Finch (2008) çalışması baz alınmıştır. Finch'in çalışmasında, hedef maddelerde kayıp veri olasılığının, hedef maddeden başka doğru madde sayısının toplamıyla ters orantılı olması durumunda kayıp veri mekanizmasının türünün RKV olduğu varsayılmıştır. Bu şekilde üretilen kayıp gözlemler, RKV değeri olarak düşünülebilir, çünkü kayıp değer olasılığı, gözlenen bir başka değişkenle ilgilidir yani hedef maddeler dışındaki doğru madde sayısının toplamıdır. RKV üretimi için daha spesifik prosedürler aşağıda anlatılmaktadır:

10 maddelik veri seti için en zor olan 3 madde, hedef (target) madde olarak seçilmiştir. Kalan 7 maddeye göre toplam puan hesaplanmıştır. Hedef maddeler dışındaki toplam puanlara dayanarak, simülasyonlar her sınıf için dört kırılma (fractile) (0 -1, 2 -3, 4 -5, 6-7) olarak ayrılmıştır (Finch, 2008).

Hedef maddeler üzerindeki kayıp cevap olasılıklarının dört farklı değeri ile dört kırılma (fractile) oluşturulmuştur. Kırılmalarda bu olasılıkların ortalaması, kayıp veri oranlarına yani %10 ve %20 eşit olacak şekilde tasarlanmıştır. Bu şekilde kayıp veriler üretilerek, hedef maddeler dışındaki maddelerin toplam puanlarına dayanarak düşük ve yüksek yetenekli bireyler için cevap örüntüleri oluşturulmuştur.

Ayrıca kayıp veri durumunun, bireylerin hedef maddelere doğru cevap verip vermemesi ile doğrudan ilgili olduğu durumda, kayıp veri mekanizmasının tipinin ROKV olduğu varsayılmıştır. Veri setinin tamamı göz önüne alındığında, maddeleri doğru cevaplayan kişilere daha düşük bir olasılıkla kayıp veri atanmış ve maddeyi yanlış yapan kişilere daha yüksek bir olasılıkla kayıp veri atanmıştır (Finch, 2008).

Bu yüksek ve düşük olasılıkların ortalaması, simülasyon çalışması için %10 ve %20 olan ROKV lerin toplam yüzdesine eşit olarak belirlenmiştir. Eksik cevapların toplam yüzdeleri her bir dosya için kontrol edilmiştir. ROKV ve ROKV üretmek için kullanılan R kodu Ek 1'de sunulmuştur. Tüm simülasyon koşulları Tablo 2’de sunulmaktadır:

Tablo 2

Simülasyon Koşulları

Örneklem Büyükliği	Karma MTK Modelleri	Sınıf Sayısı	Madde Sayısı	Kayıp Veri Türü	Kayıp Veri Yüzdesi
600	Karma Rasch	2	10	Tam	%10
1000	Karma 2PL	3	30	RKV	%20
	Karma 3PL			ROKV	

Tablo 2’de görüldüğü gibi 2 örneklem büyüklüğü (600,1000), 3 Karma MTK modeli (Karma Rasch, Karma 2PL, Karma 3PL), 2 örtük sınıf sayısı (2, 3), 2 madde sayısı (10, 30), 3 kayıp veri türü (tam, RKV, ROKV) ve 2 kayıp veri yüzdesi (%10, %20) olmak üzere simülasyon koşulları verilmektedir. Toplamda $2 \times 3 \times 2 \times 2 \times 3 \times 2 = 144$ koşul ele alınarak ve 100 replikasyon gerçekleştirilerek $144 \times 100 = 14400$ adet veri üretilmiştir.

Karma MTK Modelleri İçin Kestirim Yöntemleri

Standart MTK modellerinde parametre tahminleri, marjinal maksimum olabilirlik tahmini (marginal maximum likelihood estimation-MMLE; Bock & Aitkin, 1981), ortak maksimum olasılık tahmini (joint maximum likelihood estimation- JMLE) ve koşullu maksimum olabilirlik (conditional maximum likelihood estimation-CMLE) algoritmaları kullanılarak yapılmaktadır. Prensip olarak, madde parametreleri ve yetenek parametreleri için ML tahminleri standart sayısal optimizasyon prosedürleri (örneğin, Newton-Raphson) kullanılarak elde edilebilir. Bu prosedürler, olasılık fonksiyonunu en üst düzeye çıkaran parametre tahminlerini bulmayı amaçlayan iterasyonlara dayanmaktadır. JMLE yönteminde madde ve birey parametreleri birlikte kestirilmektedir. CMLE ve MMLE yöntemlerinde ise

birey parametrelerini tahmin etmeden madde parametreleri kestirilebilmektedir. CMLE, Rasch modellerinin ailesi ile sınırlı olmasına rağmen, MMLE, MTK çerçevesindeki tüm model türleri için kullanılabilir. Bock ve Aitkin (1981), MTK modellerindeki parametrelerin CML ve MML tahminlerini elde etmek için kullanılabilecek beklenti maksimizasyonu (expected maximization-EM; Dempster, Laird & Rubin, 1977) adlı bir algoritma önermişlerdir (von Davier & Rost, 2007). Ek olarak, MTK parametre tahminleri için Bayes çözümleri de mevcuttur.

Karma MTK modelleri için parametre kestirimi kestirilecek parametre sayısının fazla olması sebebiyle standart MTK modellerine göre daha karmaşıktır. Karma MTK modeller için parametre kestirimi Bayes algoritmaları (MCMC) veya maksimum olasılık tahmini (MLE) teknikleri kullanılarak uygulanabilir. Bu iki yaklaşımın uygulamaları arasında birtakım farklılıklar vardır. Bayes yaklaşımı, kestirilecek parametreler için önsel bilgi gerektirirken, maksimum olasılık yaklaşımı, madde parametrelerini sabit etkiler olarak ele alır ve yetenek ve madde parametreleri hakkında önceden bilgi sahibi olmayı ihmal eder (Cho & Suh, 2012). Her iki yöntem de kayıp veri olduğu durumlarda kullanılabilir ancak MLE kestirim yönteminde TRKV ve RKV türlerinde daha sağlam (robust) kestirimler elde edilmiştir (Rubin, 1976). Yapısal Eşitlik Modelleme çerçevesinde Enders ve Bandalos (2001)'de Tam Bilgi Maksimum Olabilirlik Kestirim (FIML) yöntemi ile liste bazında silme (listwise deletion), çiftler bazında silme (pairwise deletion) ve benzer cevap örüntüleri atama (similar response pattern imputation) yöntemlerinden daha etkili ve yanlı olmayan kestirimler elde etmişlerdir. Aynı zamanda Edwards ve Finch (2017) çalışmalarında MTK modellerinden 2PL ve 3PL modelleri ele alarak RKV ve ROKV durumları için parametre kestirimlerini inceledikleri çalışmadan FIML yönteminin daha iyi sonuç verdiğini gözlemlemişlerdir. Adından da anlaşılacağı gibi, Tam Bilgi Maksimum Olabilirlik (FIML) yöntemi, mevcut tüm verilerle bir maksimum olasılık uydurma (fitting) işlevi kullanarak model parametrelerini tahmin eder. Böylelikle, kayıp veriye sahip bireyler, sahip oldukları tüm bilgiler kullanılarak parametre tahminine dahil edilir ve kayıp değerlere sahip oldukları değişkenler için bunlar göz ardı edilir. Ek olarak, FIML, kayıp değerlerin atanmasını içermez, böylece bu yöntemin kullanımını diğer önerilen yaklaşımlardan bazılarına, özellikle de veri atamaya dayalı olan yöntemlere göre daha az zahmetli hale getirir. Son olarak, FIML, MTK dahil örtük değişken modellerinin uydurulması için çoğu istatistiksel yazılımda (Mplus User Guide, 2017) mevcuttur ve pratikte kullanımı çok kolaylaştırır. Mplus yazılımında kayıp veriler için kestirim yöntemi ML veya MLR olarak seçildiğinde

FIML ile kestirim yapılmaktadır. MLR kestirim yöntemi normallik ve gözlemlerin bağımsızlığı varsayımları sağlanmadığında, standart hatalar ve ki-kare test istatistiği ile güçlü parametre kestirimleri yapılmaktadır (Mplus User Guide, Version 8, 2017). Karma MTK modeli kapsamında kestirilecek parametreler; yetenek parametreleri, madde güçlük parametreleri, ayırt edicilik ve şans parametreleri ve örtük sınıf üyeliğidir. Madde güçlüğü ve yetenek dağılımı için normal bir önsel dağılım kullanılmaktadır. Cho vd., (2013), Karma MTK modellerinde normal önsel dağılım kullanımının mantıklı kestirimler yapmaya yardımcı olduğunu belirtmektedir. Örtük sınıf üyeliği ise her bir g kategorisinin çok terimli dağılıma sahip olduğu varsayımı ile kestirilmektedir (McLachlan & Peel, 2000).

Karma MTK modellerinde parametre kestirimi yapabilmek için kullanılan yöntemlere dayalı olarak çeşitli bilgisayar yazılım paketleri mevcuttur. JMLE yöntemi için R programında yer alan mixRasch (Willse, 2009) paketi, MMLE yöntemi için Mplus bilgisayar yazılımı (Muthén & Muthén, 2017), EM algoritmasının kullanıldığı Latent GOLD (Vermunt & Magidson, 2005) bilgisayar yazılımı, CMLE yöntemi için R yazılımı içindeki (von Davier, 2001) mRm (Preinerstorfer, 2011) ve psychomix (Frick, Strobl, Leisch & Zeileis, 2012) paketleri ile Winmira (von Davier, 2001) yazılım programları kullanılmaktadır. Bu araştırmada Karma MTK modellerinin tümü için kestirim yapabilen Mplus yazılım programı kullanılmıştır (Muthén & Muthén, 2017).

Mplus yazılımı esnek modelleme kapasitesine sahiptir ve çok düzeyli yapıya sahip kategorik ve sürekli değişken veriler için faktör analizi, Karma MTK modellerin ve SEM uygulamalarının analizlerini gerçekleştirebilmektedir. Mplus ayrıca maksimum olabilirlik (ML) tahmini, en küçük kareler (least square) tahmini ve Bayes gibi çeşitli model tahmin yöntemleri ile donatılmıştır (Muthén & Asparouhov, 2006; Nylund, Asparouhov & Muthén, 2007).

Etiket Değiştirme (Label Switching)

Karma MTK modellerde tahmin edilen sınıfların sayısı ve doğası hakkında bilgi sahibi olunmadığı için bazen Sınıf 1 için kestirilen parametreler Sınıf 2 olarak veya Sınıf 2 için kestirilen parametreler Sınıf 1 olarak etiketlendirilebilmektedir (Jeon & De Boeck, 2018; McLachlan & Peel, 2000; Park vd., 2016). Bu tip etiket değiştirme Bayes ve en çok olabilirlik (Maximum Likelihood-ML) kestirimlerinde meydana gelebilmektedir (Cho vd., 2010; Cho vd., 2013; Finch & French, 2012). Sınıf etiketleri veri setleri arasında değiştirildiği için, potansiyel olarak yanlış etiketlenmiş sınıflar üzerinden toplanacak parametre tahminleri

istenmeyen bir durum oluşturmaktadır. Bu durumda kestirilen madde parametre değerleri başlangıç değerleri olarak alınarak etiket değiştirme problemi çözülebilmektedir (Kutscher, Eid, & Crayen, 2019). Simülasyon çalışmalarında etiket değiştirme, her bir iterasyondaki sınıf büyüklükleri incelenerek manuel olarak kontrol edilebilir (Sen & Cohen, 2019). Bu araştırmada madde parametrelerini üretmede kullanılan değerler Mplus input dosyasına yazılarak etiket değiştirme problemi çözülmüştür. Böylece 1.sınıf için madde güçlük parametre değerleri negatif değerlerden pozitif değerlere doğru sıralanarak 1. sınıfın düşük performans gösteren sınıf olarak etiketlenmesi, 2.sınıf için madde parametre değerlerinin -0,5 değerinden +0,5 değerine göre sıralanarak 2.sınıfın orta düzey performans gösteren sınıf olarak etiketlenmesi, 3.sınıf için ise madde güçlük parametre değerlerinin pozitif değerlerden negatif değerlere göre sıralanarak 3. sınıfın yüksek performans gösteren sınıf olarak etiketlenmesi garanti altına alınmıştır.

Simülasyon Veri Analizi

Araştırmada veri üretimi için R programı kullanılmıştır. Karma MTK modelleri Mplus programı ile analiz edilebilmektedir. Ancak her bir koşul için 100 replikasyon gerçekleştirilmiş ve 144 koşul için elde edilen 14400 (144x100) veri dosyasının tek tek Mplus programında analiz edilmesinin oldukça zor olması sebebiyle veriler, Mplus programı ile R programı arasında entegrasyon sağlayabilen “Mplus Automation” paketi ile analiz edilmiştir. 100 replikasyon için kullanılacak giriş (input) dosyaları ile elde edilen çıktı (output) dosyaları da “Mplus Automation” paketi ile üretilip otomatik olarak çalıştırılmış ve analiz edilmiştir. Birden çok input dosyası üretmek için kullanılan Mplus kodu Ek 2’de sunulmaktadır.

Bu simülasyon çalışmasında, Karma MTK modellerinin performansı iki kriter temelinde değerlendirilmiştir:

1. Parametre Doğrulanması
2. Sınıflama Doğruluğu Yüzdesi

Parametre Doğrulanması (The Accuracy of Item Parameter Recovery)

Kayıp verilerin varlığında Karma MTK modellerinin performansını değerlendirmek için madde parametreleri üzerinde bir doğrulama analizi yapılmıştır. Bir tahmin edicinin kalitesini değerlendirmek için hata kareler ortalaması karekökü (root mean square error-

RMSE) ve yanlılık (BIAS) değerleri kullanılmaktadır (Feinberg & Rubright,2016). RMSE, hataların karesinin beklenen değerini temsil etmektedir. Bu çalışmada kestirilen madde güçlük parametre değerleri için madde sayısı, sınıf sayısı ve replikasyon sayısı bilgileri kullanılarak aşağıdaki eşitlikler yardımıyla hata kareler ortalaması karekökü (root mean square error-RMSE) ve yanlılık (BIAS) değerleri hesaplanmıştır:

$$RMSE(\beta_i) = \sqrt{\frac{\sum_{r=1}^R \sum_{i=1}^I \sum_{g=1}^C (\hat{\beta}_{igr} - \beta_{ig})^2}{RIC}} \quad (10)$$

$$BIAS(\beta_i) = \frac{\sum_{r=1}^R \sum_{i=1}^I \sum_{g=1}^C (\hat{\beta}_{igr} - \beta_{ig})}{RIC} \quad (11)$$

Bu eşitliklerde $\hat{\beta}_{igr}$ g sınıfındaki i maddesi için R replikasyon sonucu kestirilen madde güçlük parametresini, β_{ig} g sınıfındaki i maddesinin gerçek (true) değerini, R replikasyon sayısını, I madde sayısını ve C sınıf sayısını temsil etmektedir.

Madde güçlük parametrelerinin incelenmesine ek olarak, 3PL modelinde ayırt edicilik ve şans parametrelerinin, 2PL modelde ise ayırt edicilik parametre değerleri için RMSE ve yanlılık değerlerini de incelemek gerekmektedir. Madde ayırt edicilik parametresi için RMSE ve yanlılık (BIAS) değerleri aşağıdaki formüllerle hesaplanmıştır:

$$RMSE(\alpha_i) = \sqrt{\frac{\sum_{r=1}^R \sum_{i=1}^I \sum_{g=1}^C (\hat{\alpha}_{igr} - \alpha_{ig})^2}{RIC}} \quad (12)$$

$$BIAS(\alpha_i) = \frac{\sum_{r=1}^R \sum_{i=1}^I \sum_{g=1}^C (\hat{\alpha}_{igr} - \alpha_{ig})}{RIC} \quad (13)$$

Bu eşitliklerde $\hat{\alpha}_{igr}$ g sınıfındaki i maddesi için R replikasyon sonucu kestirilen madde ayırt edicilik parametresini, α_{ig} g sınıfındaki i maddesinin gerçek (true) değerini temsil etmektedir. Yine benzer şekilde şans parametresi için RMSE ve yanlılık (BIAS) değerleri;

$$RMSE(Y_i) = \sqrt{\frac{\sum_{r=1}^R \sum_{i=1}^I \sum_{g=1}^C (\hat{Y}_{igr} - Y_{ig})^2}{RIC}} \quad (14)$$

$$BIAS(Y_i) = \frac{\sum_{r=1}^R \sum_{i=1}^I \sum_{g=1}^C (\hat{Y}_{igr} - Y_{ig})}{RIC} \quad (15)$$

Bu eşitliklerde $\hat{Y}_{igr}$ g sınıfındaki i maddesi için R replikasyon sonucu kestirilen şans parametresini, Y_{ig} g sınıfındaki i maddesinin gerçek (true) değerini temsil etmektedir.

RMSE ve yanlılık (BIAS) değerlerini hesaplamadan önce her bir Karma MTK modeli için (Rasch, 2PL ve 3PL) gerçek parametre değerleri ile kestirilen parametre değerlerini ortak bir

ölçeğe (common scale) getirerek karşılaştırma yapabilmek için aşağıdaki eşitlik kullanılarak ortalama eşitleme (mean equating) yapılmıştır (Kolen & Brennan,2004):

$$\beta_i^* = \hat{\beta}_i - (\bar{\hat{\beta}}_T - \bar{\hat{\beta}}_B) \quad (16)$$

Bu eşitlikte B alt indisi gerçek (true) parametrelerin ölçeğini, T alt indisi kestirilen madde parametrelerinin ölçeğini, güçlük parametre değerlerinin ortalamasını, $\bar{\hat{\beta}}_T$ ve $\bar{\hat{\beta}}_B$ ise sırasıyla kestirilen ve gerçek ortalama madde güçlük parametre değerlerini ve $\hat{\beta}_i$ eşitlenmiş madde güçlük parametre değerini göstermektedir.

Ana Etkiler ve Etkileşim Etkileri İçin ANOVA

Çoğu parametrik istatistiksel modelde olduğu gibi, ANOVA normallik, varyans homojenliği ve gözlemlerin bağımsızlığı gibi çeşitli varsayımlara dayanmaktadır (Howell, 2014). Veriler hakkında doğru istatistiksel çıkarımlar yapabilmek için bu varsayımların karşılanması gerekmektedir. ANOVA'da, modelde hata terimlerinin normal dağılım sergilediği varsayılmaktadır. Ayrıca, hata terimlerinin gözlemden bağımsız olduğu varsayılmaktadır. ANOVA, normallik ve varyansların homojenliği ihlallerine karşı güçlü olmasına rağmen, bağımsızlık varsayımının ihlaline karşı güçlü değildir. Bu nedenle, analizlere geçilmeden önce verilerin bu varsayımlara uygun olmasını sağlamak önemlidir. Bu varsayımların ihlali durumunda, ham verilerin analizinden elde edilen sonuçlar güvenilir olmayabilir. Verilerin normal dağılımları, aritmetik ortalama, mod, medyan çarpıklık ve basıklık katsayıları gibi istatistikler üzerinden incelenebilmektedir (Kirk, 2008). Dağılımın normal olması için, aritmetik ortalama, mod ve medyan değerlerinin birbirine eşit ya da yakın olması yanında çarpıklık ve basıklık katsayılarının ± 1 değerleri arasında olması gerekmektedir. Ayrıca çarpıklık ve basıklık katsayılarının standart hatalarına bölünmesi ile elde edilen çarpıklık ve basıklık katsayı değerlerinin ± 2 değerleri arasında olması verilerin normal dağılım sergilediğini göstermektedir. (Tabachnick & Fidell, 2013).

Normallik varsayımı histogram ve P-P plot yardımıyla görsel olarak kontrol edilebildiği gibi, görsel olarak elde edilen grafiklerin yanında çarpıklık (skewness) ve basıklık (kurtosis) değerleri yardımıyla sonuçlar desteklenerek yorumlanması önerilmektedir (Stevens, 2009). Ayrıca normallik varsayımı Kolmogorov–Smirnov ve Shapiro–Wilk testleri kullanarak da kontrol edilebilmektedir. Ancak Kolmogorov-Smirnov testinin örneklem büyüklüğünden etkilenmesi yani örneklem büyüklüğü arttıkça p değerinin anlamlı çıkabilme eğiliminde olduğu belirtilmektedir. (Çokluk, Şekercioğlu & Büyüköztürk, 2010; Hair, Anderson,

Tatham & Black,1998). Bu çalışmada basıklık ve çarpıklık değerleri kullanılarak normallik test edilmiştir. Normallik varsayımı sağlanmadığında veride dönüştürme yapılması önerilmektedir (Çokluk vd.,2010). Pozitif çarpıklığın orta düzeyde olması durumunda karekök dönüştürme, yüksek düzeyde olması durumunda logaritmik dönüştürme, negatif çarpıklığın orta düzeyde olması durumunda yansıtma ve karekök dönüştürme; yüksek düzeyde olması durumunda ise yansıtma ve logaritmik dönüştürme yapılması ve sonrasında verilerin normal dağılıp dağılmadığının incelenmesi önerilmektedir (Tabachnick & Fidell, 2013).

ANOVA’da varyans homojenliğini Levene testi ile kontrol edilebilmektedir. Levene testte sıfır hipotezi (H_0) her grubun varyansının eşit olmasıdır. Eğer p değeri 0,05’ten yüksek çıkarsa bu hipotez kabul edilir ve varyans homojenliğinin sağlandığını söyleyebiliriz. Eğer Levene testinden elde edilen p değeri 0,05’ten küçük ise bu hipotezi reddedilir ve varyans homojenliği varsayımının sağlanmadığı ifade edilebilir.

Bir ANOVA deseninin iki veya daha fazla bağımsız değişken içermesi faktöriyel desen, bu durumda gerçekleştirilecek olan ANOVA ise faktöriyel ANOVA veya çok yönlü ANOVA olarak adlandırılmaktadır. Faktöriyel ANOVA’da tek yönlü ANOVA’da olduğu gibi sıfır hipotezinin reddedilmesine F-testi ile karar verilmektedir. Eğer F testinden elde edilen p değeri 0,05’ten küçük çıkarsa sıfır hipotezini reddederek alternatif hipotezi yani söz konusu olan bağımsız değişken gruplarının bağımlı değişken açısından farklılık gösterdiği ifade edilmektedir. Faktöriyel ANOVA ile bağımsız değişkenlerin ana etkilerinin ve etkileşim etkilerinin anlamlı olup olmadığı kontrol edilebilmektedir. İki faktörlü ANOVA için Genel Lineer Model (General Linear Model-GLM) eşitliği aşağıda verilmektedir (Rutherford,2011):

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \quad (17)$$

Eşitlik 17’de, Y_{ijk} i.bireyin puanını yani bağımlı değişkeni; μ parametresi genel ortalamayı, $\alpha_j, j = 1,2, \dots, p$ düzeyden oluşan A faktörünün etkisini, $\beta_k, k = 1,2, \dots, q$ düzeyden oluşan B faktörünün etkisini, $(\alpha\beta)_{jk}, j$ ve k düzeyden oluşan A ve B faktörlerinin etkileşim etkisini, ε_{ijk} ise kontrol edilemeyen hata kaynaklarını göstermektedir. Genel ortalama μ ,

$$\mu = \frac{\sum_{j=1}^p \sum_{k=1}^q \mu_{jk}}{pq} \quad (18)$$

eşitliği ile tanımlanmaktadır. A faktörünün etkisi,

$$\alpha_j = \mu_j - \mu \quad (19)$$

μ_j , A faktörü için marjinal ortalama, μ genel ortalama olarak tanımlanmaktadır. Dolayısıyla A faktörünün etkisi marjinal ortalama ile genel ortalama arasındaki farktır. A faktörü için marjinal ortalama μ_j ,

$$\mu_j = \frac{\sum_{k=1}^q \mu_{jk}}{q} \quad (20)$$

eşitliği ile ifade edilmektedir.

B faktörünün etkisi,

$$\beta_k = \mu_k - \mu \quad (21)$$

μ_k , B faktörü için marjinal ortalama, μ genel ortalama olarak tanımlanmaktadır. Dolayısıyla B faktörünün etkisi marjinal ortalama ile genel ortalama arasındaki farktır. B faktörü için marjinal ortalama μ_j ,

$$\mu_k = \frac{\sum_{j=1}^p \mu_{jk}}{p} \quad (22)$$

eşitliği ile tanımlanmaktadır.

A faktörü ile B faktörünün etkileşim etkileri ise

$$(\alpha\beta)_{jk} = \mu_{jk} - (\mu + \alpha_j + \beta_k) \quad (23)$$

eşitliği ile temsil edilmektedir. İki faktörlü genel lineer model için kestirilen RMSE (veya yanlılık) değerleri aşağıdaki eşitlikte verilmektedir:

$$\hat{Y}_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$$

Hata terimi ε_{ijk} ise gerçek RMSE (yanlılık) değerleri ile gözlenen RMSE (veya yanlılık) değerleri arasındaki farktır ve

$$\hat{\varepsilon}_{ijk} = Y_{ijk} - \hat{Y}_{ijk} \quad (24)$$

eşitliği ile hesaplanmaktadır.

Bu çalışmada Karma MTK modelleri ile kestirilen madde güçlük ve ayırt edicilik parametrelerinin ortalama RMSE ve yanlılık değerlerinin nasıl değiştiğini görmek bağlamında ele alınan faktörlerin ana etkilerini ve etkileşim etkilerini incelemek için altı-yönlü gruplar arası (between-subject) ANOVA analizleri yapılmıştır. Şans parametresi sadece 3 parametrelili modelde yer aldığı için 3 parametrelili model baz alınarak ele alınan

faktörlerin ana etkilerini ve etkileşim etkilerini incelemek içinse beş-yönlü gruplar arası (between-subject) ANOVA analizleri yapılmıştır. Ortalama RMSE değerleri bağımlı değişken, ele alınan faktörler ise bağımsız değişken olarak alınıp ana etkilerin anlamlılığı F testi ile kontrol edilmiştir. Daha sonra da ortalama yanlılık değerleri bağımlı değişken olarak alınıp analizler gerçekleştirilmiştir. Bu analizler sonrasında anlamlı bulunan etkileşimlerin her bir faktörün hangi düzeyinde istatistiksel olarak anlamlı olup olmadığını belirlemek için SPSS syntaxı ile basit etkiler (simple effects) test edilmiştir. Daha sonra anlamlı etkileşimlerin grafikleri verilerek her bir faktörün diğer faktörlerin düzeylerinde ortalama RMSE değerlerine göre sahip olduğu farklılıklar ortaya çıkarılmıştır (Gamst, Meyers, & Guarino, 2008).

İstatistiksel Anlamlılık ve Etki Büyüklüğü

Örneklem büyüklüğü arttıkça p değerinin anlamlılığı da artacağı için istatistiksel anlamlılığı değerlendirmede p değeri yeterli bir indeks olarak kabul edilememektedir. Bu nedenle, p değeri ile etki büyüklüğünün de raporlanması tavsiye edilmektedir. İstatistiksel verilerin analizinin yorumlanmasında, “istatistiksel anlamlılık ölçütü olan p değeri, ne kadar küçük olursa, uygulamanın etkisi ya da gücü de o kadar büyük olur” şeklinde bir kavram yanılgısı vardır. Çünkü istatistiksel anlamlılık testleri, örneklemde elde edilen sonucun şans faktörü ile elde edilme ihtimalini değerlendirir; etki büyüklüğü ise pratik anlamlılığın bir göstergesidir. Etki büyüklüğü ANOVA desenindeki her bir ana etkiyi, ilişkiyi ve hatayı açıklayan varyans oranı olarak tanımlanır ve 0,00 ile 1,00 arasında değer alır. Değişkenler arasında doğrusallık varsayımını gerektirmeyen eta-kare, bağımsız değişkenin bağımlı değişken üzerinde ne derece etkili olduğunu gösterir (Lee,2012). Tek yönlü ANOVA testlerinde etki büyüklüğünü bulmak için gruplar arası kareler toplamı ($SS_{gruplararası}$), toplam kareler toplamına (SS_{toplam}) bölünür. Sonuçlar, her bir etkideki (ilişkideki) varyans yüzdesini ve bu etki tarafından açıklanan hatayı göstermektedir.

$$\eta^2 = \frac{SS_{gruplararası}}{SS_{toplam}} \quad (25)$$

Araştırmacılar etki büyüklüğünü grup ortalamaları farkına göre ve hesaplanan varyansa göre iki kategoride hesaplamaktadırlar. Grup ortalamaları farkına göre hesaplanan etki büyüklüğü ölçümleri Cohen's d, Glass's g ve Hedge's d ile gösterilmektedir. Varyansa göre hesaplanan ilişki gücü ölçümleri ise bağımsız değişken ile bağımlı değişken arasındaki korelasyona göre hesaplanmakta, R-kare (R^2) ölçümlmeleriyle ifade edilmektedir (Kirk, 2008). Cohen'e

(1988) göre, düşük etki büyüklüğü değeri için 0,01; orta etki büyüklüğü değeri için 0,06; yüksek etki büyüklüğü değeri için ve 0,14 değerleri alt sınır değerler olarak önerilmektedir.

Birden fazla kestiricinin varlığında, kısmi etakare (η_p^2), diğer kestiriciler tarafından açıklanan varyans sabit tuttuktan sonra, belirli bir kestirici tarafından açıklanan toplam varyansın oranını ölçer. Çok yönlü (multiple-way) veya faktöriyel ANOVA desenlerinde etkileşim etkilerini belirlemek için kısmi eta-karenin kullanılması önerilmektedir (Richardson, 2011; Norouzian & Plonsky, 2018). Cohen (1988)'de çok yönlü ANOVA desenleri için klasik eta-kare (η^2) formülüne benzer olan kısmi eta-kare (η_p^2) formülünü önermiştir:

$$\eta_p^2 = \frac{SS_{gruplararası}}{SS_{gruplararası} + SS_{hata}} \quad (26)$$

Kısmi etki büyüklüğünü bulmak için gruplar arası kareler toplamı ($SS_{gruplararası}$), gruplar arası kareler toplamı ($SS_{gruplararası}$), ile hatanın toplamı ile toplanarak (SS_{hata}) bölünür.

Tek yönlü ANOVA desenlerinde tek bir etki kaynağı olduğu için, η^2 ve η_p^2 formüllerinin paydasında bir fark oluşmamaktadır. Diğer bir deyişle, tek yönlü ANOVA desenlerinde kısmen çıkarılacak başka etkiler olmadığı için, η^2 ve η_p^2 aynıdır. Bununla birlikte, iki veya çok yönlü desenler için η_p^2 değerleri, η^2 değerlerinden genellikle çok daha büyüktür. Bu durum η_p^2 formülünün diğer faktörleri ve etkileşim etkilerini kısmen ortadan kaldırması nedeniyle oluşmaktadır. Bu nedenle tek yönlü ANOVA desenlerinde η^2 , çok yönlü ANOVA desenlerinde ise η_p^2 değerlerinin yorumlanması önemli görülmektedir (Grissom & Kim, 2012; Pedhazur, 1997). Bu çalışmada öncelikle kestirilen madde parametrelerinden elde edilen ortalama RMSE değerleri bağımlı değişken; madde sayısı, sınıf sayısı, örneklem büyüklüğü, Karma MTK modelleri, kayıp veri türü ve kayıp veri oranı faktörleri de bağımsız değişken olarak alınıp yapılan ANOVA analizleri sonucu ele alınan bu faktörlerin ana etkileri için η^2 değerleri; faktörlerin etkileşim etkileri içinse η_p^2 değerleri Cohen (1988)'in önerdiği değerler doğrultusunda istatistiksel anlamlılık ile birlikte yorumlanmıştır.

Sınıflama Doğruluğu

Sınıflama doğruluğu için üretilen veri setleri içerisinde her bir kişi sınıf olasılık değerlerine göre örtük sınıflardan birine atanmaktadır. Bu veri setleri kullanılarak kestirim yapıldığında elde edilen sınıf üyeliklerinin, gerçek sınıf üyelikleri ile karşılaştırılması gerekmektedir (Sen & Cohen, 2020). Örneğin 1000 kişilik bir veri setinde 980 kişinin kestirilen sınıflama

doğruluğu gerçek sınıfı ile aynı ise o veri setinde sınıflama doğruluğunun %98 olduğu söylenebilmektedir. Bu çalışmada da R programında üretilen veri setleri içerisindeki her bir kişinin ait olduğu örtük sınıflar, Mplus çıktı dosyalarından biri olan CPROB dosyasında yer alan kestirilen örtük sınıf olasılıkları ile karşılaştırılarak hesaplanmıştır.

Gerçek Veri

TIMSS 2015 uygulaması ile matematik ve fen alanlarında ülkeler arası karşılaştırmalar yapılmaktadır. Ancak hem ülkeler eğitimsel, kültürel ve sosyolojik açıdan farklılık göstermekte hem de ülkelerde bulunan alt gruplar demografik açıdan farklı koşullara sahip olabilmektedir. Örtük alt grupların varlığının incelenebilmesi için TIMSS uygulamalarından elde edilen verilerin uygun yapıda olduğu düşünülebilir. Bu nedenle araştırma kapsamında TIMSS 2015 verisi kullanılmıştır. TIMSS 2015 8.sınıf fen alt testi fizik, kimya, biyoloji ve yer bilimleri olmak üzere dört farklı konu alanı ile bilme, uygulama ve akıl yürütme olmak üzere üç farklı bilişsel alandan oluşan bir çerçeveye sahiptir. Öğrenme alanları katılımcı ülkelerin öğretim programları dikkate alınarak belirlenmiştir. Bilme, uygulama ve akıl yürütme bilişsel süreçlerine uygun olarak maddeler geliştirilmiştir. Fen testindeki maddelerin yaklaşık yarısı çoktan seçmeli, diğer yarısı ise uzun/kısa cevaplı maddelerden oluşmaktadır. Fen ve matematik testlerinin yer aldığı TIMSS 2015'te başarı testindeki maddeler, 14'ü fen ve 14'ü matematik olmak üzere toplam 28 bloktan oluşmaktadır. 14 fen bloğun altı bloğu TIMSS 2015 uygulaması için geliştirilen yeni maddelerden oluşmakta iken sekiz blok daha önceki TIMSS uygulamalarında kullanılan maddelerden oluşmaktadır. Kitapçıklardaki çoktan seçmeli madde sayıları 11 ile 18 arasında değişmektedir (Martin vd., 2016). Bu çalışmada simülasyon verisi ile gerçek veri koşullarının örtüşebilmesi bağlamında 600 kişilik örnekleme ulaşabilmek için Türkiye verisinin, fen alt testinin 4.ve 5. kitapçıkları birleştirilerek kitapçıklarda yer alan 14 ortak madde üzerinden analizler gerçekleştirilmiştir.

Verilerin Elde Edilmesi

Araştırmanın gerçek verilerini TIMSS 2015 uygulamasına katılan ülkelerdeki sekizinci sınıf düzeyindeki öğrenciler oluşturmaktadır. TIMSS uygulamasında veriler iki aşamalı tabakalı örnekleme yöntemi kullanılarak elde edilmektedir. Buna göre örneklem seçiminin birinci aşamasında okullar, ikinci aşamasında ise bu okullarda yer alan sınıflar seçilmektedir. İlk olarak tabakalı seçkisiz örnekleme yöntemi kullanılarak ülkelerden okullar seçilmektedir.

İkinci aşamada ise seçilen okullarda yer alan bütün sınıfların eşit seçilme şanslarının olduğu basit seçkisiz örnekleme yöntemi kullanılarak öğrenciler seçilmektedir (LaRoche, Joncas & Foy, 2015).

Gerçek Veri Analizi

TIMSS 2015 8.sınıf Türkiye fen alt testinde yer alan çoktan seçmeli iki kategorili puanlanan maddelerin analizi Mplus programı ile gerçekleştirilmiştir.

Model Veri Uyumunun Belirlenmesi

Model seçimi, Karma MTK modelleri ve diğer örtük değişken modellerinde önemli bir konudur. MTK kapsamında, model belirleme için olasılık oranı testleri veya bilgi kriteri indeksleri kullanılmaktadır. Karma MTK modelleri, alternatif modeller arasından en iyi modelin seçildiği bir yaklaşım izler. Yaygın olarak kullanılan üç bilgi ölçütü indeksi: Akaike'nin (1974) bilgi kriterleri (AIC), Schwarz'ın (1978) Bayes bilgisi kriteri (BIC) ve BIC'in örneklem büyüklüğü ayarlı versiyonu (SABIC; Sclove, 1987). AIC, BIC ve SABIC tanımları aşağıda verilmiştir:

$$AIC = -2\text{Log}L + 2p$$

$$BIC = -2\text{Log}L + p * \ln(N)$$

$$SABIC = -2\log L + p * \ln(N + 24)/2)$$

L, olabirlik fonksiyonudur, p kestirilecek parametrelerin sayısını ve N örneklem büyüklüğünü ifade eder. Uyum indeksleri için parametre sayısı, tüm grupların madde parametrelerinin toplamı, grup ortalama tahminlerinin sayısı ve Karma olasılık tahminlerinin sayısı (Li vd., 2009) eklenerek hesaplanır. Bu indekslerin daha küçük değerleri daha iyi uyum göstermektedir. Görüldüğü gibi, her indeks, olasılık fonksiyonuna farklı düzeltmeler (penalty) uygulamaktadır. Böylece, bu indeksler birlikte kullanıldığında aynı veriler için farklı çözümler elde edilebilir. Bu uyum indekslerinde bir takım farklı özellikler mevcuttur. Örneğin, AIC'nin tutarsızlıktan dolayı sınıf sayısını olduğundan daha fazla kestirmeye eğilimli olduğu bilinmesine rağmen (McLachlan & Peel, 2000) etkin olarak kabul edilir (Vrieze, 2012). Görüldüğü gibi, BIC ve SABIC, olasılık fonksiyonuna daha fazla düzeltme uygulamaktadır. Böylece, BIC ve SABIC, AIC'den daha basit modeller seçmeye eğilimlidir. BIC ve SABIC denklemlerinde örneklem büyüklüğü terimi, örneklem büyüklüğü arttıkça istatistiksel olarak bir anlam ifade etmektedir (Vrieze, 2012). Buna asimptotik tutarlılık

denir. Başka bir ifade ile örneklem büyüklüğü arttıkça, BIC tutarlı bir şekilde doğru sınıfları seçmeye eğilimlidir (Vrieze, 2012). Dolayısıyla, bu indekslerin faydası, gerçek modelin yapısına bağlı olabilir. BIC ve SABIC'in gerçek modelin basit bir yapıya sahip olması durumunda daha iyi performans göstermesi beklenirken, AIC karmaşık yapıya sahip gerçek modeller için daha iyi çalışır. Bu indekslerin performansları, tahmini parametrelerin sayısı, bireylerin sayısı, tahmin teknikleri, model tipi ve karmaşıklık gibi çeşitli faktörlere bağlıdır.

BÖLÜM 4

BULGULAR

Bu bölümde çalışmanın araştırma problemleri doğrultusunda elde edilen bulgular tablo ve grafikler yardımıyla sunulmuştur.

Örneklem büyüklüğü, madde sayısı, sınıf sayısı, Karma MTK modelleri, kayıp veri oranı ve kayıp veri türü değişkenleri ele alınarak üretilen verilerin Karma MTK modelleri ile analiz edilmesi sonucu elde edilen ortalama RMSE ve yanlılık değerleri verilmiştir. Elde edilen bulgular madde güçlüğüne ait bulgular, madde ayırt ediciliğine ait bulgular ve şans parametresine ait bulgular olmak üzere sırasıyla sunulmuştur. Madde parametre kestirimlerine ait elde edilen ortalama RMSE ve yanlılık değerlerine göre seçilen faktörlerin ana etkileri ve etkileşim etkileri gruplar arası ANOVA tekniği kullanılarak incelenmiştir. Sonrasında ise ele alınan değişkenlerin kombinasyonundan elde edilen sınıflama doğruluğu yüzdelere ait bulgulara yer verilmiştir. En son olarak ise gerçek veri seti üzerinden Karma MTK modellerinin incelenmesine ait bulgular sunulmuştur.

Madde Güçlük Parametrelerine Ait Bulgular

Bu kısımda öncelikli olarak araştırma kapsamında incelenen simülasyon koşullarına (kayıp veri türü, madde sayısı, örneklem büyüklüğü, sınıf sayısı, kayıp veri yüzdesi, kullanılan Karma MTK modeli) göre madde güçlük parametresi için hesaplanan ortalama RMSE ve yanlılık değerleri sunulmuştur. Daha sonra madde güçlük parametresi için hesaplanan ortalama RMSE ve yanlılık değerlerinin simülasyon koşullarında nasıl farklılaştığını incelemek amacıyla gruplar arası ANOVA analizleri gerçekleştirilmiştir.

Madde güçlük parametresi, bu çalışmada dikkate alınan her üç Karma MTK modeli için ortak madde parametresidir. Karma Rasch modeli için kestirilen madde güçlük parametreleri

değerlerinin ortalama RMSE ve yanlılık değerleri Tablo 3'te sunulmaktadır. Simülasyon koşulları için bu tablodaki gösterimler, belirli bir durum için faktörlerin kombinasyonunu temsil etmek üzere tasarlanmıştır. Simülasyon koşullarını belirtmek için 10-14 haneden oluşan bir kodlama yapılmıştır ve bu kodlamada ilk karakter modeli (Rasch, 2PL, 3PL), ikinci karakter sınıf sayısını (2, 3), üçüncü karakter kayıp veri yüzdesini (10P, 20P), dördüncü karakter örneklem büyüklüğünü (600, 1000), beşinci karakter ise madde sayısını (10,30) göstermektedir. Örneğin R210P60010 gösteriminde Rasch model (R), sınıf sayısı (2), kayıp veri yüzdesi (10P), örneklem büyüklüğü (600), madde sayısı (10) ile ifade edilmektedir.

Tablo 3

Karma Rasch Modeli İçin Madde Güçlük Parametre Değerlerinin Ortalama RMSE ve Yanlılık Değerleri

Koşullar	TAM		Rastgele Kayıp Veri (RKV)		Rastgele Olmayan Kayıp Veri (ROKV)	
	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE
R210P60010	-0,0003	0,045	0,0047	0,052	-0,0100	0,075
R210P60030	0,0002	0,025	0,0006	0,026	-0,0050	0,054
R210P100010	0,0003	0,035	0,0048	0,043	-0,0100	0,067
R210P100030	-0,0000	0,021	0,0001	0,022	-0,0050	0,050
R220P60010	0,0007	0,046	0,0138	0,083	-0,0140	0,082
R220P60030	-0,0002	0,025	0,0012	0,028	-0,0260	0,251
R220P100010	-0,0010	0,035	0,0120	0,071	-0,0340	0,540
R220P100030	-0,0001	0,023	0,0013	0,024	-0,0200	0,188
R310P60010	0,0247	0,139	0,0017	1,160	-0,0120	0,236
R310P60030	0,0002	0,105	0,0049	0,435	-0,0130	0,170
R310P100010	-0,0008	0,102	0,0029	0,206	-0,0080	0,125
R310P100030	0,0001	0,082	0,0008	0,069	-0,0020	0,036
R320P60010	-0,0007	0,148	0,0321	1,350	-0,0510	1,690
R320P60030	-0,0007	0,081	0,0083	0,397	-0,0100	0,191
R320P100010	-0,0007	0,090	-0,0001	0,884	-0,0220	0,383
R320P100030	0,0000	0,024	0,0089	0,058	-0,0080	0,088

Tablo 3 incelendiğinde Karma Rasch modeli için elde edilen madde güçlük parametrelerinin ortalama RMSE değerlerinin madde sayısı arttıkça azaldığı, sınıf sayısı arttıkça ise arttığı görülmektedir. Karma Rasch modeli için elde edilen ortalama yanlılık değerlerinin ise 0,05'ten düşük olduğu görülmektedir. Tam veri koşulunda madde sayısı ve örneklem büyüklüğü arttıkça RMSE değerlerinin düştüğü, sınıf sayısı ve kayıp veri yüzdesi arttıkça ise RMSE değerlerinin arttığı görülmüştür. RKV ve ROKV koşullarında ise madde sayısı ve örneklem büyüklüğü arttıkça genellikle RMSE değerlerinin azaldığı, sınıf sayısı ve kayıp

veri yüzdesi arttıkça RMSE değerlerinin genellikle arttığı gözlenmiştir. Madde güçlük parametre değerlerinin en yüksek RMSE değerlerinin tam, RKV ve ROKV durumlarının 3 sınıf, %20 kayıp veri yüzdesi, 600 örneklem büyüklüğü ve 10 madde koşulunda sahip olduğu gözlenmiştir. En düşük RMSE değeri ise tam veri, 2 sınıf, %10 kayıp veri yüzdesi, 1000 örneklem büyüklüğü ve 30 madde koşulunda gözlenmiştir. Tablo 4 ise Karma 2PL modeli için kestirilen madde güçlük parametrelerinin ortalama RMSE ve yanlılık değerlerini özetlemektedir.

Tablo 4

Karma 2PL Modeli İçin Madde Güçlük Parametre Değerlerinin Ortalama RMSE ve Yanlılık Değerleri

Koşullar	TAM		Rastgele Kayıp Veri (RKV)		Rastgele Olmayan Kayıp Veri (ROKV)	
	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE
2PL210P60010	0,0005	0,041	0,0030	0,065	-0,0120	0,089
2PL210P60030	0,0002	0,025	0,0010	0,039	-0,0040	0,077
2PL210P100010	0,0004	0,032	0,0030	0,038	-0,0120	0,070
2PL210P100030	0,0001	0,021	0,0010	0,022	-0,0050	0,052
2PL220P60010	0,0012	0,057	0,0120	0,121	-0,0390	0,241
2PL220P60030	0,0005	0,037	0,0020	0,061	-0,0410	0,070
2PL220P100010	-0,0002	0,034	0,0140	0,233	-0,0360	0,177
2PL220P100030	-0,0005	0,023	0,0010	0,024	-0,0320	0,919
2PL310P60010	0,0460	1,551	0,0680	1,775	0,0821	1,950
2PL310P60030	0,0018	0,070	0,0080	0,032	-0,0017	0,064
2PL310P100010	0,0387	0,973	0,0390	1,256	0,0776	1,778
2PL310P100030	0,0018	0,051	0,0030	0,096	-0,002	0,075
2PL320P60010	0,0600	1,761	0,0720	1,994	-0,063	2,570
2PL320P60030	0,0042	0,279	0,0210	1,436	0,0120	1,709
2PL320P100010	0,0499	1,641	0,0800	1,832	0,0690	2,237
2PL320P100030	0,0029	0,226	0,0080	0,439	-0,0076	0,113

Tablo 4 incelendiğinde Karma 2PL modeli için elde edilen madde güçlük parametrelerinin ortalama RMSE değerlerinin Karma Rasch model için elde edilen madde güçlük parametrelerinin ortalama RMSE değerlerinden daha yüksek olduğu görülmektedir. Madde sayısı ve örneklem büyüklüğü arttıkça RMSE değerlerinin azaldığı, sınıf sayısı ve kayıp veri yüzdesi arttıkça ise RMSE değerlerinin arttığı görülmüştür. Kayıp veri türlerine göre ortalama RMSE değerleri incelendiğinde ise Rastgele Olmayan Kayıp Veri (ROKV) koşulu için daha yüksek RMSE değerleri elde edilmiştir. Karma 2PL model için madde güçlük parametre değerlerinin en yüksek RMSE değerlerine tam, RKV ve ROKV koşullarının 3 sınıf, %20 kayıp veri yüzdesi, 600 örneklem büyüklüğü ve 10 madde durumlarında elde

edildiği gözlenmiştir. En düşük RMSE değeri ise tam veri, 2 sınıf, %10 kayıp veri yüzdesi, 1000 örneklem büyüklüğü ve 30 madde koşulunda gözlenmiştir. Karma 3PL modeli için kestirilen madde güçlük parametrelerinin ortalama RMSE ve yanlılık değerleri ise Tablo 5’te verilmektedir:

Tablo 5

Karma 3PL Modeli İçin Madde Güçlük Parametre Değerlerinin Ortalama RMSE ve Yanlılık Değerleri

Koşullar	TAM		Rastgele Kayıp Veri (RKV)		Rastgele Olmayan Kayıp Veri (ROKV)	
	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE
3PL210P60010	-0,0320	0,367	-0,0460	0,560	-0,0320	0,373
3PL210P60030	-0,0060	0,112	-0,0140	0,225	-0,0260	0,303
3PL210P100010	-0,0170	0,219	-0,0160	0,265	-0,0180	0,263
3PL210P100030	-0,0050	0,090	-0,0110	0,168	-0,0320	0,352
3PL220P60010	-0,0250	0,348	-0,0450	0,654	-0,0650	0,592
3PL220P60030	-0,0070	0,130	-0,0130	0,213	-0,0200	0,222
3PL220P100010	-0,0160	0,219	-0,0480	0,586	-0,0610	0,520
3PL220P100030	-0,0040	0,078	-0,0080	0,124	-0,0210	0,217
3PL310P60010	-0,0140	1,386	0,0339	1,898	0,0100	2,831
3PL310P60030	-0,0110	0,210	-0,0110	0,221	-0,0130	0,257
3PL310P100010	-0,0120	1,075	-0,0060	1,704	0,0030	2,636
3PL310P100030	-0,0070	0,160	-0,0080	0,196	-0,0130	0,195
3PL320P60010	-0,0030	1,160	0,0324	4,471	0,0308	2,357
3PL320P60030	-0,0100	0,246	-0,0180	0,292	-0,0170	0,283
3PL320P100010	-0,0430	2,652	-0,1140	2,673	0,0150	2,341
3PL320P100030	-0,0060	0,141	-0,0090	0,253	-0,0130	0,212

Tablo 5 incelendiğinde Karma 3PL modeli için elde edilen madde güçlük parametrelerinin ortalama RMSE değerlerinin Karma 2PL model için elde edilen madde güçlük parametrelerinin ortalama RMSE değerlerinden daha yüksek olduğu görülmektedir. Modelin karmaşıklığı arttıkça ortalama RMSE değerlerinin arttığı görülmektedir. Madde sayısı ve örneklem büyüklüğü arttıkça RMSE değerlerinin azaldığı, sınıf sayısı ve kayıp veri yüzdesi arttıkça ise RMSE değerlerinin arttığı görülmüştür. Kayıp veri türlerine göre ortalama RMSE değerleri incelendiğinde ise RKV koşulu için daha yüksek RMSE değerleri elde edilmiştir. Karma 3PL model için madde güçlük parametre değerlerinin en yüksek RMSE değerinin RKV, 3 sınıf, %20 kayıp veri yüzdesi, 600 örneklem büyüklüğü ve 10 madde koşulu olduğu gözlenmiştir. En düşük RMSE değeri ise tam veri, 2 sınıf, %10 kayıp veri yüzdesi, 1000 örneklem büyüklüğü ve 30 madde koşulunda gözlenmiştir.

Madde Güçlük Parametrelerinden Elde Edilen Ortalama RMSE Değerleri İçin ANOVA Sonuçları

Normallik ve varyansların homojenliği varsayımları ANOVA analizleri yapılmadan önce test edilmiştir. Normallik testi ile elde edilen çarpıklık katsayısı 2,123 ve standart hatası 0,202 ve basıklık katsayısı ise 4,77 ve standart hatası 0,401 olarak elde edilmiştir. RMSE değerleri pozitif çarpık olduğu için verilerin doğal logaritması ($\log[\text{RMSE}]$) alınması sonucu, çarpıklık katsayısı 0,27 ve standart hatası 0,202 ve basıklık katsayısı ise -0,927 ve standart hatası 0,401 olarak elde edilerek verilerin normal dağılım sergilemesi sağlanmıştır. Levene testi ile p değeri 0,05 ten büyük olduğu için ($p=0,162$) varyansların homojenliği varsayımı sağlanmıştır. Madde güçlük değerleri için $\log [\text{RMSE}]$ değerleri alınarak gerçekleştirilen ANOVA sonuçları Tablo 6'da sunulmaktadır:

Tablo 6

Madde Güçlük Parametrelerinden Elde edilen log[RMSE] değerleri için ANOVA Sonuçları

Kaynak	η^2	η_p^2	sd	Kareler Toplamı	Ortalama kare	F	* $p<0,05$
Tür	0,070	0,832	2	3,881	1,940	54,318*	0,000
Madde	0,167	0,922	1	9,313	9,313	260,685*	0,000
Örneklem	0,017	0,540	1	0,922	0,922	25,798*	0,000
Sınıf	0,276	0,951	1	15,340	15,340	429,413*	0,000
Yüzde	0,033	0,702	1	1,854	1,854	51,902*	0,000
Model	0,164	0,921	2	9,126	4,563	127,724*	0,000
Tür*madde	0,001	0,085	2	0,073	0,037	1,024	0,376
Tür*örneklem	0,001	0,039	2	0,032	0,016	0,444	0,647
Tür*sınıf	0,028	0,663	2	1,543	0,771	21,593*	0,000
Tür*yüzde	0,010	0,426	2	0,582	0,291	8,151*	0,002
Tür*model	0,007	0,324	4	0,377	0,094	2,635	0,062
Madde*örneklem	0,002	0,109	1	0,096	0,096	2,690	0,115
Madde*sınıf	0,046	0,767	1	2,581	2,581	72,259*	0,000
Madde*yüzde	0,000	0,005	1	0,004	0,004	0,106	0,747
Madde*model	0,014	0,489	2	0,752	0,376	10,531*	0,001
Örneklem*sınıf	0,003	0,173	1	0,164	0,164	4,591*	0,043
Örneklem*yüzde	0,000	0,016	1	0,013	0,013	0,360	0,555
Örneklem*model	0,001	0,077	2	0,066	0,033	0,920	0,413
Sınıf*yüzde	0,000	0,007	1	0,006	0,006	0,165	0,689
Sınıf*model	0,031	0,687	2	1,723	0,862	24,119*	0,000
Yüzde*model	0,012	0,459	2	0,667	0,333	9,334*	0,001
Tür*madde*sınıf	0,011	0,440	2	0,617	0,309	8,637*	0,002
Tür*sınıf*model	0,017	0,548	4	0,954	0,239	6,676*	0,001
Madde*sınıf*model	0,011	0,439	2	0,615	0,307	8,605*	0,002
Örneklem*sınıf*model	0,005	0,268	2	0,288	0,144	4,025*	0,032
Hata	0,014		2	0,786	0,036		
			2				

Tablo 6'daki tür, kayıp veri türünü; madde, madde sayısını; örneklem, örneklem büyüklüğünü; sınıf, sınıf sayısını; yüzde, kayıp verilerin toplam yüzdesini; model, kullanılan Karma MTK modeli faktörlerini temsil etmektedir. Kayıp veri türü faktörünün ana etkisinin $F(2,22)=54,318$; $p<0,05$ ve $\eta^2=0,070$ değerlerine dayalı olarak anlamlı ve orta düzey bir etkiye sahip olduğu söylenebilir. Madde faktörünün $F(2,22)=260,685$; $p<0,05$ ve $\eta^2=0,167$; sınıf faktörünün $F(2,22)=429,413$; $p<0,05$ ve $\eta^2=0,276$; model faktörünün $F(2,22)=127,724$; $p<0,05$ ve $\eta^2=0,164$ ana etkilerinin anlamlı ve yüksek etki büyüklüğü değerlerine sahip olduğu görülmektedir. Örneklem faktörünün ana etkisinin $F(2,22)=25,798$; $p<0,05$ ve $\eta^2=0,017$ değerleri ile anlamlı ve düşük düzey bir etkiye sahip olduğu söylenebilir. Kayıp veri yüzdesi faktörünün ana etkisinin $F(2,22)=51,902$; $p<0,05$ ve $\eta^2=$

0,033 değerleri ile anlamlı ve orta düzey bir etkiye sahip olduğu söylenebilir. İkili etkileşimler incelendiğinde ise kayıp veri türü faktörü ile sınıf faktörünün $F(2,22)= 21,593$; $p<0,05$ ve $\eta_p^2= 0,663$; kayıp veri türü faktörü ile kayıp veri yüzdesi faktörünün $F(2,22)= 8,151$; $p<0,05$ ve $\eta_p^2= 0,426$ değerlerine dayalı olarak etkileşim etkilerinin anlamlı ve yüksek etki büyüklüğü değerlerine sahip olduğu söylenebilir.

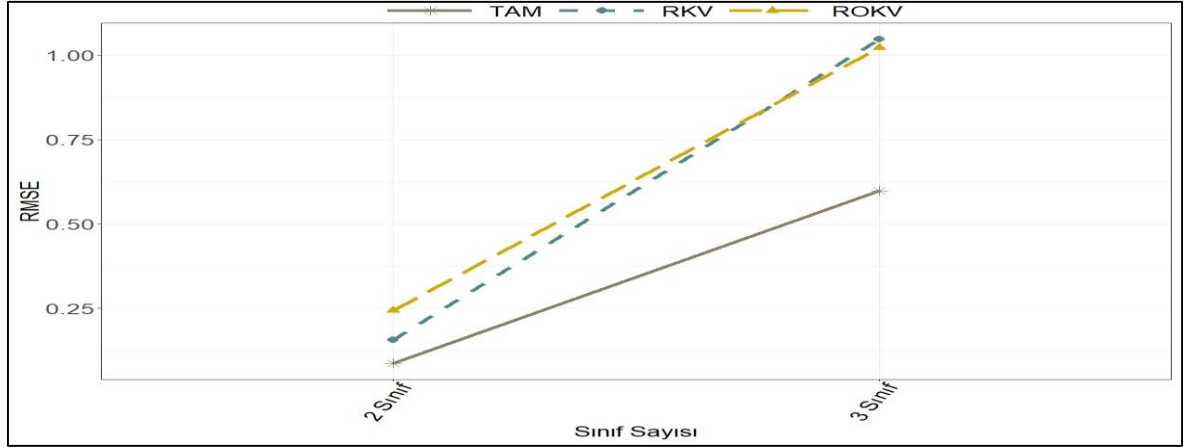
Kayıp veri türü ile sınıf faktörlerinin ikili etkileşimlerinin anlamlı bulunması sonucu sınıf faktörünün her bir düzeyinin kayıp veri türü faktörünün hangi düzeyinde istatistiksel olarak anlamlı olup olmadığını belirlemek için SPSS sözdizimi ile basit etkiler (simple effects) test edilmiştir. Basit etkilere ait Tablo 7 aşağıda verilmektedir:

Tablo 7

Sınıf Faktörünün Her Bir Koşulunda Kayıp Veri Türlerine Ait Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
2 sınıf					
Kayıp veri türü	3,547	2	1,774	49,28*	0,0007
Hata	0,786	22	0,036		
3 sınıf					
Kayıp veri türü	1,848	2	0,924	25,67*	0,0001
Hata	0,786	22	0,036		

Tablo 7 incelendiğinde sınıf faktörünün 2 sınıf koşulu ve kayıp veri türü faktörü ile etkileşimleri $F(2,22)=49,28$; $p<0,05$ ve sınıf faktörünün 3 sınıf koşulu ve kayıp veri türü faktörünün etkileşimleri $F(2,22)=25,67$; $p<0,05$ arasında anlamlı farklılık olduğu görülmektedir. Etkileşimlerin faktör düzeylerindeki ortalama farklılıklarını (mean difference) görebilmek için grafiklerden yardım alınmıştır. Kayıp veri türü ve sınıf faktörlerinin etkileşimlerinin ortalama RMSE değerlerine göre incelendiği grafik Şekil 2’de verilmektedir:



Şekil 2. Kayıp veri türü ve sınıf faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Kayıp veri türü ve sınıf faktörlerinin etkileşimleri faktörlerin her bir düzeyi için incelendiğinde tam veri ve 2 sınıf faktörlerinin etkileşimi ve tam veri 3 sınıf faktörlerinin etkileşimleri için daha düşük RMSE değeri elde edilirken, 2 sınıf ve 3 sınıf koşullarının RKV ve ROKV etkileşimleri için daha yüksek RMSE değerleri elde edilmiştir. 2 sınıf koşulunda RKV etkileşimi için daha düşük RMSE değeri gözlenirken, 3 sınıf koşulunda ise ROKV etkileşimi için daha düşük RMSE değerleri gözlenmiştir. RMSE değerinin 3 sınıf ve RKV etkileşimi için en yüksek değerine ulaştığı tespit edilmiştir. Ancak 2 sınıf ve RKV etkileşiminde elde edilen RMSE değeri 2 sınıf ve ROKV etkileşimi için daha düşüktür.

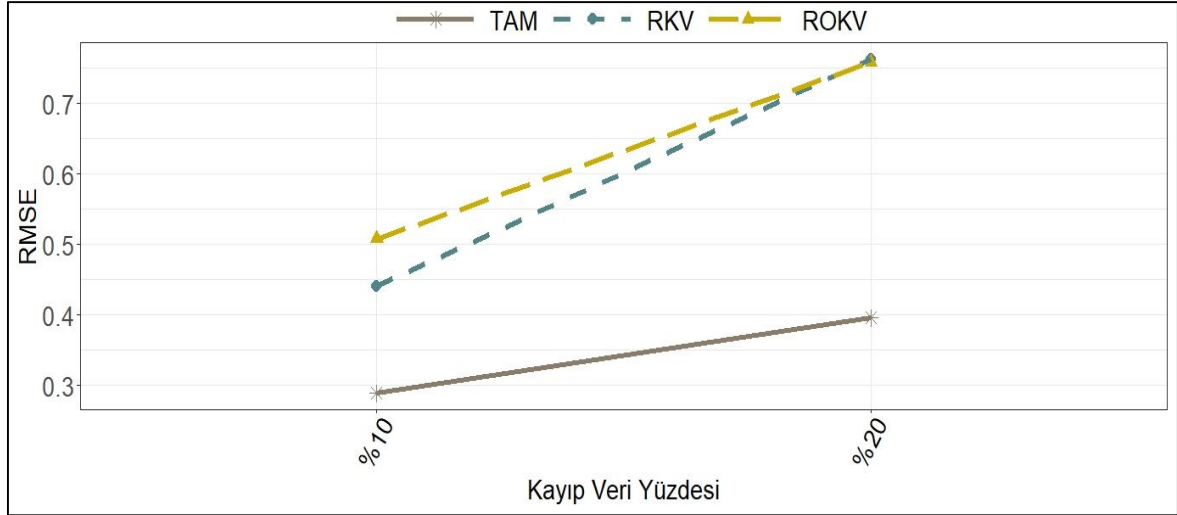
Kayıp veri türü ile kayıp veri yüzdesinin ikili etkileşimlerinin anlamlı bulunması sonucu kayıp veri yüzdesi faktörünün her bir düzeyinin kayıp veri türünün faktörün hangi düzeyinde istatistiksel olarak anlamlı olup olmadığını belirlemek için basit etkiler test edilmiştir. Kayıp veri yüzdesi faktörünün koşullarının kayıp veri faktörünü ile etkileşiminin anlamlılığının incelendiği basit etkilere ait Tablo 8 aşağıda verilmektedir:

Tablo 8

Kayıp Veri Yüzdesi Faktörünün Her Bir Koşulunda Kayıp Veri Türlerinin Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
%10 yüzde					
Kayıp veri türü	0,744	2	0,372	10,33*	0,0007
Hata	0,786	22	0,036		
%20 yüzde					
Kayıp veri türü	3,703	2	1,851	51,41*	0,000
Hata	0,786	22	0,036		

Tablo 8 incelendiğinde %10 kayıp veri yüzdesi ve kayıp veri türü faktörünün etkileşimlerinin $F(2,22) = 10,33$; $p < 0,05$ ve %20 kayıp veri yüzdesi ve kayıp veri türü faktörünün etkileşimlerinin $F(2,22) = 51,41$; $p < 0,05$ anlamlı olduğu görülmektedir. Etkileşimlerin faktör düzeylerinde sahip olduğu ortalama farklılıkları görebilmek amacıyla elde edilen grafik Şekil 5'te verilmektedir:



Şekil 3. Kayıp veri türü ile kayıp veri yüzdesi faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Şekil 3 incelendiğinde en düşük RMSE değerine tam veri ve %10 kayıp veri yüzdesi etkileşimlerinde ulaşılırken, en yüksek RMSE değerine ise RKV ve ROKV koşullarının %20 kayıp veri etkileşimlerinde ulaşılmıştır. RKV ve %10 kayıp veri yüzdesi etkileşimi için ROKV ve %10 kayıp veri yüzdesi etkileşimine göre daha düşük RMSE değeri elde edilmiştir.

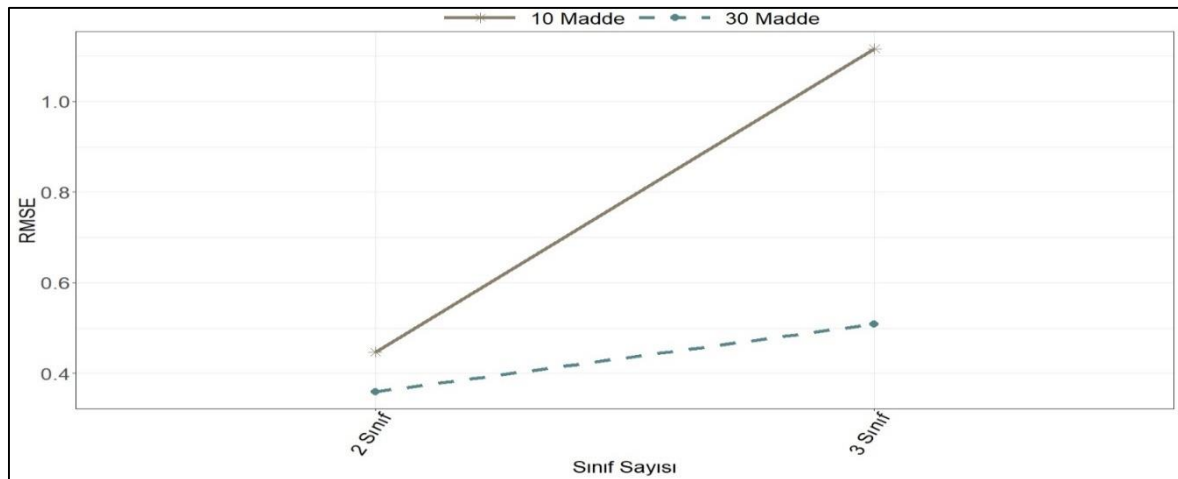
Faktörlerin ikili etkileşimleri incelendiğinde madde faktörü ile sınıf faktörünün $F(2,22) = 72,259$; $p < 0,05$ ve $\eta_p^2 = 0,767$; madde faktörü ile model faktörünün $F(2,22) = 8,151$; $p < 0,05$ ve $\eta_p^2 = 0,489$ etkileşimlerinin anlamlı ve yüksek etki büyüklüğü değerlerine sahip olduğu söylenebilir. Madde ile sınıf faktörlerinin ikili etkileşimlerinin anlamlı bulunması sonucu sınıf faktörünün her bir düzeyinin madde faktörünün düzeyinde istatistiksel olarak anlamlılığının incelendiği basit etkilere ait Tablo 9 aşağıda verilmektedir:

Tablo 9

Madde Faktörünün Her Bir Koşulunda Sınıf Faktörünün Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
2 sınıf					
Madde sayısı	1,045	1	1,045	29,02*	0,000
Hata	0,786	22	0,036		
3 sınıf					
Madde sayısı	10,875	1	10,875	302,083*	0,000
Hata	0,786	22	0,036		

Tablo 9 incelendiğinde sınıf faktörünün 2 sınıf koşulunda madde faktörü ile olan etkileşimlerin $F(1,22) = 29,02$; $p < 0,05$ ve sınıf faktörünün 3 sınıf koşulunda madde faktörü ile olan etkileşimlerin $F(1,22) = 302,083$; $p < 0,05$ anlamlı olduğu sonucuna ulaşılmıştır. Etkileşimler arasındaki ortalama farklılıkları faktör düzeylerinde etkileşimlerinin incelendiği Şekil 4 ise aşağıda verilmektedir:



Şekil 4. Madde ve sınıf faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Şekil 4'teki grafikte görüldüğü gibi 2 sınıf koşulunun 10 ve 30 madde etkileşimlerinde madde güçlük parametre kestirimi için hesaplanan ortalama RMSE değerlerinin birbirine yakın olduğu, 3 sınıf koşulunda ise 10 madde etkileşimi için elde edilen RMSE değerinin 30 madde etkileşimi için elde edilen RMSE değerine göre belirgin bir şekilde yüksek olduğu görülmektedir. Madde sayısı arttıkça RMSE değerleri azalırken, sınıf sayısı arttıkça RMSE değerlerinin arttığı görülmektedir.

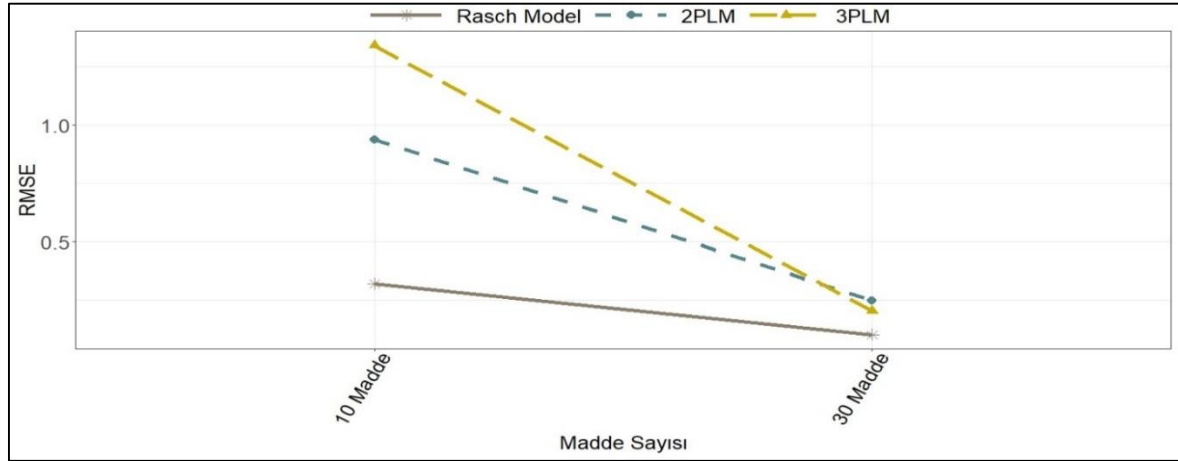
Madde faktörü ile model faktörünün $F(2,22)= 8,151$; $p<0,05$ ve $\eta_p^2= 0,489$ etkileşimleri anlamlı bulunmuştur. Madde faktörünün koşulları ile model faktörünün etkileşiminin anlamlılığının incelendiği basit etkilere ait Tablo 10 aşağıda verilmektedir:

Tablo 10

Madde Faktörünün Her Bir Koşulunda Modellerin Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
10 Madde					
Modeller	7,525	2	3,762	104,5*	0,000
Hata	0,786	22	0,036		
30 Madde					
Modeller	2,350	2	1,175	32,63*	0,000
Hata	0,786	22	0,036		

Tablo 10 incelendiğinde madde faktörünün 10 madde koşulu ile model faktörünün $F(2,22)= 104,5$; $p<0,05$ etkileşimlerinin ve madde faktörünün 30 madde koşulu ile model faktörünün $F(2,22)=32,63$; $p<0,05$ etkileşimlerinin anlamlı olduğu sonucuna ulaşılmıştır. Etkileşimlerin faktör düzeylerinde sahip olduğu ortalama farklılıkları görebilmek amacıyla elde edilen grafik Şekil 5’te verilmektedir:



Şekil 5. Madde ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Şekil 5 incelendiğinde 10 madde ve Rasch model etkileşimi için elde edilen RMSE değerinin, 2PL ve 3PL model etkileşimleri için hesaplanan ortalama RMSE değerlerine göre daha düşük olduğu görülmektedir. 10 madde ve 2PL etkileşimlerinde elde edilen ortalama RMSE değerinin 10 madde ve 3PL etkileşimlerinde elde edilen ortalama RMSE değerine göre daha düşük olduğu görülmektedir. 30 madde koşulunda ise Rasch model ile olan

etkileşimlerde elde edilen ortalama RMSE değerlerine göre daha düşük olduğu tespit edilmiştir. 30 madde ve 3PL etkileşimlerinde elde edilen ortalama RMSE değerinin 30 madde ve 2PL etkileşimlerinde elde edilen ortalama RMSE değerine göre daha düşük olduğu görülmektedir. Şekilde modellere ait eğrilerin eğimi dikkate alındığında 10 maddeli durumdan 30 maddeli duruma geçerken madde güçlük parametre kestirimi için hesaplanan ortalama RMSE değerlerinde en fazla azalışı 3PL modelin daha sonra ise 2PL modelin gösterdiği görülmekte iken Rasch model için elde edilen RMSE değerlerinin ise daha düşük bir eğimle azalma gösterdiği söylenebilir. Dolayısıyla en düşük RMSE değerine sahip modelin 30 maddeli Karma Rasch model olduğu görülebilmektedir.

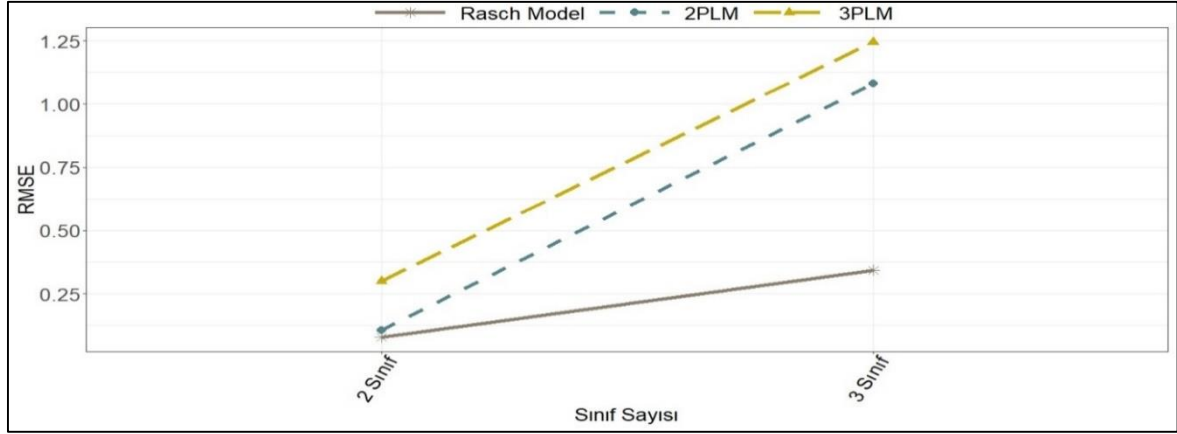
İkili etkileşimler incelendiğinde örneklem faktörü ile sınıf faktörünün $F(2,22)= 4,591$; $p<0,05$ ve $\eta_p^2= 0,173$ değerlerine göre anlamlı ve yüksek etki büyüklüğü değerine sahip olduğu gözlenmiştir. Sınıf faktörü ile model faktörünün $F(2,22)=24,119$; $p<0,05$ ve $\eta_p^2= 0,687$; kayıp veri yüzdesi ile model faktörünün $F(2,22)=9,334$; $p<0,05$ ve $\eta_p^2= 0,687$ değerlerine göre etkileşim etkilerinin anlamlı ve yüksek etki büyüklüğü değerlerine sahip olduğu söylenebilir. Sınıf faktörünün koşulları ile model faktörünün etkileşiminin anlamlılığının incelendiği basit etkilere ait Tablo 11 aşağıda verilmektedir:

Tablo 11

Sınıf Faktörünün Her Bir Koşulunda Modellerin Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
2 Sınıf					
Modeller	6,732	2	3,366	93,5*	0,000
Hata	0,786	22	0,036		
3 Sınıf					
Modeller	4,111	2	2,056	57,11*	0,000
Hata	0,786	22	0,036		

Tablo 11 incelendiğinde sınıf faktörünün 2 sınıf koşulu ile model faktörünün $F(2,22)= 93,5$; $p<0,05$ ve sınıf faktörünün 3 sınıf koşulu ile model faktörünün $F(2,22)=57,11$; $p<0,05$ etkileşimlerinin anlamlı olduğu sonucuna ulaşılmıştır. Her bir faktör düzeyinde etkileşimlere ait ortalama farklılıkları görebilmek amacıyla elde edilen RMSE değerlerine ait grafik Şekil 6'da verilmiştir:



Şekil 6.Sınıf ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Şekil 6 incelendiğinde 2 sınıf koşulunda Rasch model ve Karma 2PL model etkileşimlerinde madde güçlüğü parametreleri için hesaplanan ortalama RMSE değerlerinin birbirine oldukça yakın, Karma 3PL model etkileşimi için hesaplanan ortalama RMSE değerine göre belirgin bir şekilde düşük olduğu görülmektedir. 3 sınıf koşulunda ise 2PL ve 3PL model etkileşimlerinde, hesaplanan ortalama RMSE değerlerinin birbirine oldukça yakın olduğu, 3 sınıf ve Rasch model etkileşimi için bulunan ortalama RMSE değerine göre belirgin bir şekilde yüksek olduğu görülmektedir. Şekilde modellere ait eğrilerin eğimi dikkate alındığında 2 sınıf ve 3 sınıf etkileşimlerinde hesaplanan ortalama RMSE değerlerinde en yüksek artışı 2PL ve 3PL modelin gösterdiği, Rasch modelin ise daha düşük bir eğimle artış gösterdiği söylenebilir.

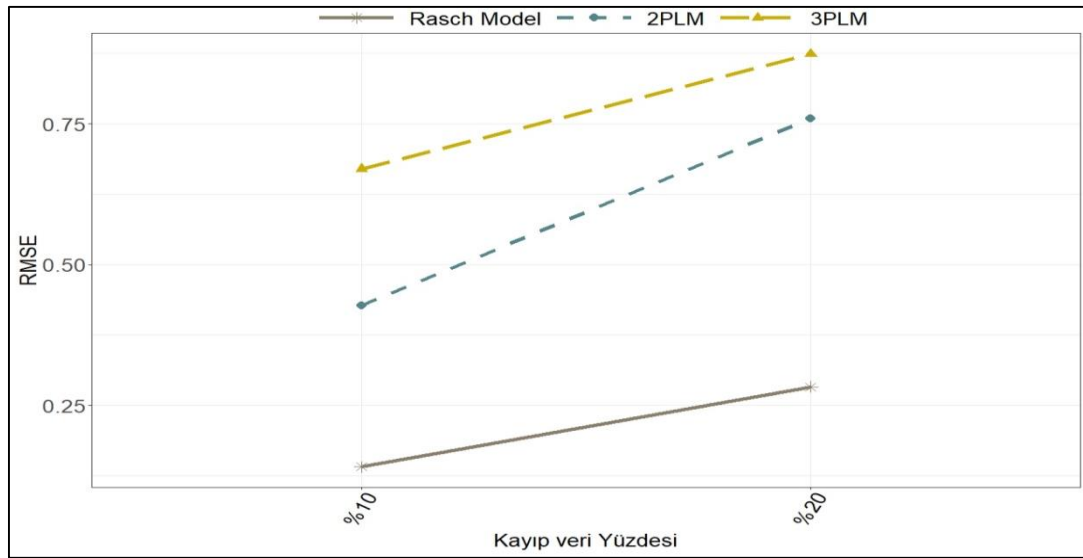
Kayıp veri yüzdesi ile model faktörünün $F(2,22)=9,334$; $p<0,05$ ve $\eta_p^2=0,687$ değerlerine göre etkileşim etkileri anlamlı bulunmuştur. Bu etkileşimi kayıp veri yüzdesi faktörünün koşullarında görebilmek için yapılan basit etki ANOVA sonuçları Tablo 12’de verilmektedir:

Tablo 12

Kayıp Veri Yüzdesi Faktörünün Her Bir Koşulunda Modellerin Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
%10 yüzde					
Modeller	6,159	2	3,079	85,54*	0,000
Hata	0,786	22	0,036		
%20 yüzde					
Modeller	3,643	2	1,821	50,58*	0,000
Hata	0,786	22	0,036		

Tablo 12 incelendiğinde %10 kayıp veri yüzdesi faktörünün model faktörü ile olan etkileşimlerinin $F(2,22)=85,54$; $p<0,05$ ve %20 kayıp veri yüzdesi faktörünün model faktörü ile olan etkileşimlerinin $F(2,22)=50,58$; $p<0,05$ anlamlı olduğu sonucuna ulaşılmıştır. Etkileşimlerin faktör düzeylerinde sahip olduğu ortalama farklılıkları görebilmek amacıyla elde edilen grafik Şekil 7’de verilmektedir:

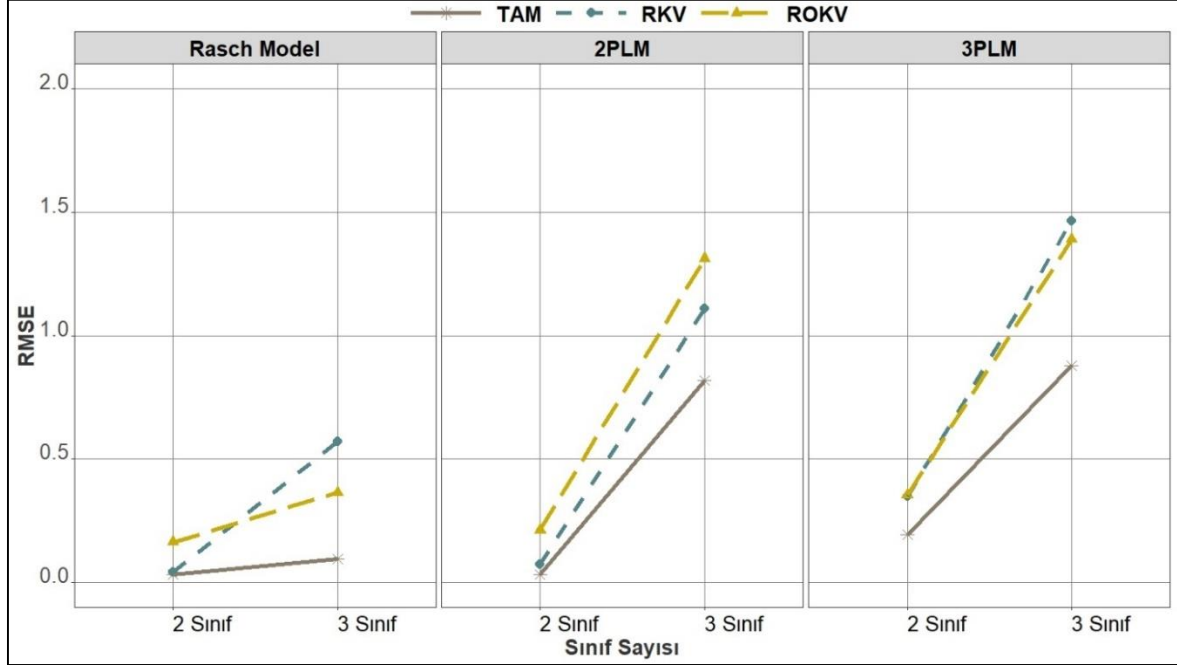


Şekil 7. Model ile kayıp veri yüzdesi faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Şekil 7’ e göre Rasch modelinde %10 kayıp veri etkileşiminde daha düşük RMSE değeri elde edilirken %20 kayıp veri etkileşiminde daha yüksek RMSE değeri elde edilmiştir. 2PL model için ise yine %10 kayıp veri yüzdesi etkileşiminde daha düşük RMSE değeri elde edilirken %20 kayıp veri yüzdesi etkileşiminde eğimde yüksek bir artış görülmüştür. 3PL modelde ise her iki kayıp veri yüzdesi etkileşiminde yüksek RMSE değerleri elde edilmiştir. Her üç modelde de %20 kayıp veri koşulunda madde güçlüklerine ilişkin RMSE değerlerinin %10 kayıp veri yüzdesi koşuluna göre daha yüksek olduğu belirlenmiştir. Eğimler incelendiğinde ise RMSE değerlerindeki en yüksek artışın Karma 2PL için elde edildiği görülmüştür. Kayıp veri yüzdesi faktörünün koşulları ile model faktörünün etkileşiminin anlamlılığının incelendiği Basit etkilere ait Tablo 12 aşağıda verilmektedir:

Üçlü etkileşimler incelendiğinde, kayıp veri türü, madde ve sınıf faktörlerinin $F(2,22)=8,637$; $p<0,05$ ve $\eta_p^2=0,440$; kayıp veri türü, sınıf ve model faktörlerinin $F(4,22)=6,676$;

$p < 0,05$ ve $\eta_p^2 = 0,548$ değerlerine dayalı olarak etkileşim etkilerinin anlamlı ve yüksek etki büyüklüğü değerlerine sahip olduğu söylenebilir. Kayıp veri türü, sınıf ve model faktörlerinin etkileşimlerinin incelendiği Şekil 8 ise aşağıda verilmektedir:

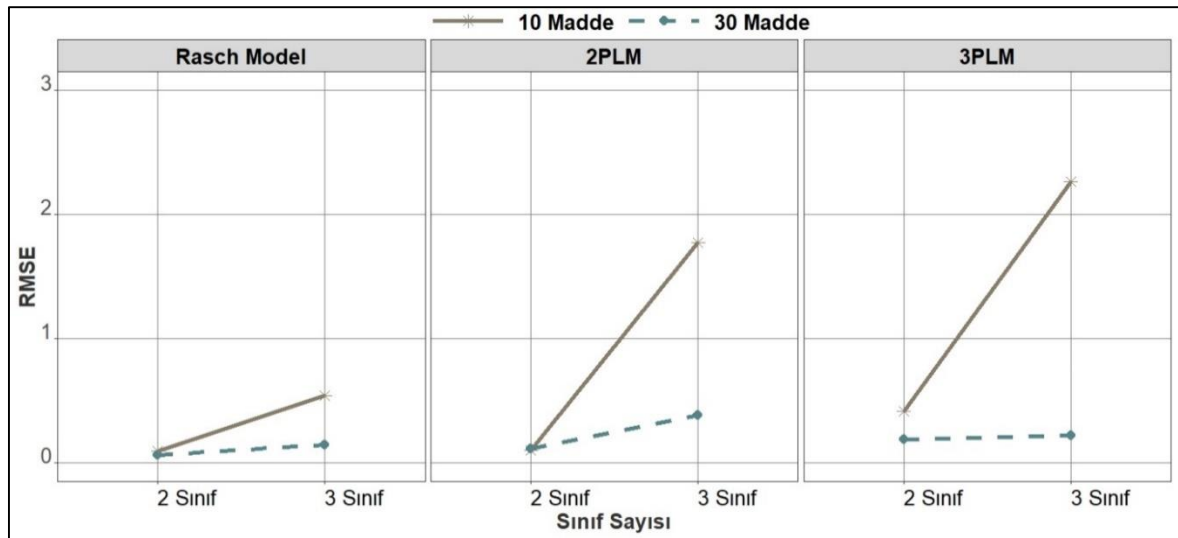


Şekil 8. Kayıp veri türü, sınıf ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafikler

Şekil 8 incelendiğinde Rasch modelin 2 sınıf koşulunda tam ve ROKV etkileşimlerinde madde güçlük parametre değerleri için kestirilen RMSE değerlerinin ROKV etkileşiminden elde edilen RMSE değerine göre daha düşük olduğu görülmektedir. 3 sınıf içinse tam veri etkileşiminde madde güçlük parametre değerleri için kestirilen RMSE değerlerinin ROKV ve ROKV etkileşimlerinde kestirildiği durumlara göre daha düşük olduğu, ROKV etkileşiminde elde edilen RMSE değerlerinin ise ROKV etkileşiminde elde edilen RMSE değerlerine göre daha düşük olduğu tespit edilmiştir. Karma 2PL modelin 2 sınıf koşulunda tam ve ROKV etkileşimlerinde madde güçlük parametre değerleri için kestirilen RMSE değerlerinin, ROKV etkileşimine göre daha düşük olduğu görülmektedir. 3 sınıf durumu içinse tam veri etkileşiminde madde güçlük parametre değerleri için kestirilen RMSE değerlerinin ROKV ve ROKV etkileşimlerinde kestirildiği durumlara göre daha düşük olduğu, ROKV etkileşiminde elde edilen RMSE değerlerinin ise ROKV etkileşiminde elde edilen RMSE değerlerine göre daha düşük olduğu tespit edilmiştir. Karma 3PL modelin 2 sınıf

koşulunda ise tam veri etkileşiminde madde güçlük parametre değerleri için kestirilen RMSE değerlerinin RKV ve ROKV koşullarına göre daha düşük olduğu görülmektedir. 3 sınıf durumu içinse tam veri etkileşiminde madde güçlük parametre değerleri için kestirilen RMSE değerlerinin RKV ve ROKV etkileşimlerinde kestirildiği durumlara göre daha düşük olduğu, RKV ve ROKV etkileşimlerinde elde edilen RMSE değerlerinin ise birbirine oldukça yakın olduğu görülmektedir.

Faktörlerin üçlü etkileşim etkileri incelendiğinde madde, sınıf ve model faktörlerinin $F(2,22)= 8,606$; $p<0,05$ ve $\eta_p^2= 0,439$ etkileşim etkilerinin anlamlı ve yüksek etki büyüklüğü değerlerine sahip olduğu tespit edilmiştir. Madde, sınıf ve model faktörlerinin etkileşimleri Şekil 9’da verilmektedir:

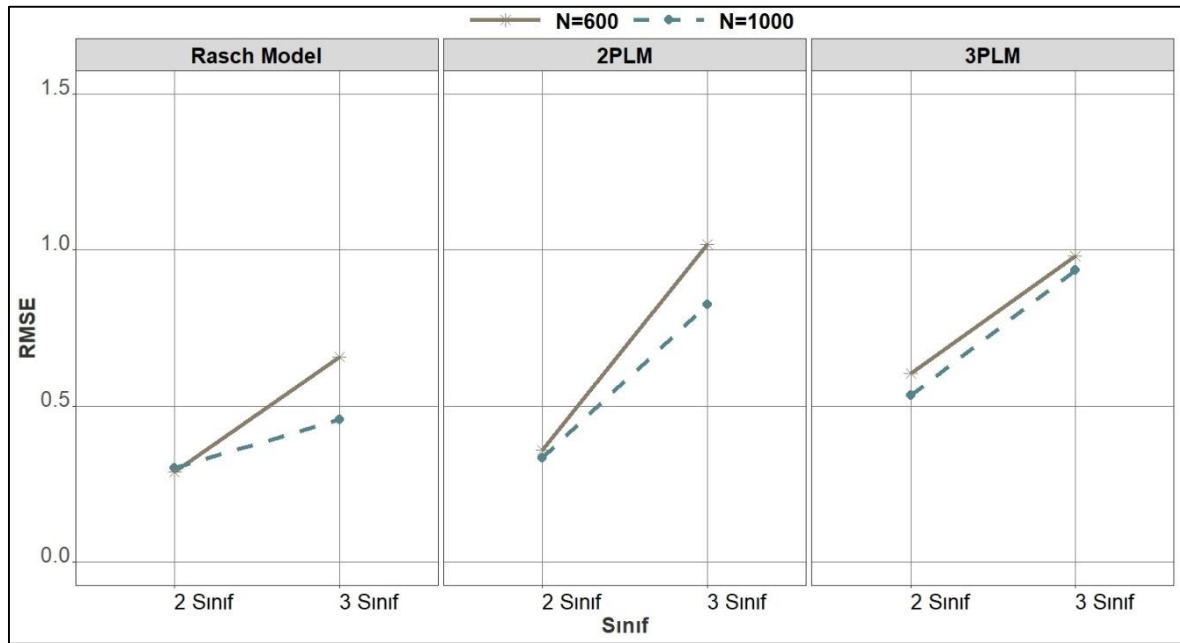


Şekil 9. Madde, sınıf ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafikler

Şekil 9 incelendiğinde Rasch model, 2 sınıf faktörlerinin 10 madde ve 30 madde etkileşimlerinde RMSE değerleri birbirine daha yakın değer alırken; Rasch model, 3 sınıf faktörlerinin 10 madde ve 30 madde etkileşimlerinde RMSE değerleri aralarındaki farkın arttığı görülmektedir. 2PL model için ise 2 sınıf faktörlerinin 10 madde ve 30 madde etkileşimlerinde RMSE değerleri aynı iken 3 sınıf olduğunda RMSE değerleri aralarındaki fark belirgin bir şekilde artmıştır. 3PL modelde ise 2 sınıf olduğunda 10 madde ve 30 madde etkileşimlerinde RMSE değerleri birbirine daha yakın değer alırken 3 sınıf olduğunda RMSE değerleri aralarındaki farklılık daha çok artmıştır. Her üç modelde de 2 ve 3 sınıf faktörlerinin 10 madde etkileşimlerinde madde güçlüklerine ilişkin RMSE değerlerinin 30 madde etkileşimlerine göre daha yüksek olduğu belirlenmiştir. 2 sınıf koşulunda 10 ve 30

madde arasındaki RMSE değerleri arasındaki farklılık ile 3 sınıf koşulunda 10 ve 30 madde arasındaki RMSE değerleri arasındaki farklılık arasındaki artışın Rasch model için en az, 3PL model için en fazla olduğu söylenebilir.

Bir diğer anlamlı bulunan üçlü etkileşimin örneklem, sınıf ve model faktörlerinin etkileşimleri olduğu görülmektedir. Örneklem, sınıf ve model faktörlerinin $F(2,22)= 4,025$; $p<0,05$ ve $\eta_p^2= 0,268$ etkileşim etkilerinin anlamlı ve yüksek etki büyüklüğü değerlerine sahip olduğu görülmektedir. Bu faktörlerin etkileşimlerine ait grafikler Şekil 10'da verilmektedir:



Şekil 10.Örneklem, sınıf ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama RMSE değerlerine ait grafikler

Şekil 10 incelendiğinde Karma Rasch ve Karma 2PL modellerin 2 sınıf koşulunda, örneklem büyüklüğü faktörünün etkileşimlerinde madde güçlük parametre değerleri için elde edilen RMSE değerlerinin birbirine yakın olduğu, 3 sınıf koşulunda ise 1000 örneklem büyüklüğü etkileşiminde elde edilen RMSE değerinin 600 örneklem büyüklüğü etkileşiminde elde edilen RMSE değerine göre daha düşük olduğu görülmektedir. Karma 3PL de ise 2 ve 3 sınıf faktörlerinin 1000 örneklem büyüklüğü etkileşimlerinde, madde güçlük parametre değerleri için kestirilen RMSE değerlerinin 600 örneklem büyüklüğü etkileşiminde kestirilen RMSE değerine göre daha düşük olduğu görülmektedir.

Sonuç olarak kayıp veri türü, sınıf sayısı, madde sayısı, model, örneklem büyüklüğü ve kayıp veri yüzdesi faktörlerinin ana etkilerinin ve etkileşim etkilerinin anlamlı ve yüksek etki

büyükklüğü değlerlerine sahip olması faktörlerin araştırmanın amacına uygun olarak seçildiğı ve Karma MTK modellerinde madde güçlük parametre kestirimi için önemli faktörler olduğı söylenebilir.

Madde Güçlük Parametrelerinden Elde Edilen Ortalama Yanlılık Değerleri İçin ANOVA Sonuçları

Madde güçlük parametre kestirimlerinden elde edilen ortalama yanılılık değleri kullanarak normallik varsayımı ANOVA analizleri yapılmadan önce test edilmiştir. Normallik testi ile elde edilen çarpıklık katsayısı 0,707 ve standart hatası 0,202 ve basıklık katsayısı ise 1,173 ve standart hatası 0,401 olarak elde edilmiştir. Yanlılık değlerinin normal dağılım göstermesi için yapılan uygun veri dönüştürme işlemleri sonucu veri dağılımının normalleşmediğı tespit edilmiştir. Verilerin basıklık ve çarpıklık katsayılarının -1 ile +1 değleri arasında olması ve varyans analizinin normal dağılım varsayımı sağlanmadığında güçlü bir yöntem olduğı (Rutherford, 2011) göz önüne alınarak yanılılık değlerinin farklı koşullar altındaki değişiminin incelenmesi için ANOVA kullanımının uygun olacağına karar verilmiştir. Madde güçlük değleri için yanılılık değleri alınarak gerçekleştirilen ANOVA sonuçları Tablo 13'te sunulmaktadır:

Tablo 13

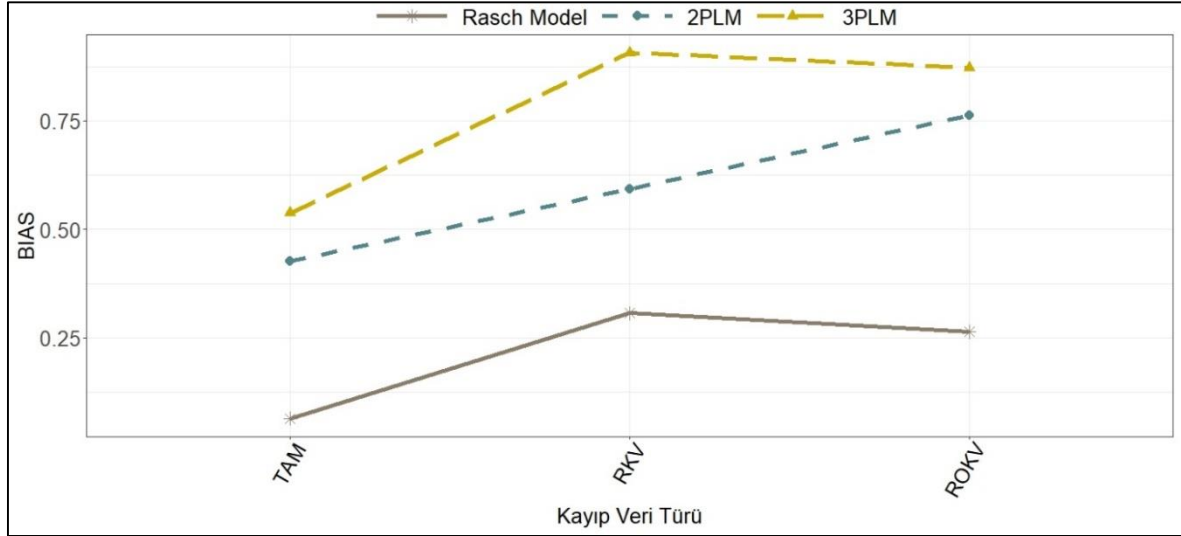
Madde Güçlük Parametrelerinden Elde Edilen Ortalama Yanlılık değerleri için ANOVA Sonuçları

Kaynak	η^2	η_p^2	sd	Kareler Toplamı	Ortalama kare	F	* $p<0,05$
Tür	0,047	0,476	2	0,004	0,002	9,987*	0,001
Madde	0,012	0,190	1	0,001	0,001	5,159	0,033
Örneklem	0,000	0,004	1	0,000	0,000	0,080	0,779
Sınıf	0,116	0,693	1	0,011	0,011	49,566*	0,000
Yüzde	0,001	0,024	1	0,000	0,000	0,539	0,470
Model	0,102	0,663	2	0,010	0,005	21,594*	0,000
Tür*madde	0,002	0,043	2	0,000	0,000	0,491	0,619
Tür*örneklem	0,013	0,201	2	0,001	0,001	2,775	0,084
Tür*sınıf	0,022	0,301	2	0,002	0,001	4,745*	0,019
Tür*yüzde	0,023	0,308	2	0,002	0,001	4,893*	0,017
Tür*model	0,069	0,572	4	0,007	0,002	7,362*	0,001
Madde*örneklem	0,003	0,059	1	0,000	0,000	1,377	0,253
Madde*sınıf	0,039	0,430	1	0,004	0,004	16,565*	0,001
Madde*yüzde	0,003	0,058	1	0,000	0,000	1,358	0,256
Madde*model	0,008	0,136	2	0,001	0,000	1,738	0,199
Örneklem*sınıf	0,005	0,089	1	0,000	0,000	2,146	0,157
Örneklem*yüzde	0,001	0,013	1	0,000	0,000	0,280	0,602
Örneklem*model	0,009	0,149	2	0,001	0,000	1,932	0,169
Sınıf*yüzde	0,003	0,047	1	0,000	0,000	1,076	0,311
Sınıf*model	0,044	0,457	2	0,004	0,002	9,262*	0,001
Yüzde*model	0,003	0,047	2	0,000	0,000	0,545	0,587
Tür*madde*örneklem	0,023	0,305	2	0,002	0,001	4,818*	0,018
Tür*madde*model	0,068	0,569	4	0,007	0,002	7,266*	0,001
Tür*sınıf*model	0,040	0,436	4	0,004	0,001	4,246*	0,011
Madde*sınıf*model	0,019	0,269	2	0,002	0,001	4,045*	0,032
Örneklem*sınıf*model	0,020	0,278	2	0,002	0,001	4,225*	0,028
Hata	0,052		22	0,005	0,000		

Tablo 13'teki tür, kayıp verilerin türünü; madde, verideki madde sayısını; örneklem, örneklem büyüklüğünü; sınıf, sınıf sayısını; yüzde, kayıp verilerin toplam yüzdesini; model, kullanılan Karma MTK modelini; temsil etmektedir. Kayıp veri türü faktörünün ana etkisinin $F(2,22)=9,987$; $p<0,05$ ve $\eta^2=0,047$ değerlerine dayalı olarak anlamlı ve orta düzey bir etkiye sahip olduğu söylenebilir. Madde faktörünün $F(1,22)=5,159$; $p<0,05$ ve $\eta^2=0,012$ değerlerine dayalı olarak anlamlı ve düşük bir etki büyüklüğü değerine; sınıf faktörünün $F(1,22)=49,566$; $p<0,05$ ve $\eta^2=0,116$ ve model faktörünün $F(2,22)=21,594$; $p<0,05$ ve $\eta^2=0,102$ değerlerine göre ana etkilerinin anlamlı ve yüksek etki büyüklüğü

değerlerine sahip olduğu görülmektedir. Örneklem ve kayıp veri yüzdesi faktörünün ana etkilerinin ise anlamlı olmadığı tespit edilmiştir.

Faktörlerin ikili etkileşimleri incelendiğinde ise kayıp veri türü ile sınıf faktörlerinin etkileşim etkileri için $F(2,22) = 4,745$, $p < 0,05$ ve $\eta_p^2 = 0,301$; kayıp veri türü ile kayıp veri yüzdesi faktörlerinin etkileşim etkileri için $F(2,22) = 4,893$; $p < 0,05$ ve $\eta_p^2 = 0,308$; kayıp veri türü ile model faktörlerinin etkileşim etkileri için $F(4,22) = 7,362$; $p < 0,05$ ve $\eta_p^2 = 0,572$ değerleri ile anlamlı ve yüksek etki büyüklüğü değerlerinin elde edildiği görülmektedir. Elde edilen ortalama RMSE değerlerine göre yapılan faktöriyel ANOVA analizleri sonucunda da kayıp veri türü ile sınıf, kayıp veri türü ile yüzde faktörlerinin etkileşimleri anlamlı bulunmuştur. Bu analizlerde bulunan sonuçlardan farklı olarak ortalama yanlılık değerleri alınarak yapılan faktöriyel ANOVA analizi sonucunda kayıp veri türü ile model etkileşimleri anlamlı bulunmuştur. Model ile kayıp veri türü faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama yanlılık değerlerine ait grafik Şekil 11’de verilmektedir:



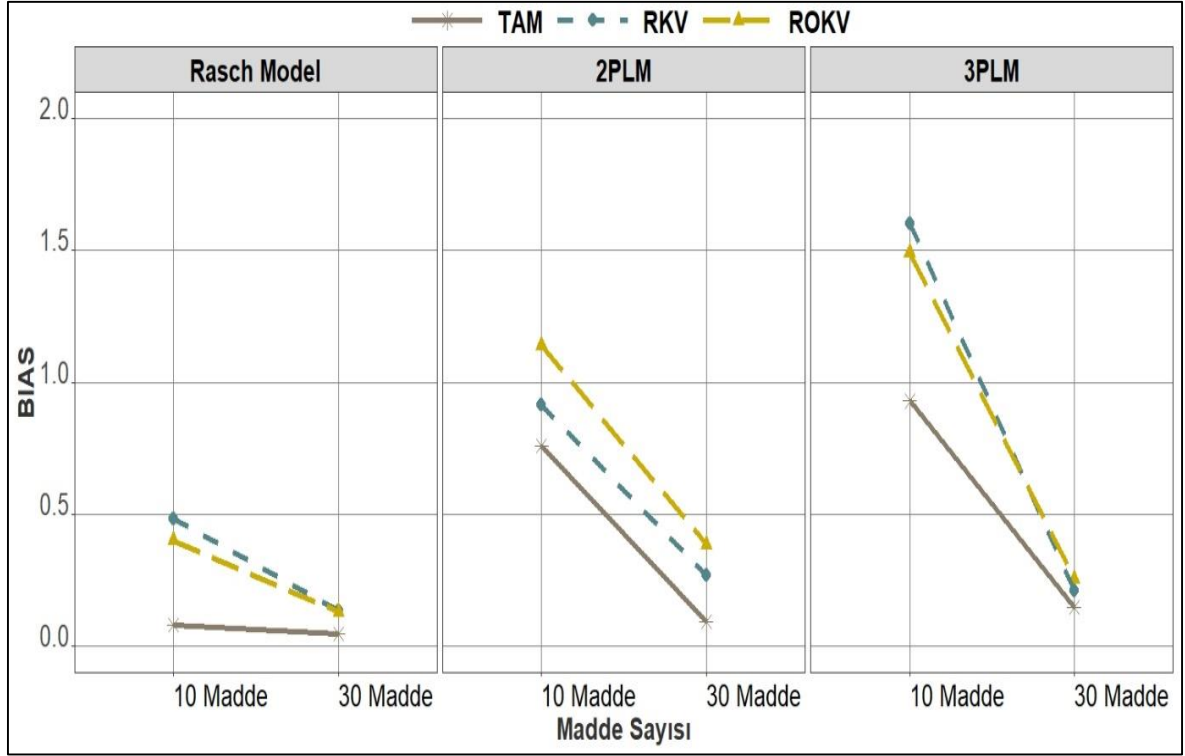
Şekil 11. Model ile kayıp veri türü faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama yanlılık değerlerine ait grafik

Şekil 11 incelendiğinde Rasch model ve tam veri etkileşiminde madde güçlük parametre değerleri için kestirilen yanlılık değerlerinin RKV ve ROKV etkileşimlerinden elde edilen yanlılık değerine göre daha düşük olduğu görülmektedir. Yine Rasch model için ROKV etkileşiminde madde güçlük parametre değerleri için kestirilen yanlılık değerlerinin RKV etkileşimlerinde kestirildiği durumlara göre daha düşük olduğu tespit edilmiştir. Karma 2PL

model ve tam veri etkileşiminde madde güçlük parametre değerleri için kestirilen yanlılık değerlerinin RKV ve ROKV etkileşimlerinden elde edilen RMSE değerine göre daha düşük olduğu görülmekte iken RKV etkileşiminde ise madde güçlük parametre değerleri için kestirilen yanlılık değerlerinin ROKV etkileşimlerinde kestirildiği durumlara göre daha düşük olduğu tespit edilmiştir. Karma 3PL tam veri etkileşiminde, madde güçlük parametre değerleri için kestirilen yanlılık değerlerinin RKV ve ROKV etkileşimlerinden elde edilen yanlılık değerine göre daha düşük olduğu, ROKV etkileşiminde ise madde güçlük parametre değerleri için kestirilen yanlılık değerlerinin RKV etkileşimlerinde kestirildiği durumlara göre daha düşük olduğu tespit edilmiştir. En düşük ortalama yanlılık değerine Rasch modelin tam veri koşulunda ulaşılırken en yüksek ortalama yanlılık değerine 3PL modelin RKV etkileşiminde ulaşılmıştır.

Elde edilen ortalama yanlılık değerleri ile yapılan faktöriyel ANOVA analizleri sonucu madde ile sınıf faktörlerinin ikili etkileşim etkilerinin $F(1,22) = 16,565$; $p < 0,05$ ve $\eta_p^2 = 0,430$; sınıf ile model faktörlerinin etkileşim etkilerinin $F(2,22) = 9,262$; $p < 0,05$ ve $\eta_p^2 = 0,457$ değerleri ile anlamlı ve yüksek etki büyüklüğü değerlerine sahip olduğu tespit edilmiştir. Bu sonuçlar ortalama RMSE değerleri ile yapılan ANOVA analizleri sonuçları ile benzerlik göstermektedir.

Üçlü etkileşimler incelendiğinde, kayıp veri türü, madde ve örneklem faktörlerinin $F(2,22) = 4,818$; $p < 0,05$ ve $\eta_p^2 = 0,305$; kayıp veri türü, madde ve model faktörlerinin $F(4,22) = 7,266$; $p < 0,05$ ve $\eta_p^2 = 0,569$; kayıp veri türü, sınıf ve model faktörlerinin $F(4,22) = 4,246$; $p < 0,05$ ve $\eta_p^2 = 0,436$ etkileşim etkilerinin anlamlı ve yüksek etki büyüklüğü değerlerine sahip olduğu söylenebilir. Ortalama RMSE değerleri ile yapılan ANOVA analizleri sonuçlarında kayıp veri türü, madde ve sınıf faktörlerinin; kayıp veri türü, sınıf ve model faktörlerinin etkileşim etkileri anlamlı ve yüksek etki büyüklüğüne sahip olduğu bulunmuştur. Ortalama RMSE değerlerinden farklı olarak ortalama yanlılık değerleri ile gerçekleştirilen ANOVA analizleri sonucunda kayıp veri türü, madde ve model faktörleri için elde edilen grafik Şekil 12’de verilmektedir:



Şekil 12. Kayıp veri türü, madde ve model faktörlerinin etkileşimi sonucu madde güçlük parametresi için elde edilen ortalama yanlılık değerlerine ait grafik

Şekil 12 incelendiğinde Rasch model ve 10 madde faktörleri ile Rasch model ve 30 madde faktörlerinin tam veri faktörü ile etkileşimlerinde elde edilen ortalama yanlılık değerlerinin ROKV ve RKV etkileşimlerinden elde edilen ortalama yanlılık değerlerine göre daha düşük olduğu, Rasch model, 10 madde ve ROKV faktörlerinin etkileşimleri için elde edilen yanlılık değerlerinin; RKV etkileşimine göre daha düşük olduğu, Rasch model, 30 madde faktörlerinin ROKV ve RKV etkileşimlerinde elde edilen ortalama yanlılık değerlerinin birbirine yakın olduğu görülmektedir. 2PL model, 10 madde ve 30 madde koşullarında; tam veri durumunda elde edilen ortalama yanlılık değerlerinin ROKV ve RKV koşullarında elde edilen ortalama yanlılık değerlerine göre daha düşük olduğu, 2PL model, 10 madde ve 30 madde koşullarında elde edilen ortalama yanlılık değerlerinin ROKV koşulunda RKV koşuluna göre daha yüksek olduğu gözlenmiştir. 3PL model ve 10 madde koşullarında; tam veri durumunda elde edilen ortalama yanlılık değerlerinin ROKV ve RKV koşullarında elde edilen ortalama yanlılık değerlerine göre daha düşük olduğu, 3PL model ve 10 madde koşullarında elde edilen ortalama yanlılık değerlerinin ROKV koşulunda RKV koşuluna göre daha düşük olduğu gözlenmiştir. 3PL model ve 30 madde koşullarında; tam veri

durumunda elde edilen ortalama yanlılık değerlerinin ROKV ve RKV koşullarında elde edilen ortalama yanlılık değerlerine göre daha düşük olduğu, 3PL model ve 30 madde koşullarında elde edilen ortalama yanlılık değerlerinin RKV koşulunda ROKV koşuluna göre daha düşük olduğu gözlenmiştir.

Faktörlerin üçlü etkileşim etkileri incelendiğinde madde, sınıf ve model faktörlerinin $F(2,22)= 4,045$; $p<0,05$ ve $\eta_p^2= 0,269$; örneklem, sınıf ve model faktörlerinin üçlü etkileşimleri de $F(2,22)= 4,225$; $p<0,05$ ve $\eta_p^2= 0,278$ değerlerine göre yüksek ve anlamlı etkileşimlere sahip olduğu söylenebilir. Bulunan sonuçlar ortalama RMSE değerleri alınarak elde edilen sonuçlarla benzerlik göstermektedir.

Madde Ayırt Edicilik Parametrelerine Ait Bulgular

Bu kısımda öncelikli olarak araştırma kapsamında incelenen simülasyon koşullarına (kayıp veri türü, madde sayısı, örneklem büyüklüğü, sınıf sayısı, kayıp veri yüzdesi, kullanılan Karma MTK modeli) göre madde ayırt edicilik parametresi için hesaplanan ortalama RMSE ve yanlılık değerleri sunulmuştur. Daha sonra simülasyon koşullarında madde ayırt edicilik parametresi için hesaplanan ortalama RMSE ve yanlılık değerlerinin ve bu değerlere göre faktörlerin etkileşim etkilerinin nasıl farklılaştığını incelemek amacıyla gruplar arası ANOVA analizi yapılmıştır.

Karma Rasch modelde ayırt edicilik parametresi 1 olduğu için Karma 2PL ve Karma 3PL MTK modellerinin madde ayırt edicilik parametreleri için ortalama RMSE ve yanlılık değerleri elde edilmiştir. Simülasyon koşulları için gösterimler, belirli bir durum için faktörlerin kombinasyonunu temsil etmek üzere tasarlanmıştır. Simülasyon koşullarını belirtmek için 10-14 haneden oluşan bir kodlama yapılmıştır ve bu kodlamada ilk karakter modeli (2PL, 3PL), ikinci karakter sınıf sayısını (2, 3), üçüncü karakter kayıp veri yüzdesini (10P, 20P), dördüncü karakter örneklem büyüklüğünü (600, 1000), beşinci karakter ise madde sayısını (10,30) göstermektedir. Örneğin 2PL210P60010 gösteriminde 2PL model (2PL), sınıf sayısı (2), kayıp veri yüzdesi (10P), örneklem büyüklüğü (600), madde sayısı (10) ile ifade edilmektedir. Tablo 14'te Karma 2PL model için madde ayırt edicilik parametre değerlerinin ortalama RMSE ve yanlılık değerleri verilmiştir:

Tablo 14

Karma 2PL Model İçin Madde Ayırt Edicilik Parametre Değerlerinin Ortalama RMSE ve Yanlılık Değerleri

Koşullar	TAM		Rastgele Kayıp Veri (RKV)		Rastgele Kayıp Veri (ROKV)	
	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE
2PL210P60010	-0,0014	0,165	-0,0060	0,212	-0,0267	0,497
2PL210P60030	-0,0053	0,179	-0,0087	0,310	-0,0078	0,277
2PL210P100010	-0,0017	0,057	0,0012	0,061	-0,0019	0,063
2PL210P100030	-0,0006	0,024	-0,0067	0,038	-0,0017	0,052
2PL220P60010	-0,0280	0,234	-0,0020	0,261	-0,0313	0,563
2PL220P60030	-0,0035	0,259	-0,0142	0,256	-0,0289	0,523
2PL220P100010	-0,0027	0,183	-0,0176	0,197	-0,0089	0,208
2PL220P100030	-0,0008	0,032	-0,0070	0,041	-0,0029	0,055
2PL310P60010	-0,0139	0,843	-0,0151	0,937	0,0995	1,222
2PL310P60030	-0,0103	0,725	-0,0709	0,873	-0,0308	0,916
2PL310P100010	-0,0141	0,866	-0,0515	0,910	-0,1330	1,469
2PL310P100030	-0,0056	0,675	-0,0090	0,784	-0,0186	0,833
2PL320P60010	-0,0246	0,975	0,1069	1,277	0,1190	1,366
2PL320P60030	-0,032	0,837	-0,0635	0,927	-0,0427	0,982
2PL320P100010	-0,0565	0,922	-0,0733	0,981	0,0890	1,032
2PL320P100030	-0,0509	0,786	-0,0140	0,854	-0,0220	0,967

Tablo 14 incelendiğinde madde ayırt edicilik parametre kestirimlerinin tam veri durumu için ortalama RMSE değerlerinin daha düşük, RKV koşulundaki durumlar için biraz daha yüksek, ROKV durumu içinse en yüksek değerlerine ulaşıldığı görülmektedir. En düşük RMSE değerleri tam, RKV ve ROKV koşullarının 2 sınıf, %10 kayıp veri yüzdesi, 1000 örneklem büyüklüğü ve 30 madde koşulları için elde edilirken, en yüksek RMSE değerine ise ROKV koşulunun 3 sınıf, %10 kayıp veri yüzdesi, 1000 örneklem büyüklüğü ve 10 madde koşulunda ulaşılmıştır. Madde ayırt edicilik parametre kestirimleri için en yüksek RMSE değerlerinin elde edildiği koşullar ile madde güçlük parametre kestirim değerleri için en yüksek ortalama RMSE değerlerinin elde edildiği durumlarla tutarlı olduğu görülmektedir. Tablo 15'te ise Karma 3PL model için madde ayırt edicilik parametre değerlerinin ortalama RMSE ve yanlılık değerleri verilmiştir:

Tablo 15

Karma 3PL Model İçin Madde Ayırt Edicilik Parametre Değerlerinin Ortalama RMSE ve Yanlılık Değerleri

Koşullar	TAM		Rastgele Kayıp Veri (RKV)		Rastgele Olmayan Kayıp Veri (ROKV)	
	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE
3PL210P60010	0,0042	0,331	0,0056	0,645	0,0051	0,634
3PL210P60030	0,0028	0,161	0,0046	0,234	0,0078	0,171
3PL210P100010	0,0189	0,264	0,0383	0,320	0,0282	0,216
3PL210P100030	0,0062	0,094	0,0078	0,115	0,0096	0,120
3PL220P60010	0,0028	0,371	-0,045	0,654	-0,065	0,662
3PL220P60030	0,0037	0,192	0,0046	0,229	-0,020	0,262
3PL220P100010	0,0189	0,337	-0,048	0,405	-0,019	0,417
3PL220P100030	0,0078	0,101	0,0102	0,151	0,009	0,176
3PL310P60010	0,1064	1,140	0,1080	1,293	0,1230	1,337
3PL310P60030	-0,0180	0,776	-0,0110	0,821	-0,0130	0,857
3PL310P100010	-0,0124	0,988	0,1040	1,113	0,0947	1,228
3PL310P100030	-0,012	0,581	-0,0080	0,696	-0,0130	0,705
3PL320P60010	0,1010	1,262	0,1320	1,463	-0,1290	1,472
3PL320P60030	-0,0280	0,920	-0,023	0,943	-0,026	0,952
3PL320P100010	0,0860	1,023	0,0980	1,242	0,1130	1,281
3PL320P100030	-0,0390	0,723	-0,042	0,817	-0,076	0,832

Tablo 15 incelendiğinde madde ayırt edicilik parametre kestirimlerinin tam veri durumu için ortalama RMSE değerlerinin daha düşük, RKV ve ROKV koşulları içinse daha yüksek değerler elde edildiği görülmektedir. En düşük RMSE değerlerine tam, RKV ve ROKV koşullarının 2 sınıf, %10 kayıp veri yüzdesi, 1000 örneklem büyüklüğü ve 30 madde koşulları için elde edilirken, en yüksek RMSE değerine ise RKV ve ROKV koşulunun 3 sınıf, %20 kayıp veri yüzdesi, 600 örneklem büyüklüğü ve 10 madde koşulunda ulaşılmıştır. Bu sonuçlar ile Karma 2PL modelde madde ayırt edicilik ve madde güçlük parametre kestirim değerleri için en yüksek RMSE değerlerinin elde edildiği koşulların benzerlik gösterdiği ifade edilebilir.

Madde Ayırt Edicilik Parametrelerinden Elde Edilen Ortalama RMSE Değerleri İçin ANOVA Sonuçları

ANOVA analizleri yapılmadan önce normallik ve varyansların homojenliği varsayımları test edilmiştir. Çarpıklık ve basıklık katsayıları incelendiğinde çarpıklık katsayısı 0,235 ve standart hatası 0,246; basıklık katsayısı ise -1,146 ve standart hatası 0,488 olarak

bulunmuştur. RMSE değerlerinin normal dağılım göstermesi için yapılan uygun veri dönüştürme işlemleri sonucu veri dağılımının normalleşmediği tespit edilmiştir. Verilerin basıklık ve çarpıklık katsayılarının -1 ile +1 değerleri arasında olması ve varyans analizinin normal dağılım varsayımının ihlaline karşı güçlü bir yöntem olmasından (Rutherford, 2011) dolayı RMSE değerlerinin farklı koşullar altındaki değişiminin incelenmesi için ANOVA kullanımının uygun olacağına karar verilmiştir. Madde ayırt edicilik değerleri için RMSE değerleri alınarak gerçekleştirilen ANOVA analiz sonuçları Tablo 16'da sunulmaktadır.

Tablo 16

Madde Ayırt Edicilik Parametrelerinden Elde edilen RMSE değerleri için ANOVA Sonuçları

Kaynak	η^2	η_p^2	sd	Kareler Toplamı	Ortalama kare	F	* $p<0,05$
Tür	0,026	0,905	2	0,445	0,223	57,014*	0,000
Madde	0,085	0,969	1	1,442	1,442	369,207*	0,000
Örneklem	0,033	0,923	1	0,565	0,565	144,609*	0,000
Sınıf	0,769	0,996	1	13,065	13,065	3345,500*	0,000
Yüzde	0,009	0,760	1	0,149	0,149	38,052*	0,000
Model	0,010	0,782	1	0,168	0,168	43,128*	0,000
Tür*madde	0,004	0,593	2	0,068	0,034	8,756*	0,005
Tür*örneklem	0,002	0,379	2	0,029	0,014	3,658	0,058
Tür*sınıf	0,002	0,428	2	0,035	0,018	4,483*	0,035
Tür*yüzde	0,000	0,012	2	0,001	0,000	0,072	0,931
Tür*model	0,003	0,558	2	0,059	0,030	7,570*	0,007
Madde*örneklem	0,000	0,099	1	0,005	0,005	1,318	0,273
Madde*sınıf	0,010	0,777	1	0,163	0,163	41,723*	0,000
Madde*yüzde	0,000	0,007	1	0,000	0,000	0,084	0,777
Madde*model	0,016	0,856	1	0,279	0,279	71,379*	0,000
Örneklem*sınıf	0,002	0,404	1	0,032	0,032	8,141*	0,015
Örneklem*yüzde	0,000	0,123	1	0,007	0,007	1,686	0,219
Örneklem*model	0,000	0,021	1	0,001	0,001	0,258	0,621
Sınıf*yüzde	0,000	0,103	1	0,005	0,005	1,379	0,263
Sınıf*model	0,001	0,185	1	0,011	0,011	2,716	0,125
Yüzde*model	0,000	0,022	1	0,001	0,001	0,268	0,614
Tür*örneklem*sınıf	0,003	0,489	2	0,045	0,022	5,738*	0,018
Örneklem*sınıf*model	0,003	0,495	1	0,046	0,046	11,774*	0,005
Hata	0,003		12	0,047	0,004		

Tablo 16'da tür, kayıp verilerin türünü; madde, verideki madde sayısını; örneklem, örneklem büyüklüğünü; sınıf, sınıf sayısını; yüzde, kayıp verilerin toplam yüzdesini; model, kullanılan Karma MTK modelini; temsil etmektedir. Kayıp veri türü faktörünün ana etkisinin $F(2,12)=57,014$; $p<0,05$ ve $\eta^2=0,026$ değerlerine dayalı olarak anlamlı ve düşük düzey bir etkiye sahip olduğu söylenebilir. Madde faktörünün ana etkisinin $F(1,12)=369,207$; $p<0,05$ ve $\eta^2=$

0,085 değerlerine göre anlamlı ve orta düzey bir etki büyüklüğü değerine; sınıf faktörünün $F(1,12)= 3345,500$; $p<0,05$ ve $\eta^2= 0,769$ ana etkilerinin anlamlı ve yüksek etki büyüklüğü değerine ve model faktörünün $F(1,12)= 43,128$; $p<0,05$ ve $\eta^2= 0,010$ değerlerine göre ana etkilerinin anlamlı ve düşük etki büyüklüğü değerlerine sahip olduğu görülmektedir. Örneklem faktörünün ana etkisinin $F(1,12)= 144,609$; $p<0,05$ ve $\eta^2= 0,033$ değerlerine göre anlamlı ve düşük düzey bir etki büyüklüğü değerine; kayıp veri yüzdesi faktörünün ana etkisinin $F(1,12)= 38,052$; $p<0,05$ ve $\eta^2= 0,009$ değerlerine göre anlamlı ve düşük düzey bir etki büyüklüğü elde edilmiştir.

Faktörlerin ikili etkileşimleri incelendiğinde ise kayıp veri türü ile madde faktörlerinin etkileşim etkileri için $F(2,12)= 8,756$; $p<0,05$ ve $\eta_p^2= 0,593$ değerleri ile anlamlı ve yüksek etki büyüklüğü değerlerinin elde edildiği görülmektedir. Kayıp veri türü ile madde faktörlerinin ikili etkileşimlerinin anlamlı bulunması sonucu madde faktörünün her bir düzeyinin kayıp veri türünün faktörün hangi düzeyinde istatistiksel olarak anlamlı olup olmadığını belirlemek için SPSS sözdizimi ile basit etkiler test edilmiştir. Basit etkilere ait Tablo 17 aşağıda verilmektedir:

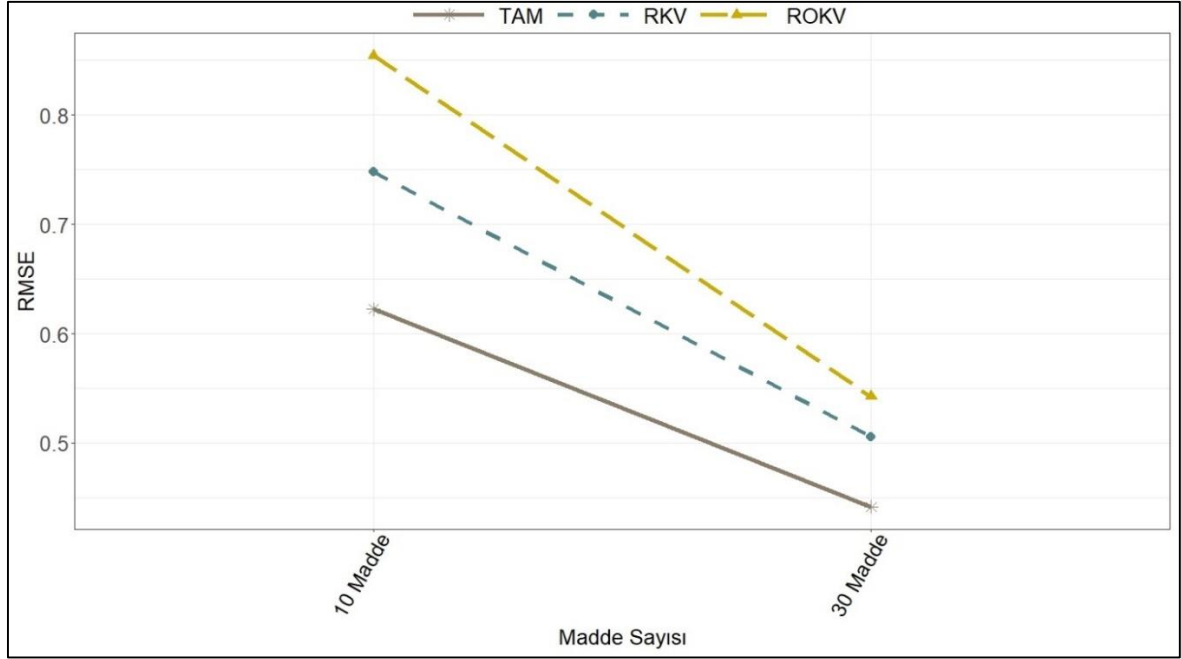
Tablo 17

Madde Faktörünün Her Bir Koşulunda Kayıp Veri Türlerine Ait Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
10 madde					
Kayıp veri türü	0,430	2	0,215	53,75*	0,000
Hata	0,047	12	0,004		
30 madde					
Kayıp veri türü	0,083	2	0,042	10,5*	0,0023
Hata	0,047	12	0,004		

* $p<0,05$

Tablo 17 incelendiğinde 10 madde ve kayıp veri türü faktörünün etkileşimleri arasında $F(2,12)= 53,75$; $p<0,05$ anlamlı farklılık olduğu; 30 madde ve kayıp veri türü faktörünün etkileşimleri $F(2,12)= 10,5$; $p<0,05$ arasında da anlamlı farklılık olduğu görülmektedir. Etkileşimlerin faktör düzeylerindeki ortalama farklılıklarını incelemek amacıyla faktörlerinin etkileşimi sonucu elde edilen RMSE değerlerine ait grafik Şekil 13'te verilmektedir:



Şekil 13. Kayıp veri türü ve madde faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Şekil 13 incelendiğinde 10 madde koşulunda tam veri için hesaplanan madde ayırt edicilik parametre kestirimlerinin RMSE değerinin en düşük, ROKV etkileşiminde ise RMSE değerinin biraz daha yüksek, ROKV etkileşiminde ise RMSE değerinin en yüksek olduğu görülmektedir. Benzer şekilde 30 madde koşulunda tam veri için hesaplanan madde ayırt edicilik parametre kestirimlerinin RMSE değerinin en düşük, ROKV etkileşiminde ise RMSE değerinin biraz daha yüksek, ROKV etkileşiminde ise RMSE değerinin en yüksek olduğu görülmektedir. Madde sayısı arttıkça tüm veri türlerinin eğrilerinde düşüş olduğu görülmekte ve RMSE değerlerinin madde sayısı arttıkça azaldığı sonucuna ulaşılmıştır.

Faktörlerin ikili etkileşimleri incelendiğinde kayıp veri türü ile sınıf faktörlerinin etkileşim etkileri için $F(2,12) = 4,483$; $p < 0,05$ ve $\eta_p^2 = 0,428$ değerleri ile anlamlı ve yüksek etki büyüklüğü değerlerinin elde edildiği görülmektedir. Kayıp veri türü ile sınıf faktörlerinin ikili etkileşimlerinin anlamlı bulunması sonucu faktör düzeylerinde istatistiksel olarak anlamlılığı belirlemek için yapılan basit etki Tablo 18’de verilmiştir:

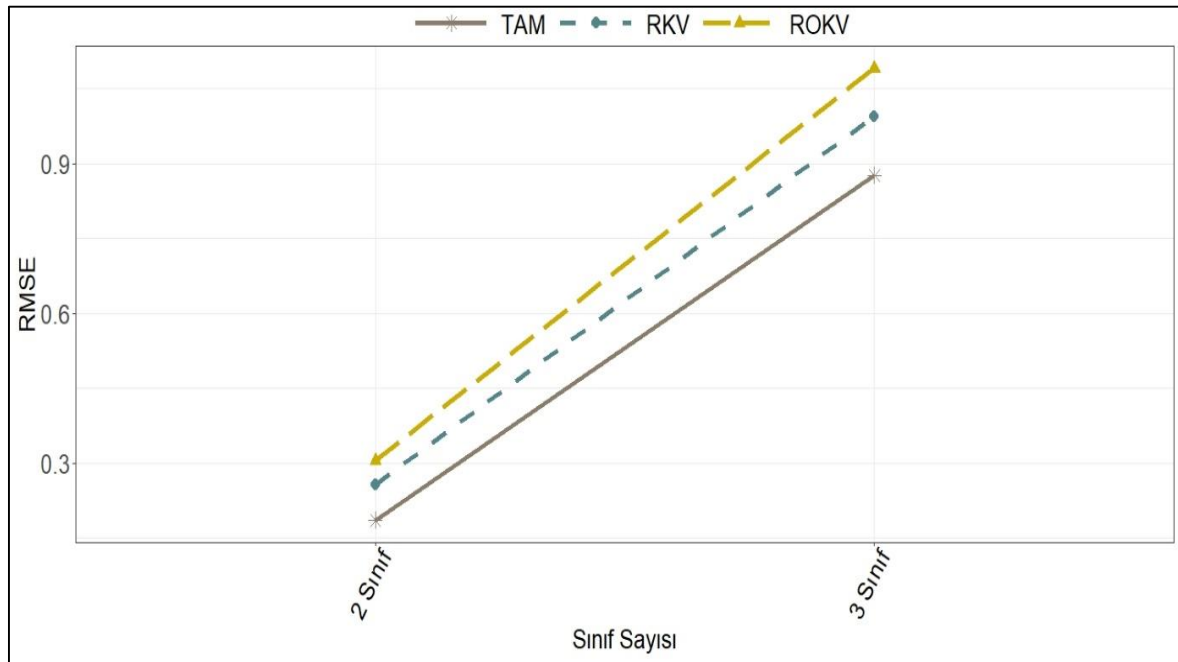
Tablo 18

Sınıf Faktörünün Her Bir Koşulunda Kayıp Veri Türlerine Ait Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
2 Sınıf					
Kayıp veri türü	0,116	2	0,058	14,5*	0,0006
Hata	0,047	12	0,004		
3 Sınıf					
Kayıp veri türü	0,365	2	0,182	45,5	0,0000
Hata	0,047	12	0,004		

*p<0,05

Tablo 18 incelendiğinde 2 sınıf koşulunda kayıp veri türü faktörü ile etkileşimlerin $F(2,12)=14,5$; $p<0,05$ ve 3 sınıf koşulunda kayıp veri türü faktörü ile etkileşimlerin $F(2,12)=45,5$; $p<0,05$ arasında anlamlı farklılık olduğu görülmektedir. Etkileşimlerin faktör düzeylerindeki ortalama farklılıklarını görebilmek için amacıyla faktörlerinin etkileşimi sonucu elde edilen RMSE değerlerine ait grafik Şekil 14’te verilmektedir:



Şekil 14. Kayıp veri türü ve sınıf faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Şekil 14 incelendiğinde, 2 sınıf koşulunda kayıp veri türleri için madde ayırt edicilik parametre kestiriminde elde edilen ortalama RMSE değerlerinin 3 sınıf koşulu için elde edilen ortalama RMSE değerlerinden daha düşük olduğu görülmektedir. Ayrıca 2 sınıf ve 3

sınıf koşullarında en düşük ortalama RMSE değerleri tam veri etkileşimlerinde, en yüksek ortalama RMSE değerleri ise ROKV etkileşiminde elde edilmiştir.

Kayıp veri türü ile model faktörlerinin etkileşim etkileri için $F(2,12) = 7,570$; $p < 0,05$ ve $\eta_p^2 = 0,558$ değerleri ile anlamlı ve yüksek etki büyüklüğü değerlerinin elde edildiği görülmektedir. Kayıp veri türü faktörünün koşulları ile model faktörünün anlamlılığının incelendiği basit etkilere ait Tablo 19 aşağıda verilmektedir:

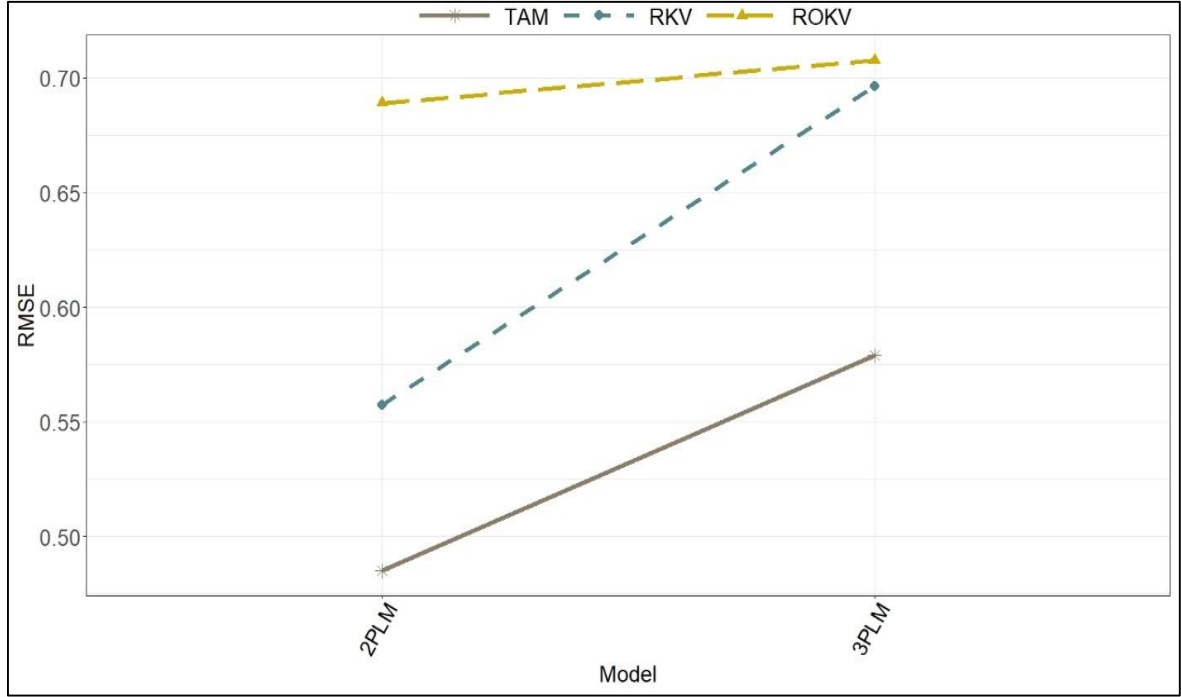
Tablo 19

Kayıp Veri Türü Faktörünün Her Bir Koşulunda Modellerin Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
Tam Veri					
Modeller	0,071	1	0,071	17,75*	0,0002
Hata	0,047	12	0,004		
RKV					
Modeller	0,154	1	0,154	38,5*	0,001
Hata	0,047	12	0,004		
ROKV					
Modeller	0,003	1	0,003	0,75	0,49
Hata	0,047	12	0,004		

* $p < 0,05$

Tablo 19 incelendiğinde tam veri ve model faktörünün etkileşimleri $F(1,12) = 17,75$; $p < 0,05$ ile RKV ve model faktörünün etkileşimleri $F(1,12) = 38,5$; $p < 0,05$ arasında anlamlı farklılık olduğu, ROKV ve model faktörünün etkileşimleri $F(1,12) = 0,75$; $p > 0,05$ arasında anlamlı farklılık olmadığı görülmektedir. Etkileşimlerin faktör düzeylerindeki ortalama farklılıklarını görebilmek için elde edilen RMSE değerlerine ait grafik Şekil 15'te verilmektedir:



Şekil 15. Kayıp veri türü ve model faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Şekil 15 incelendiğinde Karma 2PL için tam veri koşulunda madde ayırt edicilik parametre kestirimi için en düşük RMSE değerinin elde edildiği, RKV koşulunda RMSE değerlerinin arttığı, ROKV koşulunda ise en yüksek RMSE değerlerinin elde edildiği gözlenmiştir. Karma 3PL içinse tam veri etkileşiminde madde ayırt edicilik parametre kestiriminde en düşük RMSE değerinin elde edildiği, RKV ve ROKV etkileşimlerinde ise birbirine yakın RMSE değerleri elde edilmiştir.

Faktörlerin ikili etkileşimleri incelendiğinde ise madde ile sınıf faktörlerinin etkileşim etkileri için $F(2,12) = 41,723$; $p < 0,05$ ve $\eta_p^2 = 0,777$ değerleri ile anlamlı ve yüksek etki büyüklüğü değerlerinin elde edildiği görülmektedir. Madde ile sınıf faktörlerinin ikili etkileşimlerinin anlamlı bulunması sonucu sınıf faktörünün her bir düzeyinin madde faktörünün düzeyinde istatistiksel olarak anlamlılığının incelendiği basit etkilere ait Tablo 20 aşağıda verilmektedir:

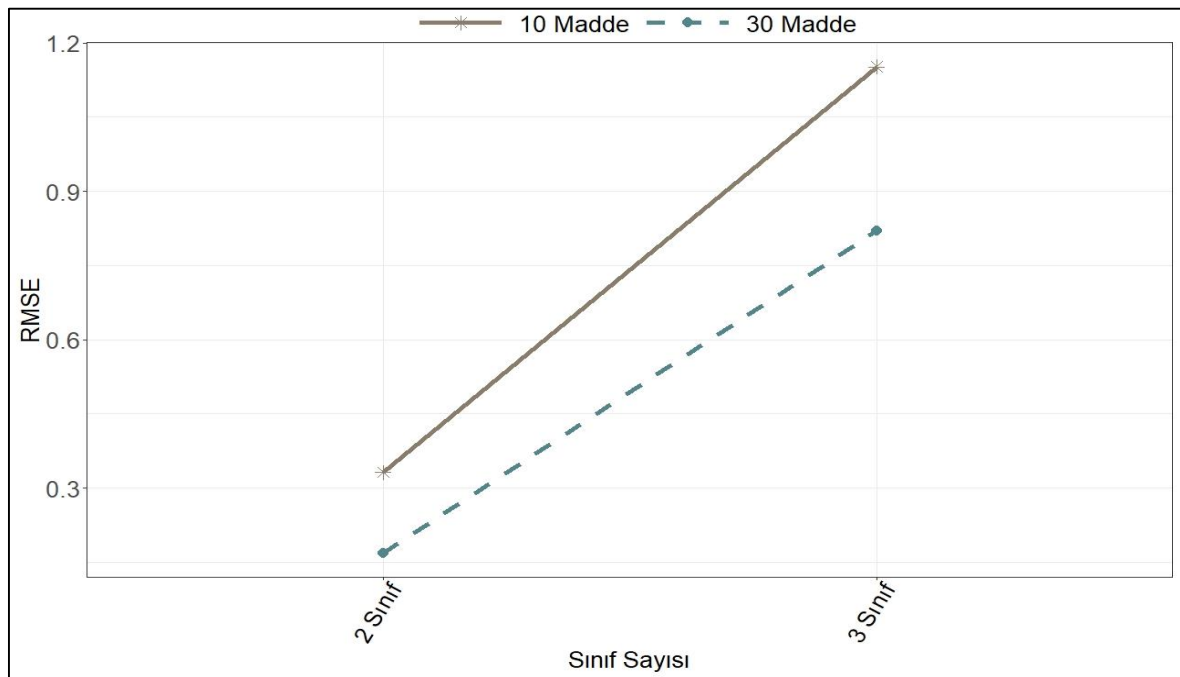
Tablo 20

Madde Faktörünün Her Bir Koşulunda Sınıf Faktörünün Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
2 sınıf					
Madde sayısı	0,318	1	0,318	79,5*	0,000
Hata	0,047	12	0,004		
3 sınıf					
Madde sayısı	1,287	1	1,287	321,75*	0,000
Hata	0,047	12	0,004		

*p<0,05

Tablo 20 incelendiğinde sınıf faktörünün 2 sınıf koşulunda madde faktörü ile olan etkileşimlerin $F(1,12)=79,5$; $p<0,05$ ve sınıf faktörünün 3 sınıf koşulunda madde faktörü ile olan etkileşimlerin $F(1,12)=321,75$; $p<0,05$ anlamlı olduğu sonucuna ulaşılmıştır. Etkileşimler arasındaki ortalama farklılıkları faktör düzeylerinde görebilmek için grafikten yardım alınmıştır. Madde ile sınıf faktörlerinin etkileşimi sonucu elde edilen RMSE değerlerine ait grafik Şekil 16’da verilmektedir:



Şekil 16. Madde ve sınıf faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Şekil 16 incelendiğinde, 2 sınıf ve 30 madde etkileşiminde madde ayırt edicilik parametre kestiriminde elde edilen ortalama RMSE değerinin, 2 sınıf ve 10 madde etkileşimi için elde edilen RMSE değerine daha düşük olduğu görülmektedir. 3 sınıf ve 30 madde etkileşimi için elde edilen RMSE değerinin ise 2 sınıf ve 30 madde etkileşimi için elde edilen RMSE değerinden daha büyük olduğu, 3 sınıf ve 10 madde etkileşimi için elde edilen RMSE değerinden ise daha küçük olduğu tespit edilmiştir.

Faktörlerin ikili etkileşimleri incelendiğinde madde ve model faktörlerinin ikili etkileşimleri için $F(2,12) = 71,379$; $p < 0,05$ ve $\eta_p^2 = 0,856$ değerleri ile anlamlı ve yüksek etki büyüklüğüne sahip olduğu görülmektedir. Madde ile model faktörlerinin ikili etkileşimlerinin anlamlı bulunması sonucu madde faktörünün her bir düzeyinin model faktörünün düzeyinde istatistiksel olarak anlamlılığının incelendiği basit etkilere ait Tablo 21 aşağıda verilmektedir:

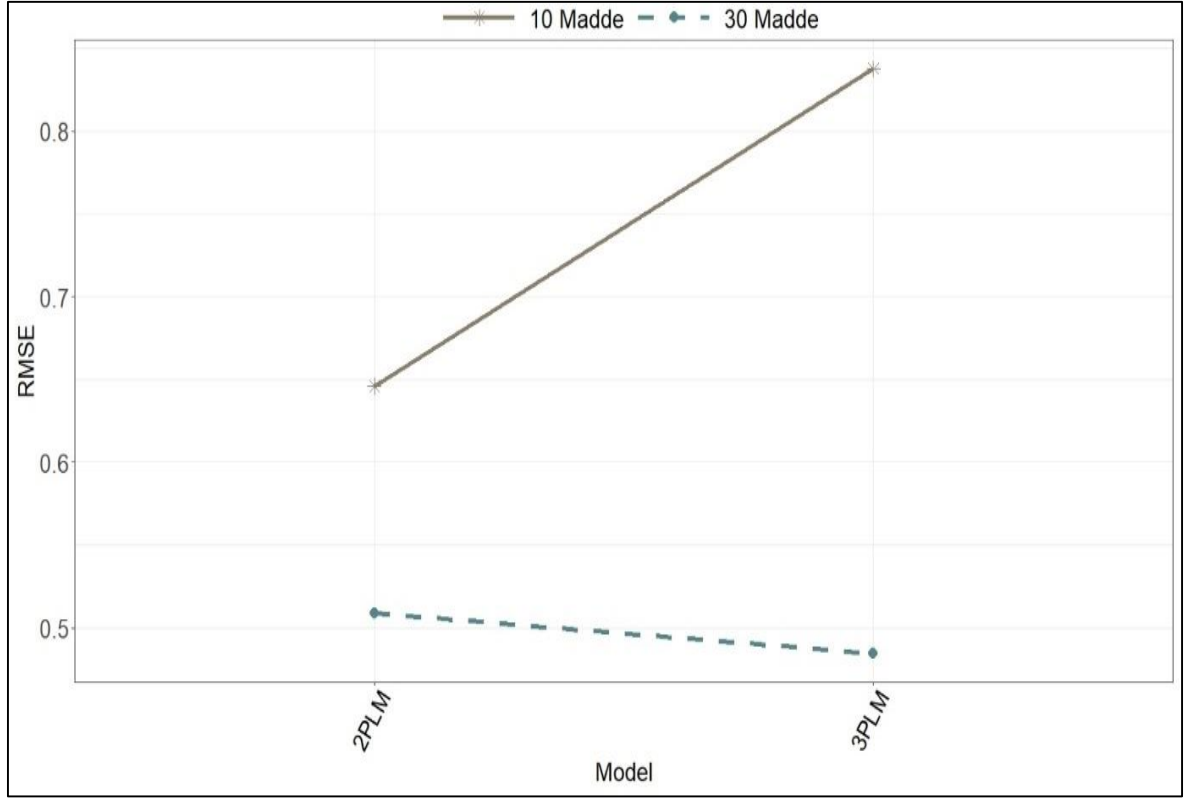
Tablo 21

Madde Faktörünün Her Bir Koşulunda Model Faktörünün Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
10 madde					
Modeller	0,440	1	0,440	110*	0,000
Hata	0,047	12	0,004		
30 madde					
Modeller	0,007	1	0,007	1,75	0,215
Hata	0,047	12	0,004		

* $p < 0,05$

Tablo 21 incelendiğinde 10 madde koşulunda model faktörü ile olan etkileşimlerin $F(1,12) = 110$; $p < 0,05$ anlamlı olduğu, 30 madde koşulunda model faktörü ile olan etkileşimlerin $F(1,12) = 1,75$; $p > 0,05$ anlamlı olmadığı sonucuna ulaşılmıştır. Etkileşimler arasındaki ortalama farklılıkları faktör düzeylerinde görebilmek için model ve madde faktörlerinin etkileşimi sonucu elde edilen RMSE değerlerine ait grafik Şekil 17’de sunulmaktadır:



Şekil 17. Madde ve model faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Şekil 17 incelendiğinde 10 madde koşulunda 2PL için elde edilen RMSE değerinin, 3PL modelde madde ayırt edicilik parametreleri için hesaplanan RMSE değerlerine göre daha düşük olduğu görülmektedir. 30 madde koşulunda ise 3PL modelde madde ayırt edicilik parametreleri için hesaplanan RMSE değerinin 2PL için hesaplanan RMSE değerine göre daha düşük olduğu görülmektedir. Şekilde modellere ait eğrilerin eğimi dikkate alındığında 10 maddeli durumdan 30 maddeli duruma geçerken madde ayırt edicilik parametre kestirimi için hesaplanan RMSE değerlerinin 3PL için daha fazla değişim gösterdiği söylenebilir.

Faktörlerin ikili etkileşimleri incelendiğinde örneklem ve sınıf faktörlerinin ikili etkileşimlerinin $F(1,12) = 8,141$; $p < 0,05$ ve $\eta_p^2 = 0,404$ değerleri ile anlamlı ve yüksek etki büyüklüğüne sahip olduğu görülmektedir. İkili etkileşimlerinin anlamlı bulunması sonucu örneklem faktörünün her bir düzeyinin sınıf faktörünün düzeyinde istatistiksel olarak anlamlılığının incelendiği Basit etkilere ait Tablo 22 aşağıda verilmektedir:

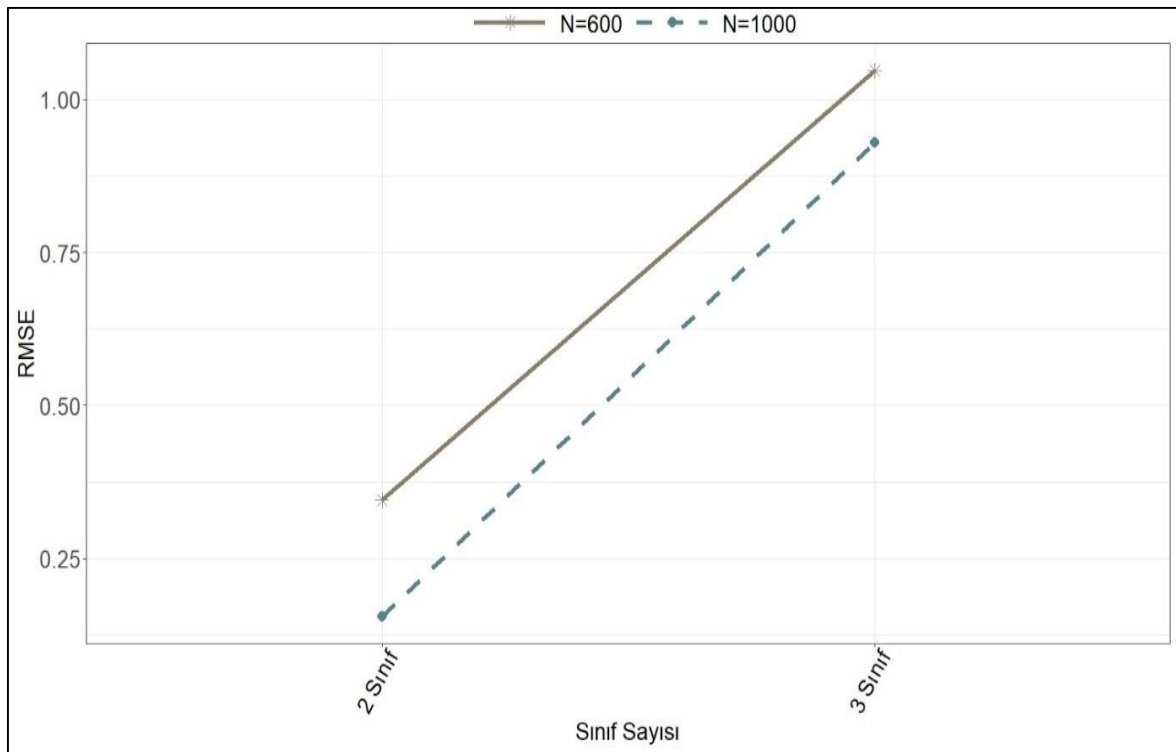
Tablo 22

Sınıf Faktörünün Her Bir Koşulunda Örneklem Faktörünün Basit Etki ANOVA Sonuçları

Koşullar	Kareler Toplamı	Serbestlik Derecesi	Ortalama Kare	F	p
2 sınıf					
Örneklem	0,432	1	0,432	108*	0,00
Hata	0,047	12	0,004		
3 sınıf					
Örneklem	0,164	1	0,164	41*	0,00
Hata	0,047	12	0,004		

*p<0,05

Tablo 22 incelendiğinde 2 sınıf koşulunda örneklem faktörü ile olan etkileşimlerin $F(1,12)=108$; $p<0,05$ anlamlı ve 3 sınıf koşulunda örneklem faktörü ile olan etkileşimlerin $F(1,12)=41$; $p<0,05$ anlamlı olduğu sonucuna ulaşılmıştır. Etkileşimler arasındaki ortalama farklılıkları faktör düzeylerinde görebilmek için sınıf ve örneklem faktörlerinin etkileşimi sonucu elde edilen RMSE değerlerine ait grafik Şekil 18’de sunulmaktadır:

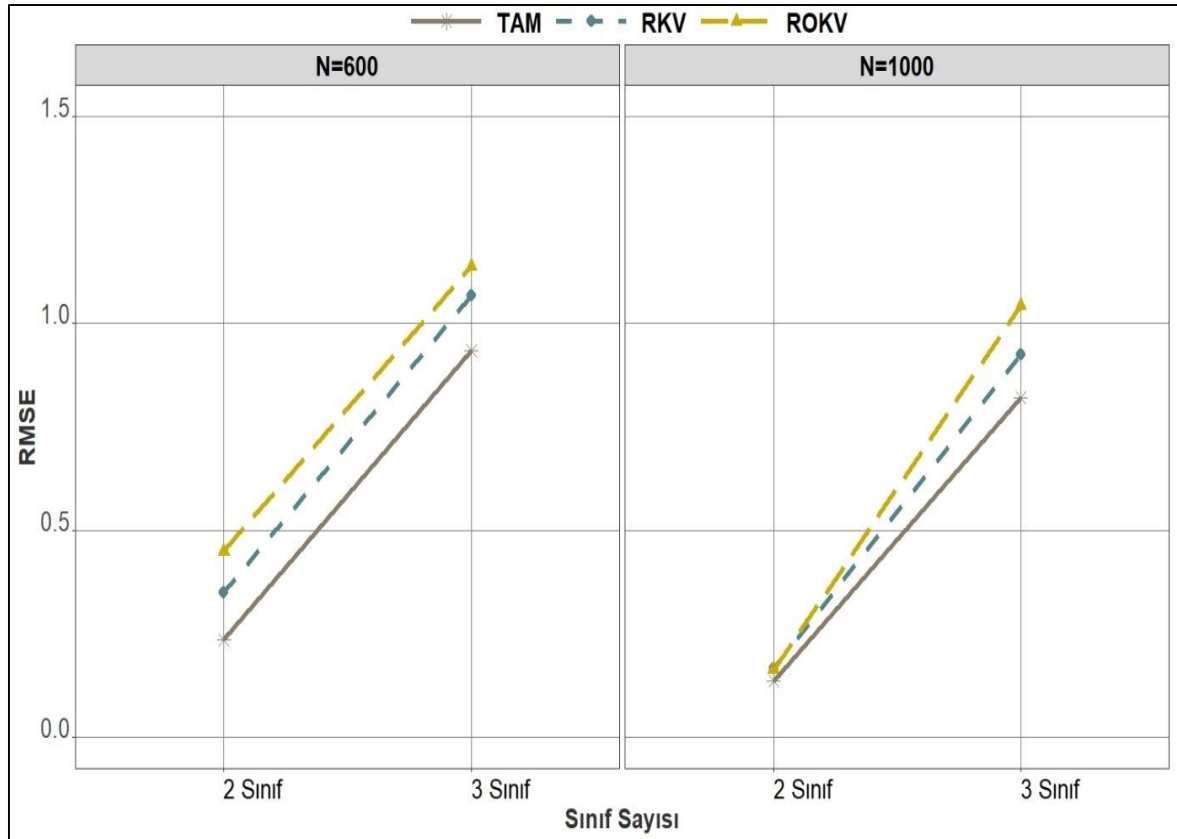


Şekil 18. Örneklem ve sınıf faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafik

Şekil 18 incelendiğinde 2 sınıf ve 1000 örneklem büyüklüğü etkileşiminde madde ayırt edicilik parametreleri kestirimi için elde edilen ortalama RMSE değerinin, 2 sınıf ve 600

örneklem büyüklüğü etkileşiminde hesaplanan ortalama RMSE değerlerine göre daha düşük olduğu görülmektedir. 3 sınıf koşulunda ise 1000 örneklem büyüklüğü etkileşiminde madde ayırt edicilik parametreleri için hesaplanan ortalama RMSE değerinin 600 örneklem büyüklüğü etkileşimi için hesaplanan ortalama RMSE değerine göre daha düşük olduğu görülmekte ancak 2 sınıf ve 1000 örneklem büyüklüğü etkileşimi için elde edilen ortalama RMSE değerine göre ise oldukça yüksek olduğu görülmektedir. Sınıf sayısı arttıkça ortalama RMSE değeri de artmakta; örneklem büyüklüğü arttıkça ise ortalama RMSE değerinin azaldığı gözlenmiştir.

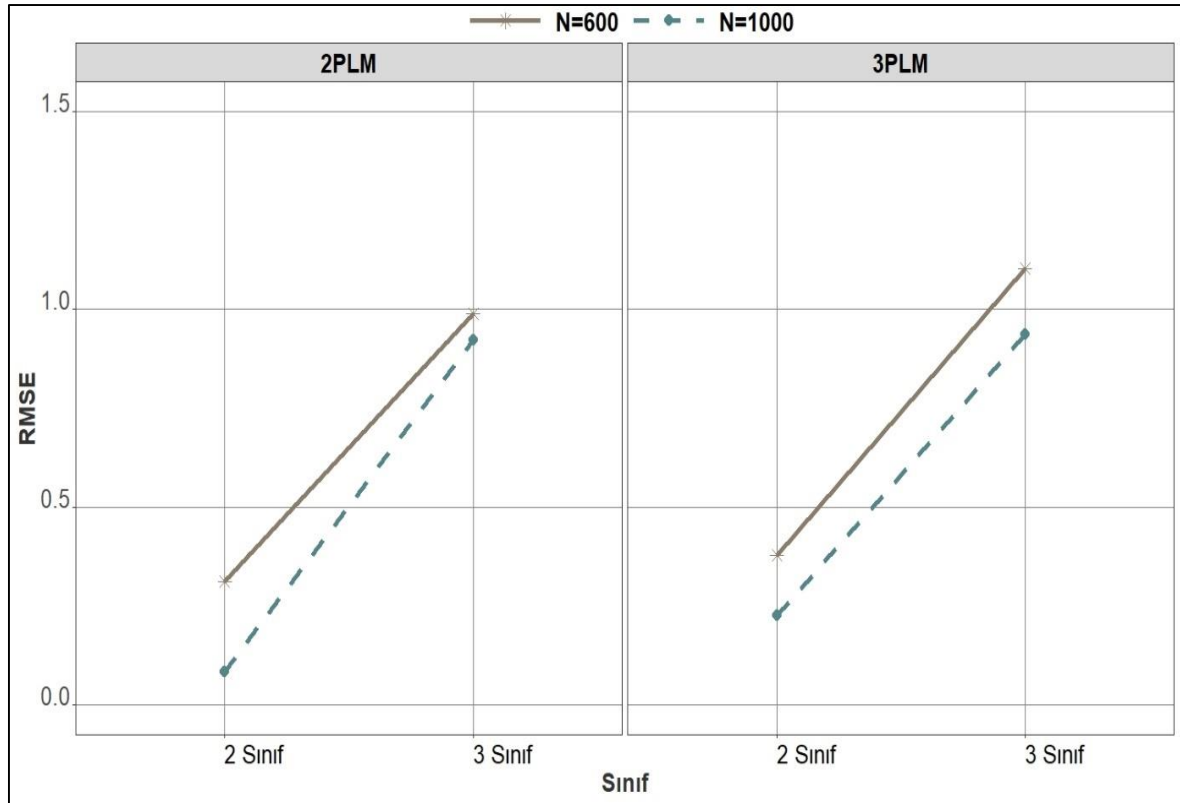
Faktörlerin üçlü etkileşimleri incelendiğinde kayıp veri türü, örneklem ve sınıf faktörlerinin ikili etkileşimleri için $F(2,12) = 5,738$; $p < 0,05$ ve $\eta_p^2 = 0,489$ değerleri ile anlamlı ve yüksek etki büyüklüğüne sahip olduğu görülmektedir. Kayıp veri türü, örneklem ve sınıf faktörlerinin etkileşimi sonucu elde edilen RMSE değerlerine ait grafik Şekil 19’da sunulmaktadır:



Şekil 19. Kayıp veri türü, örneklem ve sınıf faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafikler

Şekil 19 incelendiğinde 2 sınıf ve 3 sınıf koşullarının 600 örneklem büyüklüğü faktörü ile etkileşimlerinde, tam veri için en düşük RMSE değerinin, ROKV içinse en yüksek RMSE değerinin elde edildiği görülmektedir. 2 sınıf koşulu ile 1000 örneklem büyüklüğü faktörü etkileşimlerinde, tam veri için en düşük RMSE değeri elde edilirken, RKV ve ROKV etkileşimlerinde ise birbirine yakın ortalama RMSE değerleri elde edilmiştir. 3 sınıf koşulunun 1000 örneklem büyüklüğü etkileşiminde tam veri için en düşük ortalama RMSE değerine ulaşılmıştır. 3 sınıf koşulunun 1000 örneklem büyüklüğü etkileşiminde RKV için elde edilen ortalama RMSE değerinin ise tam veri etkileşiminden daha yüksek ROKV etkileşiminden ise daha düşük olduğu tespit edilmiştir.

Faktörlerin üçlü etkileşimleri incelendiğinde örneklem, sınıf ve model faktörlerinin etkileşimleri için $F(1,12) = 11,774$; $p < 0,05$ ve $\eta_p^2 = 0,495$ değerleri ile anlamlı ve yüksek etki büyüklüğüne sahip olduğu görülmektedir. Bu faktörlerinin etkileşimi sonucu elde edilen RMSE değerlerine ait grafik Şekil 20’de sunulmaktadır:



Şekil 20.Örneklem, sınıf ve model faktörlerinin etkileşimi sonucu madde ayırt edicilik parametresi için elde edilen ortalama RMSE değerlerine ait grafikler

Şekil 20 incelendiğinde Karma 2PL modellerin 2 sınıf koşulunun 1000 örneklem büyüklüğü etkileşiminde elde edilen ortalama RMSE değerinin 600 örneklem büyüklüğü etkileşimine göre daha düşük olduğu; 3 sınıf koşulunun 600 ve 1000 örneklem büyüklüğü etkileşimleri içinse elde edilen ortalama RMSE birbirine yakın olduğu görülmektedir. Karma 3PL modellerin 2 ve 3 sınıf koşullarında 1000 örneklem büyüklüğü etkileşimlerinde, madde ayırt edicilik parametre değerleri için kestirilen ortalama RMSE değerlerinin, 600 örneklem büyüklüğü etkileşiminde elde edilen RMSE değerlerine göre daha düşük olduğu belirlenmiştir. Sonuç olarak kayıp veri türü, madde, model ve sınıf faktörlerinin anlamlı etki büyüklüğü değerlerine sahip olması Karma MTK modellerinde madde ayırt ediciliği parametre kestiriminde seçilen faktörlerin önemli birer faktör olduğunu göstermektedir.

Madde Ayırt Edicilik Parametrelerinden Elde Edilen Ortalama Yanlılık Değerleri İçin ANOVA Sonuçları

ANOVA analizleri yapılmadan önce normallik varsayımı test edilmiştir. Çarpıklık ve basıklık katsayıları incelendiğinde çarpıklık katsayısı 0,897 ve standart hatası 0,246; basıklık katsayısı ise 1,078 ve standart hatası 0,488 olarak bulunmuştur. Yanlılık değerlerinin normal dağılım göstermesi için yapılan uygun veri dönüştürme işlemleri sonucu veri dağılımının normalleşmediği tespit edilmiştir. Verilerin basıklık ve çarpıklık katsayılarının -1 ile +1 değerleri arasında olması ve varyans analizinin normal dağılım varsayımının ihlaline karşı güçlü bir yöntem olmasından dolayı yanlılık değerlerinin farklı koşullar altındaki değişiminin incelenmesi için ANOVA kullanımının uygun olacağına karar verilmiştir (Rutherford, 2011). Madde ayırt edicilik değerleri için yanlılık değerleri alınarak gerçekleştirilen ANOVA analiz sonuçları Tablo 23'te sunulmaktadır. Tablo 23'te tür, kayıp verilerin türünü; madde, verideki madde sayısını; örneklem, örneklem büyüklüğünü; sınıf, sınıf sayısını; yüzde, kayıp verilerin toplam yüzdesini; model, kullanılan Karma MTK modelini temsil etmektedir.

Tablo 23

Madde Ayırt Edicilik Parametrelerinden Elde Edilen Yanlılık Değerleri İçin ANOVA Sonuçları

Kaynak	η^2	η_p^2	sd	Kareler Toplamı	Ortalama kare	F	p
Tür	0,018	0,413	2	0,005	0,002	4,217*	0,041
Madde	0,170	0,866	1	0,042	0,042	77,717*	0,000
Örneklem	0,001	0,020	1	0,000	0,000	0,245	0,630
Sınıf	0,044	0,624	1	0,011	0,011	19,880*	0,001
Yüzde	0,008	0,237	1	0,002	0,002	3,725	0,078
Model	0,057	0,683	1	0,014	0,014	25,814*	0,000
Tür*madde	0,031	0,544	2	0,008	0,004	7,147*	0,009
Tür*örneklem	0,005	0,154	2	0,001	0,001	1,092	0,367
Tür*sınıf	0,048	0,646	2	0,012	0,006	10,928*	0,002
Tür*yüzde	0,003	0,100	2	0,001	0,000	0,667	0,531
Tür*model	0,019	0,423	2	0,005	0,002	4,405*	0,037
Madde*örneklem	0,008	0,222	1	0,002	0,002	3,431	0,089
Madde*sınıf	0,224	0,895	1	0,056	0,056	102,261*	0,000
Madde*yüzde	0,001	0,029	1	0,000	0,000	0,361	0,559
Madde*model	0,024	0,475	1	0,006	0,006	10,875*	0,006
Örneklem*sınıf	0,020	0,433	1	0,005	0,005	9,174*	0,010
Örneklem*yüzde	0,002	0,076	1	0,001	0,001	0,989	0,340
Örneklem*model	0,000	0,018	1	0,000	0,000	0,223	0,645
Sınıf*yüzde	0,003	0,110	1	0,001	0,001	1,490	0,246
Sınıf*model	0,022	0,461	1	0,006	0,006	10,266*	0,008
Yüzde*model	0,001	0,042	1	0,000	0,000	0,522	0,484
Tür*madde*sınıf	0,047	0,642	2	0,012	0,006	10,773*	0,002
Tür*sınıf*model	0,021	0,441	2	0,005	0,003	4,738*	0,030
Madde*örneklem*sınıf	0,016	0,379	1	0,004	0,004	7,329*	0,019
Hata	0,026		12	0,007	0,001		

* $p < 0,05$

Kayıp veri türü faktörünün ana etkisinin $F(2,12)=4,217$; $p < 0,05$ ve $\eta^2 = 0,018$ değerlerine dayalı olarak anlamlı ve düşük düzey bir etkiye sahip olduğu söylenebilir. Madde faktörünün ana etkisinin $F(1,12)= 77,717$; $p < 0,05$ ve $\eta^2 = 0,17$ değerlerine göre anlamlı ve yüksek düzey bir etki büyüklüğü değerine; sınıf faktörünün $F(1,12)= 19,880$; $p < 0,05$ ve $\eta^2 = 0,044$ anlamlı ve orta düzey etki büyüklüğü değerine ve model faktörünün ana etkisinin $F(1,12)= 25,814$; $p < 0,05$ ve $\eta^2 = 0,057$ anlamlı ve orta düzey etki büyüklüğü değerlerine sahip olduğu görülmektedir. Örneklem ve kayıp veri yüzdesi faktörlerinin ana etkileri ise anlamlı bulunmamıştır

Faktörlerin ikili etkileşimleri incelendiğinde ise kayıp veri türü ile madde faktörlerinin etkileşim etkileri için $F(2,12) = 7,147$; $p < 0,05$ ve $\eta_p^2 = 0,544$; kayıp veri türü ile sınıf faktörlerinin etkileşim etkileri için $F(2,12) = 10,928$; $p < 0,05$ ve $\eta_p^2 = 0,646$; kayıp veri türü ile model faktörlerinin etkileşim etkileri için $F(2,12) = 4,405$; $p < 0,05$ ve $\eta_p^2 = 0,423$ değerleri ile anlamlı ve yüksek etki büyüklüğü değerlerinin elde edildiği görülmektedir. Bu sonuçlar ortalama RMSE değerleri alınarak yapılan ANOVA sonuçları ile benzerlik göstermektedir.

Faktörlerin ikili etkileşimleri incelendiğinde madde ile sınıf faktörlerinin etkileşim etkileri için $F(1,12) = 102,261$; $p < 0,05$ ve $\eta_p^2 = 0,895$; madde ile model faktörlerinin etkileşim etkileri için $F(1,12) = 10,875$; $p < 0,05$ ve $\eta_p^2 = 0,475$ değerleri ile anlamlı ve yüksek etki büyüklüğü değerlerinin elde edildiği görülmektedir. Örneklem ile sınıf faktörlerinin etkileşim etkileri için $F(1,12) = 9,174$; $p < 0,05$ ve $\eta_p^2 = 0,433$ etkileşim etkileri için de anlamlı ve yüksek etki büyüklüğü değerleri bulunmuştur. Bu sonuçlar ortalama RMSE değerleri alınarak yapılan ANOVA sonuçları ile benzerlik göstermektedir.

Faktörlerin üçlü etkileşimleri incelendiğinde kayıp veri türü, madde ve sınıf faktörlerinin etkileşimleri için $F(2,12) = 10,773$; $p < 0,05$ ve $\eta_p^2 = 0,642$; kayıp veri türü, sınıf ve model faktörlerinin etkileşimleri için $F(2,12) = 4,738$; $p < 0,05$ ve $\eta_p^2 = 0,441$; madde, örneklem ve sınıf faktörlerinin etkileşimleri için $F(1,12) = 10,773$; $p < 0,05$ ve $\eta_p^2 = 0,379$ değerleri ile anlamlı ve yüksek etki büyüklüğü değerlerine sahip olduğu görülmektedir. Ortalama RMSE değerleri alınarak yapılan ANOVA sonuçlarında kayıp veri türü, örneklem ve sınıf faktörleri ile örneklem, sınıf ve model faktörlerinin etkileşimleri anlamlı bulunmuştur. Ortalama yanlılık değerleri ile desteklenen sonuçlarda madde, kayıp veri türü, örneklem, sınıf ve model faktörlerinin madde ayırt edicilik parametresinin kestiriminde önemli bir faktör olduğu söylenebilir.

Şans Parametrelerine Ait Bulgular

Bu kısımda öncelikli olarak araştırma kapsamında incelenen simülasyon koşullarına (kayıp veri türü, madde sayısı, örneklem büyüklüğü, sınıf sayısı, kayıp veri yüzdesi, kullanılan Karma MTK modeli) göre şans parametresi için hesaplanan ortalama RMSE ve yanlılık değerleri sunulmuştur. Daha sonra şans parametresi için hesaplanan ortalama RMSE ve yanlılık değerlerinin simülasyon koşullarında nasıl farklılaştığını incelemek amacıyla ANOVA analizi yapılmıştır. Şans parametresi sadece Karma 3PL MTK modellerinde yer

aldığı için bu kısımda Karma 3PL için sonuçlar yer almaktadır. 3PL210P60010 gösteriminde 3 parametrelili model (3PL), Sınıf sayısı (2), kayıp veri oranı (10P), örneklem büyüklüğü (600), madde sayısı (10) ile ifade edilmektedir. Şans parametre değerleri için elde edilen ortalama RMSE ve yanlılık değerleri Tablo 24’te verilmiştir:

Tablo 24

Şans Parametre Değerleri İçin Elde Edilen Ortalama RMSE ve Yanlılık Değerleri

	Tam Veri		Rastgele Kayıp Veri (RKV)		Rastgele Olmayan Kayıp Veri (ROKV)	
	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE
3PL210P60010	0,0169	0,0767	0,0017	0,0768	0,0170	0,0768
3PL210P60030	0,0059	0,0464	0,0059	0,0462	0,0059	0,0464
3PL210P100010	0,0171	0,0771	0,0172	0,0775	0,0173	0,0781
3PL210P100030	0,0060	0,0469	0,0060	0,0466	0,0060	0,0465
3PL220P60010	0,0169	0,0767	0,0163	0,0765	0,0169	0,0777
3PL220P60030	0,0059	0,0464	0,0059	0,0462	0,0061	0,0463
3PL220P100010	0,0175	0,0787	0,0176	0,0788	0,0177	0,0791
3PL220P100030	0,0060	0,0469	0,0060	0,0466	0,0058	0,0468
3PL310P60010	0,0103	0,0593	0,0104	0,0595	0,0106	0,0600
3PL310P60030	0,0043	0,0385	0,0043	0,0386	0,0043	0,0397
3PL310P100010	0,0167	0,061	0,0100	0,0583	0,0110	0,0618
3PL310P100030	0,0040	0,0381	0,0042	0,0393	0,0041	0,0392
3PL320P60010	0,0103	0,0592	0,0095	0,0597	0,0110	0,0615
3PL320P60030	0,0041	0,0391	0,0040	0,0384	0,0041	0,0391
3PL320P100010	0,0108	0,0614	0,0109	0,06183	0,0103	0,0617
3PL320P100030	0,0040	0,03922	0,0040	0,03926	0,0041	0,0397

Tablo 24 incelendiğinde şans parametresi için madde sayısı arttıkça ortalama RMSE değerlerinin azaldığı görülmektedir. Şans parametreleri için ortalama RMSE değerleri, madde güçlüğü ve ayırt edicilik parametreleri için elde edilen ortalama RMSE değerlerinden daha düşük değerlere sahip olduğu görülmektedir. Bunun nedeni, şans parametre değerleri sıfır ile bir arasında değer alırken madde ayırt edicilik ve güçlük parametre değerlerinin daha büyük mutlak değerler alabilmesidir. Şans parametresi için ortalama yanlılık değerleri incelendiğinde ise madde sayısı arttıkça ortalama yanlılık değerlerinin azaldığı görülmektedir. Kişi sayısı arttıkça ise ortalama yanlılık değerlerinin değişmediği görülmektedir.

Şans Parametrelerinden Elde Edilen Ortalama RMSE Değerleri İçin ANOVA Sonuçları

ANOVA analizleri yapılmadan önce normallik ve varyansların homojenliği varsayımları test edilmiştir. Yapılan normallik testi sonucu çarpıklık katsayısı ve standart hatası 0,389(0,343) ve basıklık katsayısı ve standart hatası-1,337(0,674) olarak bulunmuştur. Ortalama RMSE değerlerinin normal dağılım göstermesi için yapılan uygun veri dönüştürme işlemleri sonucu veri dağılımının normalleşmediği tespit edilmiştir. Varyans analizinin normal dağılım varsayımının ihlaline karşı güçlü bir yöntem olmasından dolayı yanlılık değerlerinin farklı koşullar altındaki değişiminin incelenmesi için ANOVA kullanımının uygun olacağına karar verilmiştir (Rutherford, 2011). Şans parametreleri için ortalama RMSE değerleri alınarak gerçekleştirilen ANOVA sonuçları Tablo 25’te sunulmaktadır. Tablo 25’te tür, kayıp verilerin türünü; madde, verideki madde sayısını; örneklem, örneklem büyüklüğünü; sınıf, sınıf sayısını; yüzde, kayıp verilerin toplam yüzdesini; model, kullanılan Karma MTK modelini temsil etmektedir.

Tablo 25

Şans Parametrelerinden Elde edilen RMSE değerleri için ANOVA Sonuçları

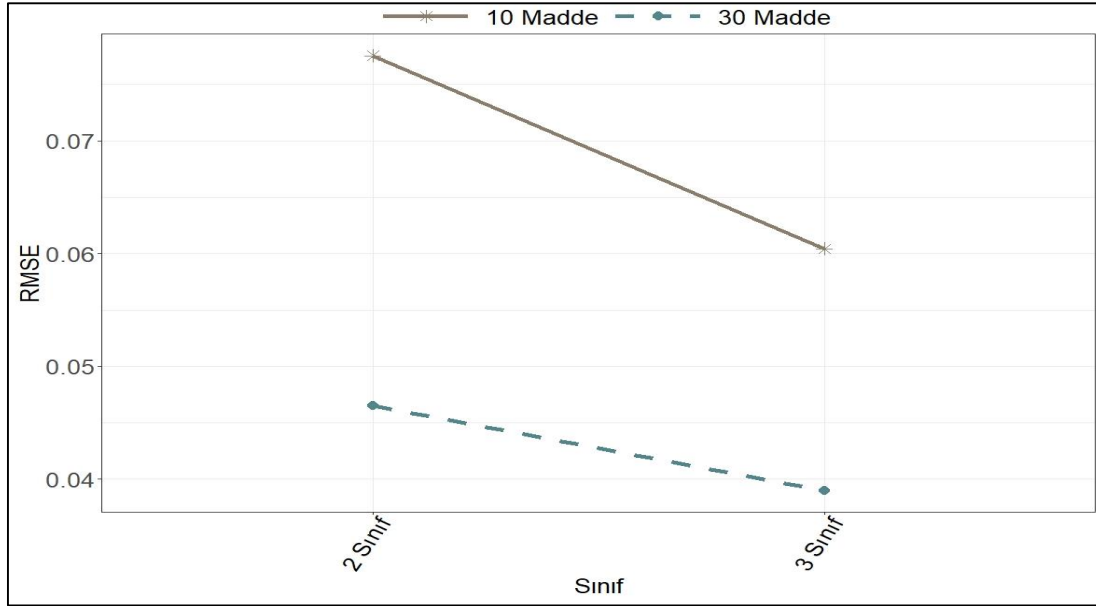
Kaynak	η^2	η_p^2	sd	Kareler Toplamı	Ortalama kare	F	p
Tür	0,001	0,951	2	0,000	0,000	19,316	0,049
Madde	0,794	1,000	1	0,008	0,008	42941,408*	0,000
Örneklem	0,001	0,64	1	0,000	0,000	54,039*	0,018
Sınıf	0,173	1,000	1	0,002	0,002	9336,320*	0,000
Yüzde	0,000	0,769	1	0,000	0,000	6,671	0,123
Madde*sınıf	0,029	0,999	1	0,000	0,000	1547,526*	0,001
Hata	0,000		2	0,000	0,000		

* $p<0,05$

Madde faktörünün ana etkisinin $F(1,2)= 42941,408$; $p<0,05$ ve $\eta^2= 0,794$ ve sınıf faktörünün ana etkisinin $F(1,2)= 9336,320$; $p<0,05$ ve $\eta^2= 0,173$ anlamlı ve yüksek etki büyüklüğü değerine ve örneklem faktörünün ana etkisinin $F(1,2)= 54,039$; $p<0,05$ ve $\eta^2= 0,001$ anlamlı ve düşük etki büyüklüğü değerlerine sahip olduğu görülmektedir. Kayıp veri türü ve kayıp veri yüzdesi faktörlerinin ana etkisi ise anlamlı bulunmamıştır

İkili etkileşimler incelendiğinde ise sadece madde ve sınıf faktörlerinin $F(1,2)= 1547,526$; $p<0,05$ ve $\eta_p^2=0,999$ etkileşim etkileri için anlamlı ve yüksek etki büyüklüğü değerlerine

ulaşılmıştır. Madde ve sınıf faktörlerinin etkileşimi sonucu elde edilen RMSE değerlerine ait grafik Şekil 21’de sunulmuştur:



Şekil 21. Madde ve sınıf faktörlerinin etkileşimi sonucu elde edilen şans parametreleri için kestirilen RMSE değerlerine ait grafikler

Şekil 21 incelendiğinde 2 sınıf ve 3 sınıf koşulunun 10 madde etkileşiminde daha yüksek RMSE değeri, 30 madde etkileşiminde ise şans parametresi kestiriminde daha düşük RMSE değerleri elde edilmiştir. Şans parametresinin kestiriminde madde ve sınıf sayısı arttıkça RMSE değerinin azaldığı görülmektedir. Sonuç olarak madde ve sınıf faktörlerinin şans parametresinin kestiriminde önemli birer faktör oldukları söylenebilir.

Şans Parametrelerinden Elde Edilen Ortalama Yanlılık Değerleri İçin ANOVA Sonuçları

ANOVA analizleri yapılmadan önce normallik ve varyansların homojenliği varsayımları test edilmiştir. Yapılan normallik testi sonucu çarpıklık katsayısı ve standart hatası 0,554 (0,343) ve basıklık katsayısı ve standart hatası -1,207(0,674) olarak bulunmuştur. Varyans analizinin normal dağılım varsayımının ihlaline karşı güçlü bir yöntem olmasından dolayı yanlılık değerlerinin farklı koşullar altındaki değişiminin incelenmesi için ANOVA kullanımının uygun olacağına karar verilmiştir (Rutherford, 2011). Şans parametreleri için ortalama yanlılık değerleri alınarak gerçekleştirilen ANOVA sonuçları Tablo 26’da sunulmaktadır.

Tablo 26

Şans Parametrelerinden Elde edilen Yanlılık değerleri için ANOVA Sonuçları

Kaynak	η^2	η_p^2	df	Kareler Toplamı	Ortalama kare	F	p
Tür	0,014	0,508	2	0,000	0,000	1,032	0,492
Madde	0,667	0,979	1	0,001	0,001	95,301*	0,010
Örneklem	0,011	0,440	1	0,000	0,000	1,570	0,337
Sınıf	0,106	0,884	1	0,000	0,000	15,172	0,060
Yüzde	0,001	0,080	1	0,000	0,000	0,174	0,717
Hata	0,014		2	0,000	0,000		

* $p<0,05$

Tablo 26 incelendiğinde sadece madde faktörünün ana etkisinin $F(1,2)= 95,301$; $p<0,05$ ve $\eta^2= 0,667$ değerlerine dayalı olarak anlamlı ve yüksek düzey bir etkiye sahip olduğu söylenebilir.

Sınıflama Doğruluğu

Sınıflama doğruluğu için üretilen veri setleri içerisinde her bir kişi sınıf olasılık değerlerine göre örtük sınıflardan birine atanmaktadır. Bu veri setleri kullanılarak kestirim yapıldığında elde edilen sınıf üyeliklerinin, gerçek sınıf üyelikleri ile karşılaştırılması gerekmektedir (Sen & Cohen, 2020). Örneğin 1000 kişilik bir veri setinde 980 kişinin kestirilen sınıflama doğruluğu gerçek sınıfı ile aynı ise o veri setinde sınıflama doğruluğunun %98 olduğu söylenebilmektedir. Tablo 27’de Karma Rasch model için elde edilen sınıflama doğruluğu yüzdeleri verilmektedir:

Tablo 27

Karma Rasch model için elde edilen sınıflama doğruluğu yüzdeleri

	Koşullar	TAM	RKV	ROKV
1	R210P60010	98,62	85,03	81,06
2	R210P60030	93,01	87,48	83,07
3	R210P100010	98,62	85,26	84,58
4	R210P100030	99,02	87,81	82,93
5	R220P60010	98,60	79,98	72,03
6	R220P60030	89,02	76,59	72,93
7	R220P100010	98,65	75,61	75,70
8	R220P100030	88,01	76,56	62,44
9	R310P60010	85,26	64,49	64,87
10	R310P60030	83,32	74,43	75,30
11	R310P100010	88,58	67,72	71,61
12	R310P100030	84,92	75,85	75,54
13	R320P60010	77,66	61,75	59,25
14	R320P60030	83,70	63,81	61,86
15	R320P100010	78,60	65,20	60,83
16	R320P100030	85,26	64,44	62,54

Ele alınan değişkenlere göre Karma Rasch modelde tam veri durumu için daha yüksek sınıflama doğruluğu yüzdeleri elde edilmiştir. Tam veri koşulunda en yüksek sınıflama doğruluğu yüzdesine 2 sınıf, %10 kayıp veri yüzdesi, 30 madde ve 1000 örneklem büyüklüğü için ulaşılrken (99,02) en düşük sınıflama doğruluğu yüzdesine ise 3 sınıf, %20 kayıp veri yüzdesi, 10 madde ve 600 örneklem büyüklüğü için ulaşılmıştır (77,67). Kayıp veri türüne göre ise RKV ve ROKV örüntüsü koşullarında daha düşük sınıflama doğruluğu yüzdeleri elde edilmiştir. RKV örüntüsü koşulunda en yüksek sınıflama doğruluğu yüzdesine 2 sınıf, %10 kayıp veri yüzdesi, 30 madde ve 1000 örneklem büyüklüğü için ulaşılrken (87,81) en düşük sınıflama doğruluğu yüzdesine ise 3 sınıf, %20 kayıp veri yüzdesi, 10 madde ve 600 örneklem büyüklüğü için ulaşılmıştır (61,75). ROKV örüntüsü koşulunda en yüksek sınıflama doğruluğu yüzdesine 2 sınıf, %10 kayıp veri yüzdesi, 10 madde ve 1000 örneklem büyüklüğü için ulaşılrken (84,58) en düşük sınıflama doğruluğu yüzdesine ise 3 sınıf, %20 kayıp veri yüzdesi, 10 madde ve 600 örneklem büyüklüğü için ulaşılmıştır (59,25). Tablo 28’de ise Karma 2PL model için elde edilen sınıflama doğruluğu yüzdeleri verilmektedir:

Tablo 28

Karma 2 parametrelı model için elde edilen sınıflama dođruluđu yüzdeleri

	Koşullar	TAM	RKV	ROKV
1	2PL210P60010	98,36	86,85	72,54
2	2PL210P60030	98,71	87,35	75,44
3	2PL210P100010	98,34	86,71	72,79
4	2PL210P100030	99,02	87,41	76,06
5	2PL220P60010	98,26	72,18	65,02
6	2PL220P60030	97,03	77,01	72,69
7	2PL220P100010	98,33	72,97	64,90
8	2PL220P100030	86,02	77,08	72,90
9	2PL310P60010	83,70	75,34	68,44
10	2PL310P60030	83,92	83,47	79,83
11	2PL310P100010	86,51	76,82	72,01
12	2PL310P100030	84,92	84,18	80,15
13	2PL320P60010	83,33	71,42	67,10
14	2PL320P60030	93,70	72,14	76,08
15	2PL320P100010	91,66	74,86	68,05
16	2PL320P100030	93,24	71,75	77,83

Tablo 28 incelendiđinde, alınan deđiřkenlere göre Karma 2PL modelde tam veri durumu için daha yüksek sınıflama dođruluđu yüzdeleri elde edilmiřtir. Tam veri koşulunda en yüksek sınıflama dođruluđu yüzdesine 2 sınıf, %10 kayıp veri yüzdesi, 30 madde koşullarının 1000 (99,02) örneklem büyüklüklerinin kombinasyonlarında ulařılırken en düşük sınıflama dođruluđu yüzdesine ise 3 sınıf, %20 kayıp veri yüzdesi, 10 madde ve 600 örneklem büyüklüđu için ulařılmıřtır (83,32). Kayıp veri türüne göre ise ROKV örüntüsü koşulunda daha düşük sınıflama dođruluđu yüzdeleri elde edilmiřtir. RKV örüntüsü koşulunda en yüksek sınıflama dođruluđu yüzdesine 2 sınıf, %10 kayıp veri yüzdesi, 30 madde ve 1000 örneklem büyüklüđu için ulařılırken (87,41) en düşük sınıflama dođruluđu yüzdesine ise 3 sınıf, %20 kayıp veri yüzdesi, 10 madde ve 600 örneklem büyüklüđu için ulařılmıřtır (71,42). ROKV örüntüsü koşulunda en yüksek sınıflama dođruluđu yüzdesine 3 sınıf, %10 kayıp veri yüzdesi, 30 madde ve 1000 örneklem büyüklüđu için ulařılırken (80,154) en düşük sınıflama dođruluđu yüzdesine ise 3 sınıf, %20 kayıp veri yüzdesi, 10 madde koşullarının 600 (65,02) ve 1000 (64,90) kombinasyonlarında ulařılmıřtır. Tablo 29’da ise Karma 3PL model için elde edilen sınıflama dođruluđu yüzdeleri verilmektedir:

Tablo 29

Karma 3 parametrelı model için elde edilen sınıflama dođruluđu yúzdeleri

	Koşullar	TAM	RKV	ROKV
1	3PL210P60010	98,51	83,71	77,93
2	3PL210P60030	98,77	88,79	79,98
3	3PL210P100010	96,90	83,87	77,92
4	3PL210P100030	99,58	89,30	81,33
5	3PL220P60010	95,66	81,20	68,47
6	3PL220P60030	97,01	72,30	73,18
7	3PL220P100010	96,22	67,25	68,26
8	3PL220P100030	99,60	75,88	70,46
9	3PL310P60010	84,37	61,45	61,54
10	3PL310P60030	83,32	70,96	68,79
11	3PL310P100010	86,76	62,20	65,25
12	3PL310P100030	84,92	74,38	71,22
13	3PL320P60010	80,21	62,64	66,75
14	3PL320P60030	81,80	62,28	68,33
15	3PL320P100010	84,65	66,06	69,67
16	3PL320P100030	82,53	63,53	70,46

Tablo 29 incelendiđinde, ele alınan deđiřkenlere göre Karma 3PL modelde de tam veri durumu için daha yüksek sınıflama dođruluđu yúzdeleri elde edilmiřtir. Tam veri koşulunda en yüksek sınıflama dođruluđu yúzdesine 2 sınıf, 30 madde ve 1000 örneklem büyüklüđu koşullarının %10 (99,58) ve %20 (99,60) kayıp veri yúzdelerinin kombinasyonlarında ulařılmıřtır. En düşük sınıflama dođruluđu yúzdesine ise 3 sınıf, %20 kayıp veri yúzdesi, 10 madde ve 600 örneklem büyüklüđu için ulařılmıřtır (80,213). Kayıp veri türüne göre ise rastgele ve rastgele olmayan kayıp veri örüntüsü koşullarında daha düşük sınıflama dođruluđu yúzdeleri elde edilmiřtir. Rastgele ve rastgele olmayan kayıp veri örüntüsü koşullarında en yüksek sınıflama dođruluđu yúzdesi 2 sınıf, %10 kayıp veri yúzdesi, 30 madde ve 1000 örneklem büyüklüđu koşulunda, en düşük sınıflama dođruluđu yúzdesine ise 3 sınıf, %10 kayıp veri yúzdesi, 10 madde ve 600 örneklem büyüklüđu koşulunda elde edilmiřtir.

Gerçek Veri

Bu çalışmada gerçek veri olarak Türkiye TIMSS 2015 8. Sınıf fen alt testinin verilerinden rastgele 600 ve 1000 kişilik örneklemeler çekilerek madde sayısı sabit tutularak tam, RKV ve ROKV türleri ile %10 ve %20 kayıp veri yúzdelerine uygun olacak şekilde veri setleri oluşturulmuřtur. 600 kişilik örneklemi elde edebilmek için fen alt testinin 4. ve 5.

kitapçıkları birleştirilerek 880 kişilik örneklem oluşturulmuştur. Her iki kitapçıkta bulunan ortak madde sayısı 14 olduğu için madde sayısı 14 olarak sabitlenmiştir. Tam veri elde edebilmek için kayıp verilere ait satırlar silinerek 600 kişilik tam veri örnekleme oluşturulmuştur. Veriye en iyi uyum sağlayan modeli tespit etmek için sırasıyla 2, 3, 4 ve 5 sınıflı Karma MTK modelleri (Rasch, 2PL, 3PL) ile analizler gerçekleştirilmiştir. Karma MTK modelleri Mplus programı kullanılarak analiz edilmiştir. Bu araştırmada parametre kestirimleri için MLR (Maximum Likelihood-ML) yani Tam Bilgi En Çok Olabilirlik (Full Information Maximum Likelihood-FIML) kestirim yöntemi kullanılmıştır. TIMSS 2015 8.sınıf fen alt test verileri için en uygun Karma MTK modelini belirlemek için yapılan analizler sonucunda Türkiye verisi için her bir Karma MTK modeline göre elde edilen bilgi kriter indeksleri incelenmiş ve Tablo 30’da verilmiştir:

Tablo 30.

TIMSS 8.sınıf fen alt testinin 4.ve 5. Kitapçığından elde edilen veriler için Karma MTK model veri uyum indeks değerleri

Model	Karma Rasch		Karma-2PL		Karma-3PL	
	AIC	BIC	AIC	BIC	AIC	BIC
2-Sınıf	10186,882	10314,393	10070,564	10321,189	10168,024	10357,092
3-Sınıf	10157,589	10351,054	10072,199	10388,778	10170,023	10363,488
4-Sınıf	10139,021	10398,439	10074,236	10522,881	10172,025	10369,886
5-Sınıf	10130,675	10456,048	10077,176	10648,777	10174,024	10376,283

Tablo 30’da BIC değerleri incelendiğinde verinin en iyi iki sınıflı Karma Rasch modele uyum sağladığı görülmektedir. Seçilen model için kestirilen madde parametre değerleri ve standart hataları Tablo 31’de verilmiştir.

Tablo 31.

İki Sınıflı Karma Rasch model için tam veriden elde edilen madde parametre değerleri

Maddeler	α	β_1	Standart hata	β_2	Standart hata
Madde1	1,000	-1,070	0,267	0,367	0,233
Madde2	1,000	-0,630	0,266	1,044	0,280
Madde3	1,000	-0,776	0,301	1,063	0,270
Madde4	1,000	0,542	0,163	0,226	0,228
Madde5	1,000	-3,033	0,560	-0,534	0,465
Madde6	1,000	-1,757	0,437	1,107	0,520
Madde7	1,000	-1,766	0,251	-0,680	0,226
Madde8	1,000	-1,292	0,268	0,657	0,377
Madde9	1,000	-0,900	0,170	-1,269	0,233
Madde10	1,000	0,551	0,152	0,510	0,219
Madde11	1,000	0,186	0,151	1,351	0,470
Madde12	1,000	-0,290	0,183	0,288	0,214
Madde13	1,000	-2,164	0,316	0,098	0,430
Madde14	1,000	-0,567	0,178	-0,118	0,205

Karma Rasch model için her iki sınıf için elde edilen madde güçlük değerleri (β_1 ve β_2) ile madde ayırt edicilik değerleri (α) Tablo 31’de bulunmaktadır. Sınıf 1’deki bireyler için kestirilen madde güçlük parametre değerlerinin -3,033 (Madde5) ile 0,551 (Madde10) arasında değiştiği yani maddelerin Sınıf 1’deki öğrencilere genellikle daha kolay geldiği ve bu sınıftaki öğrencilerin daha yüksek performans sergiledikleri ifade edilebilir. Madde güçlük parametre kestirimlerinin standart hatalarına ilişkin değerlerin ise 0,560 (Madde5) ile 0,151 (Madde14) arasında değiştiği görülmüştür. Sınıf 2’deki bireyler için kestirilen madde güçlük parametre değerlerinin -1,269 (Madde9) ile 1,351 (Madde11) arasında değiştiği yani maddelerin Sınıf 2’deki öğrencilere genellikle daha zor geldiği ve bu sınıftaki öğrencilerin daha düşük performans sergiledikleri söylenebilir. Madde güçlük parametre kestirimlerinin standart hatalarına ilişkin değerlerin ise 0,205 (Madde14) ile 0,520 (Madde6) arasında değiştiği görülmüştür. İki sınıflı Karma Rasch modeli için elde edilen olasılık değerleri ise sırasıyla Sınıf 1 için 0,66 olarak, Sınıf 2 içinse 0,34 olarak elde edilmiştir. TIMSS 2015 8.sınıf fen alt testi tam verisinden üretilen %10 RKV için kestirilen madde parametre değerleri ve standart hataları Tablo 32’de verilmiştir.

Tablo 32.

İki Sınıflı Karma Rasch model %10 RKV için elde edilen madde parametre değerleri

Maddeler	α	β_1	Standart hata	β_2	Standart hata
Madde1	1,000	-1,136	0,273	0,392	0,281
Madde2	1,000	-0,715	0,275	0,929	0,293
Madde3	1,000	-0,864	0,323	0,992	0,281
Madde4	1,000	0,624	0,207	0,205	0,254
Madde5	1,000	-3,780	0,689	-0,775	0,402
Madde6	1,000	-1,822	0,481	0,993	0,470
Madde7	1,000	-1,837	0,317	-0,699	0,234
Madde8	1,000	-1,312	0,290	0,495	0,311
Madde9	1,000	-0,830	0,189	-1,284	0,254
Madde10	1,000	0,536	0,176	0,475	0,217
Madde11	1,000	0,122	0,193	1,269	0,371
Madde12	1,000	-0,435	0,234	0,267	0,219
Madde13	1,000	-2,234	0,378	0,042	0,395
Madde14	1,000	-0,517	0,184	-0,120	0,218

Karma Rasch model 2 sınıf 600 örneklem büyüklüğü ve %10 RKV için her iki sınıf için elde edilen madde güçlük değerleri (β_1 ve β_2) ile madde ayırt edicilik değerleri (α) Tablo 32’de bulunmaktadır. Sınıf 1’deki kestirilen madde güçlük parametre değerlerinin -3,780 (Madde5) ile 0,624 (Madde4) arasında değiştiği yani maddelerin Sınıf 1’deki öğrencilere genellikle daha kolay geldiği ve bu sınıftaki öğrencilerin daha yüksek performans sergiledikleri söylenebilir. Madde güçlük parametre kestirimlerinin standart hatalarına ilişkin değerlerin ise 0,122 (Madde11) ile 0,689 (Madde5) arasında değiştiği görülmüştür. Sınıf 2’deki kestirilen madde güçlük parametre değerlerinin -1,284 (Madde9) ile 0,993 (Madde6) arasında değiştiği yani maddelerin Sınıf 2’deki öğrencilere genellikle daha zor geldiği ve bu sınıftaki öğrencilerin daha düşük performans sergiledikleri ifade edilebilir. Madde güçlük parametre kestirimlerinin standart hatalarına ilişkin değerlerin ise 0,218 (Madde14) ile 0,470 (Madde 6) arasında değiştiği görülmüştür. Madde güçlük parametre kestirimlerine ait standart hata değerlerinin tam veri durumunda daha yüksek olduğu tespit edilmiştir. İki sınıflı Karma Rasch modeli için elde edilen olasılık değerleri ise sırasıyla Sınıf 1 için 0,58 olarak, Sınıf 2 içinse 0,42 olarak elde edilmiştir. TIMSS 2015 8.sınıf fen alt testi tam verisinden üretilen %20 RKV için kestirilen madde parametre değerleri ve standart hataları Tablo 33’te verilmiştir.

Tablo 33.

İki Sınıflı Karma Rasch model %20 RKV için elde edilen madde parametre değerleri

Maddeler	α	β_1	Standart hata	β_2	Standart hata
Madde1	1,000	-1,057	0,434	0,333	0,220
Madde2	1,000	-0,548	0,293	0,786	0,332
Madde3	1,000	-0,907	0,461	1,073	0,368
Madde4	1,000	0,558	0,192	0,061	0,335
Madde5	1,000	-3,613	0,953	-0,650	0,499
Madde6	1,000	-1,970	0,613	1,437	1,148
Madde7	1,000	-1,817	0,310	-0,922	0,264
Madde8	1,000	-1,214	0,404	0,507	0,359
Madde9	1,000	-0,798	0,237	-1,463	0,305
Madde10	1,000	0,392	0,184	0,748	0,268
Madde11	1,000	0,089	0,179	1,111	0,552
Madde12	1,000	-0,342	0,187	0,160	0,268
Madde13	1,000	-2,065	0,372	-0,016	0,548
Madde14	1,000	-0,588	0,205	-0,256	0,232

Karma Rasch model RKV için her iki sınıf için elde edilen madde güçlük değerleri (β_1 ve β_2) ile madde ayırt edicilik değerleri (α) Tablo 32’de bulunmaktadır. Sınıf 1’deki kestirilen madde güçlük parametre değerlerinin -3,613 (Madde5) ile 0,558 (Madde4) arasında değiştiği yani maddelerin Sınıf 1’deki öğrencilere genellikle daha kolay geldiği ve bu sınıftaki öğrencilerin daha yüksek performans sergiledikleri söylenebilir. Madde güçlük parametre kestirimlerinin standart hatalarına ilişkin değerlerin ise 0,179 (Madde11) ile 0,953 (Madde5) arasında değiştiği görülmüştür. Sınıf 2’deki kestirilen madde güçlük parametre değerlerinin -1,462 (Madde9) ile 1,437 (Madde6) arasında değiştiği yani maddelerin Sınıf 2’deki öğrencilere genellikle daha zor geldiği ve bu sınıftaki öğrencilerin daha düşük performans sergiledikleri ifade edilebilir. Madde güçlük parametre kestirimlerinin standart hatalarına ilişkin değerlerin ise 0,220 (Madde1) ile 1,148 (Madde6) arasında değiştiği görülmüştür. Madde güçlük parametre kestirimlerine ait standart hata değerlerinin tam veri durumunda daha düşük, RKV durumunda ise daha yüksek olduğu tespit edilmiştir. İki sınıflı Karma Rasch modeli için elde edilen olasılık değerleri ise sırasıyla Sınıf 1 için 0,65 olarak, Sınıf 2 içinse 0,35 olarak elde edilmiştir. TIMSS 2015 8.sınıf fen alt testi tam verisinden üretilen %10 ROKV için kestirilen madde parametre değerleri ve standart hataları Tablo 34’te verilmiştir.

Tablo 34.

İki Sınıflı Karma Rasch model %10 ROKV için elde edilen madde parametre değerleri

Maddeler	α	β_1	Standart hata	β_2	Standart hata
Madde1	1,000	-0,924	0,213	0,532	0,249
Madde2	1,000	-0,422	0,208	1,233	0,259
Madde3	1,000	-0,621	0,224	1,385	0,297
Madde4	1,000	0,923	0,172	0,345	0,232
Madde5	1,000	-3,784	0,648	-0,382	0,323
Madde6	1,000	-1,642	0,293	1,356	0,461
Madde7	1,000	-1,651	0,232	-0,509	0,218
Madde8	1,000	-1,164	0,216	0,872	0,319
Madde9	1,000	-0,712	0,172	-1,173	0,244
Madde10	1,000	0,839	0,166	0,767	0,227
Madde11	1,000	0,476	0,165	1,653	0,331
Madde12	1,000	-0,065	0,163	0,364	0,227
Madde13	1,000	-1,910	0,251	0,140	0,291
Madde14	1,000	-0,550	0,157	-0,146	0,207

Tablo 34’te Karma Rasch model %10 ROKV için her iki sınıf için elde edilen madde güçlük değerleri (β_1 ve β_2) ile madde ayırt edicilik değerleri (α) bulunmaktadır. Sınıf 1’deki kestirilen madde güçlük parametre değerlerinin -3,784 (Madde5) ile 0,923 (Madde4) arasında değiştiği yani maddelerin Sınıf 1’deki öğrencilere genellikle daha kolay geldiği ve bu sınıftaki öğrencilerin daha yüksek performans sergiledikleri söylenebilir. Madde güçlük parametre kestirimlerinin standart hatalarına ilişkin değerlerin ise 0,157 (Madde14) ile 0,648 (Madde5) arasında değiştiği görülmüştür. Sınıf 2’deki kestirilen madde güçlük parametre değerlerinin -1,173 (Madde9) ile 1,653 (Madde11) arasında değiştiği yani maddelerin Sınıf 2’deki öğrencilere genellikle daha zor geldiği ve bu sınıftaki öğrencilerin daha düşük performans sergiledikleri ifade edilebilir. Madde güçlük parametre kestirimlerinin standart hatalarına ilişkin değerlerin ise 0,207 (Madde14) ile 0,461 (Madde6) arasında değiştiği görülmüştür. Madde güçlük parametre kestirimlerine ait standart hata değerlerinin tam veri durumunda daha düşük, ROKV durumunda ise daha yüksek olduğu tespit edilmiştir. Ayrıca RKV ve ROKV koşulları için birbirine yakın hata değerleri elde edilmiştir. İki sınıflı Karma Rasch modeli için elde edilen olasılık değerleri ise sırasıyla Sınıf 1 için 0,63 olarak, Sınıf 2 içinse 0,37 olarak elde edilmiştir. TIMSS 2015 8.sınıf fen alt testi tam verisinden üretilen %20 ROKV için kestirilen madde parametre değerleri ve standart hataları Tablo 35’te verilmiştir.

Tablo 35.

İki Sınıflı Karma Rasch model %20 ROKV için elde edilen madde parametre değerleri

Maddeler	α	β_1	Standart hata	β_2	Standart hata
Madde1	1,000	-0,999	0,281	0,868	0,278
Madde2	1,000	-0,277	0,251	1,368	0,290
Madde3	1,000	-0,493	0,267	1,384	0,301
Madde4	1,000	1,441	0,280	0,682	0,252
Madde5	1,000	-3,950	0,875	-0,551	0,321
Madde6	1,000	-1,747	0,431	1,272	0,443
Madde7	1,000	-1,691	0,341	-0,375	0,216
Madde8	1,000	-1,117	0,310	0,905	0,280
Madde9	1,000	-0,292	0,220	-1,153	0,241
Madde10	1,000	1,215	0,216	1,171	0,226
Madde11	1,000	0,906	0,219	1,685	0,325
Madde12	1,000	0,112	0,216	0,686	0,210
Madde13	1,000	-1,892	0,294	0,089	0,291
Madde14	1,000	-0,200	0,196	0,231	0,202

Tablo 35’te Karma Rasch model %20 ROKV için her iki sınıf için elde edilen madde güçlük değerleri (β_1 ve β_2) ile madde ayırt edicilik değerleri (α) bulunmaktadır. Sınıf 1’deki kestirilen madde güçlük parametre değerlerinin -3,950 (Madde5) ile 1,441 (Madde4) arasında değiştiği yani maddelerin Sınıf 1’deki öğrencilere genellikle daha kolay geldiği ve bu sınıftaki öğrencilerin daha yüksek performans sergiledikleri söylenebilir. Madde güçlük parametre kestirimlerinin standart hatalarına ilişkin değerlerin ise 0,196 (Madde14) ile 0,875 (Madde5) arasında değiştiği görülmüştür. Sınıf 2’deki kestirilen madde güçlük parametre değerlerinin -1,153 (Madde9) ile 1,685 (Madde11) arasında değiştiği yani maddelerin Sınıf 2’deki öğrencilere genellikle daha zor geldiği ve bu sınıftaki öğrencilerin daha düşük performans sergiledikleri ifade edilebilir. Madde güçlük parametre kestirimlerinin standart hatalarına ilişkin değerlerin ise 0,202 (Madde14) ile 0,443 (Madde6) arasında değiştiği görülmüştür. Madde güçlük parametre kestirimlerine ait standart hata değerlerinin tam veri durumunda daha düşük, ROKV durumunda ise daha yüksek olduğu tespit edilmiştir. Ayrıca RKV ve ROKV koşulları için birbirine yakın hata değerleri elde edilmiştir. İki sınıflı Karma Rasch modeli için elde edilen olasılık değerleri ise sırasıyla Sınıf 1 için 0,54 olarak, Sınıf 2 içinse 0,46 olarak elde edilmiştir.

BÖLÜM 5

TARTIŞMA, SONUÇ VE ÖNERİLER

Bu bölümde araştırma sonucunda elde edilen bulgular alan yazındaki bulgular ile tartışılmış ve elde edilen sonuçlar doğrultusunda araştırmacılara ileriye dönük yapılacak çalışmalar konusunda önerilerde bulunulmuştur.

Tartışma ve Sonuç

Bu araştırmada madde sayısı, örneklem büyüklüğü, örtük sınıf sayısı, kayıp veri türü ve kayıp veri yüzdesi ve Karma MTK modelleri değişkenleri ele alınarak kestirilen madde parametrelerinin ve ele alınan değişkenlerin etkileşim etkilerinin nasıl değiştiği incelenmiştir. Parametre doğrulanması için hata kareler ortalaması karekökü (RMSE) ve yanlılık (BIAS) değerleri incelenmiştir. Ayrıca kestirilen örtük sınıf üyelikleriyle gerçek sınıf üyelikleri karşılaştırılarak sınıflama doğruluğu yüzdeleri elde edilmiştir.

Karma MTK modelleri için madde güçlük ve ayırt edicilik parametrelerinin kestiriminde örneklem büyüklüğü ve madde sayısı arttıkça daha düşük RMSE değerleri elde edilirken sınıf sayısı arttıkça RMSE değerinin arttığı görülmüştür. Şans parametresinin kestiriminde ise örneklem büyüklüğü, madde ve sınıf sayısı arttıkça RMSE değerinin azaldığı görülmektedir. Literatürde Karma MTK modelleri için örneklem büyüklüğü, madde sayısı ve sınıf sayısı faktörlerinin parametre doğrulanması üzerindeki etkisini araştıran birçok çalışmada da (Alexeev vd., 2011; Cho vd.,2013; Finch & French, 2012; Li vd., 2009; Preinerstorfer & Formann, 2012; Revuelta, 2010; Sen vd., 2016) madde güçlük ve madde ayırt edicilik parametre kestiriminde örneklem büyüklüğü ve madde sayısı arttıkça RMSE değerinin azaldığı, sınıf sayısı arttıkça ise RMSE değerlerinin arttığı gözlenmiştir. Şans parametresinin kestiriminde ise örneklem büyüklüğü, madde sayısı ve sınıf sayısı arttıkça RMSE değerlerinin azaldığı görülmüştür (Finch & French, 2012; Sen vd., 2016). Bu araştırmanın sonuçlarının literatürde bulunan sonuçlarla benzerlik gösterdiği söylenebilir.

Sınıf sayısı arttığında, sınıflarda yer alan bireylerin sayısının azalmasından dolayı hata değerlerinin artmasının doğal olduğu ifade edilebilir (Finch & French, 2012).

Karma MTK modelleri için madde güçlük ve madde ayırt edicilik parametre kestirimlerinde tam veri durumları için daha düşük RMSE değerleri, kayıp veri türünün RKV ve ROKV durumlarında ise daha yüksek RMSE değerleri elde edilmiştir. Kayıp veri yüzdesinin %20 olduğu durumlarda ise daha yüksek RMSE değerlerine ulaşılmıştır. Literatürde Karma MTK modelleri ile kayıp veri türü ve yüzdesi koşullarının ele alındığı bir araştırmada da benzer sonuçlar bulunmuştur (Lee,2012). Kayıp veri türlerine göre şans parametresi için birbirine yakın RMSE değerleri elde edilmesi ise Finch (2008) in çalışması ile benzerlik göstermektedir. Finch (2008)'de MTK modelleri için şans parametresinin kestiriminde tam, RKV ve ROKV koşullarının ortalama RMSE değerleri düşük ve birbirine oldukça yakın çıkmıştır. Kayıp veri üretme mekanizması Finch (2008)'in çalışmasındaki gibi kırılmalara ayrılarak üretildiği için ve kayıp veri ataması yapılmadan FIML yöntemi ile kayıp veriler analiz edildiği için şans parametresi için ortalama RMSE değerlerinin birbirine yakın çıkması doğal görünmektedir.

Madde parametrelerinin doğrulanmasında Karma 3PL modeli için elde edilen ortalama RMSE değerlerinin Karma 2PL model için elde edilen ortalama RMSE değerlerinden daha yüksek, Karma 2PL modeli için elde edilen ortalama RMSE değerlerinin Karma Rasch model için elde edilen ortalama RMSE değerlerinden ise daha yüksek olduğu görülmüştür. Madde parametrelerinin kestiriminde modelin karmaşıklığı arttıkça RMSE değerlerinin arttığı bir RMSE değerleri örüntüsü görünmektedir. Dolayısıyla Karma Rasch model için elde edilen madde parametrelerinin Karma 2PL ve Karma 3PL modelden daha az hataya sahip olduğu söylenebilir. Bu sonuçlar Cho vd., (2013) ve Sen (2014)'ün çalışmaları ile benzerlik göstermektedir. Bu çalışmalarda da araştırmacılar yaptıkları simülasyon çalışmasında iki sınıflı Karma Rasch model için daha düşük RMSE değerleri elde etmişlerdir.

Ayrıca ele alınan değişkenlerin ana etkileri ve etkileşim etkilerinin incelenmesinde istatistiksel anlamlılık için F ve p değerleri ve pratik anlamlılık için etki büyüklüğü ölçüleri olan eta-kare değerleri yorumlanmıştır. Madde güçlük ve ayırt edicilik parametrelerinin kestiriminde ele alınan tüm faktörlerin ana etkileri anlamlı bulunmuştur.Şans parametresinin kestiriminde ise sınıf, madde ve örneklem faktörlerinin ana etkileri anlamlı bulunmuştur. Ayrıca madde güçlük parametresi kestiriminde madde faktörü ile sınıf; madde faktörü ile

model faktörünün sınıf faktörü ile model faktörünün etkileşimlerinin anlamlı ve yüksek etki büyüklüğü değerlerine sahip olduğu görülmüştür. Madde ayırt edicilik parametresi kestiriminde de madde ile sınıf faktörlerinin; madde ve model faktörlerinin etkileşimleri arasında anlamlı ve yüksek etki büyüklüğü değerleri elde edilmiştir. Şans parametresinin kestiriminde ise sadece madde ve sınıf faktörlerinin etkileşim etkileri için anlamlı ve yüksek etki büyüklüğü değerlerine ulaşılmıştır. Yüksek etki büyüklüğü değerlerine sahip olan faktörlerin önemli faktör oldukları ifade edilmektedir (Sen, 2014). Dolayısıyla bu sonuçlar Karma MTK modellerinde sınıf, madde ve model faktörlerinin madde güçlük ve ayırt edicilik parametrelerinin kestiriminde, madde ve sınıf faktörlerinin ise şans parametresinin kestiriminde önemli birer faktör olduğunu desteklemektedir (Lee,2012; Sen,2014).

Ayrıca madde güçlük parametresi kestiriminde kayıp veri türü faktörü ile sınıf faktörünün; kayıp veri türü faktörü ile kayıp veri yüzdesi faktörlerinin ikili etkileşimi yüksek ve anlamlı etki büyüklüğü değerleri bulunmuştur. Madde ayırt edicilik parametresi kestiriminde ise kayıp veri türü ile sınıf faktörlerinin etkileşim etkileri için kayıp veri türü ile model faktörlerinin etkileşim etkileri için yüksek etki büyüklüğü değerleri elde edilmiştir. Dolayısıyla bu sonuçlar Karma MTK modellerinde kayıp veri türü, sınıf ve kayıp veri yüzdesi faktörlerinin madde güçlük parametrelerinin kestiriminde, kayıp veri türü, sınıf ve model faktörlerinin ise ayırt edicilik parametrelerinin önemli birer faktör olduğunu desteklemektedir.

Ayrıca madde güçlük parametresi kestiriminde örneklem faktörü ile sınıf faktörünün örneklem, sınıf ve model faktörlerinin ve madde ayırt edicilik parametresi kestiriminde örneklem ve sınıf faktörlerinin ikili etkileşimleri için, örneklem, sınıf ve model faktörlerinin etkileşim değerleri ile anlamlı ve yüksek etki büyüklüğü değerleri bulunmuştur. Dolayısıyla madde güçlük ve ayırt edicilik parametre kestirimlerinde örneklem, sınıf ve model faktörlerinin önemli birer faktör olduğu söylenebilir.

Madde güçlük parametre kestiriminden elde edilen ortalama RMSE değerlerine göre üçlü etkileşimler incelendiğinde, kayıp veri türü, madde ve sınıf faktörlerinin kayıp veri türü, sınıf ve model faktörlerinin değerlerine göre etkileşimleri anlamlı ve yüksek bulunan etkileşimlerdir. Madde ayırt edicilik parametre kestiriminden elde edilen ortalama yanlışlık değerlerine göre ise kayıp veri türü, madde ve sınıf faktörlerinin etkileşimleri için kayıp veri türü, sınıf ve model faktörlerinin etkileşimleri için anlamlı ve yüksek etkileşim etkileri

bulunmuştur. Dolayısıyla madde güçlük ve ayırt edicilik parametre kestirimlerinden elde edilen sonuçların birbirleriyle tutarlı olduğu görülmektedir.

Sonuç olarak Karma MTK modelleri için madde parametre kestiriminde ana etkilerinin ve etkileşim etkilerinin anlamlı ve yüksek etki büyüklüğü değerlerine sahip olması seçilen faktörlerin araştırmanın amacına uygun önemli faktörler olduğu ifade edilebilir.

Elde edilen sınıflama doğruluğu yüzdeleri incelendiğinde ise tüm Karma MTK modellerinde tam veri durumu için daha yüksek sınıflama doğruluğu yüzdeleri elde edilmiştir. Tüm Karma MTK modelleri için tam veri ve rastgele kayıp veri koşulunda en yüksek sınıflama doğruluğu yüzdesine 2 sınıf, %10 kayıp veri yüzdesi, 30 madde, 1000 örneklem büyüklüğü kombinasyonlarında ulaşılırken en düşük sınıflama doğruluğu yüzdesine ise 3 sınıf, %20 kayıp veri yüzdesi, 10 madde ve 600 örneklem büyüklüğü için ulaşılmıştır. RKV ve ROKV koşullarında ise tüm modeller için daha düşük sınıflama doğruluğunun elde edildiği gözlenmiştir. Sınıflama doğruluğu yüzdelere göre RKV ve ROKV türlerinden elde edilen verilerin sahte (spurious) örtük sınıflar ürettiği söylenebilir (Lee,2012).

Gerçek veri kısmında ise geniş ölçekli uluslararası değerlendirmelerden biri olan TIMSS 2015 8. Sınıf uygulamasının Türkiye verisinin fen alt testinden elde edilen verilerin heterojenliğini araştırarak verinin en iyi iki sınıflı Karma Rasch modeline uyum sağladığı tespit edilmiştir. Bu sonuç literatürde geniş ölçekli uygulamaların Karma MTK modelleri ile analiz edildiği çalışmalarla örtüşmektedir (Sen, 2014; Toker, 2016; Zhang vd., 2015). Sen (2014)'in araştırmasında TIMSS 2011 sekizinci sınıf matematik verisi için en yüksek performans gösteren ülkelerden biri olan Güney Kore ülkesinden elde edilen veriler Karma MTK modelleri ile analiz edilmiştir. Karma MTK modelleri için 18 çoktan seçmeli maddeden oluşan ve beşinci kitapçığı alan 369 birey analize dahil edilmiştir. Verinin iki sınıflı Karma Rasch modele daha iyi uyum sağladığı sonucuna ulaşılmıştır. Zhang vd. (2015) yaptıkları çalışmada matematik ve fen alanlarında öğrencilerin nasıl sınıflandırıldığına dair bilgi sağlamak amacıyla her bir alt test ve bu alt testlerin birlikte ele alınması durumlarında Karma Rasch modeli ile iki örtük sınıf bulunmuştur.

Ayrıca simülasyon verisinden elde edilen sonuçları gerçek veri üzerinden elde edilen sonuçlarla karşılaştırmak için TIMSS 2015 Türkiye 8. Sınıf fen alt testinin verisinden tam, RKV ve ROKV kayıp veri türü koşulları ile %10 ve %20 kayıp veri yüzdesi koşullarına uygun olacak şekilde veri setleri çekilerek gerçekleştirilen analiz sonucunda, madde parametre kestirimlerinin standart hatalarının tam veri koşulunda daha düşük, RKV ve

ROKV koşullarında ise daha yüksek olduğu sonucuna ulaşılmıştır. Bu bağlamda gerçek veriden elde edilen sonuçların simülasyon verisinden elde edilen sonuçlarla örtüştüğü görülmüştür. Dolayısıyla simülasyon verisinin geçerliğine de katkı sunulmuştur.

Öneriler

Çalışma kapsamında elde edilen bulgular doğrultusunda araştırmacılara sunulan öneriler aşağıda yer almaktadır:

- Bu çalışmada madde sayısı, Karma MTK modelleri, örneklem büyüklüğü, örtük sınıf sayısı, kayıp veri türü ve kayıp veri yüzdesi faktörleri dikkate alınmıştır. Elde edilen bulgular Karma MTK modelleri ile yapılan kestirimlerde madde parametrelerinin nasıl değiştiğini açıkça göstermektedir. Madde güçlük parametre kestiriminde kayıp veri türünün etkisi, sınıf sayısı ve kayıp veri yüzdesi değiştirilerek incelenebilir. Madde ayırt edicilik parametre kestiriminde kayıp veri türünün etkisi, model türü ve sınıf sayısı değiştirilerek incelenebilir.
- Madde güçlük ve ayırt edicilik parametre kestirimlerinde örneklem, sınıf, madde ve model faktörlerinin koşulları değiştirilerek; şans parametresinin kestiriminde madde ve sınıf faktörlerinin koşulları değiştirilerek Karma MTK modellerinde madde parametre kestirimlerine etkisi incelenebilir.
- Bu araştırmada madde güçlük, madde ayırt edicilik ve şans parametrelerinin üretilmesinde kullanılan değerler Li, vd (2009) çalışmasında kullanılan değerler ile sınırlıdır. Araştırmacılar farklı dağılımlar kullanarak madde parametre üretim değerlerini değiştirebilir. Ayrıca bu araştırmada yeteneğin standart normal dağılımdan rastgele çekilerek üretildiği varsayılmıştır. Araştırmacılar farklı yetenek dağılımları kullanarak bu araştırmada ele alınan faktörlerle kestirilen madde parametrelerinin nasıl değiştiğini inceleyebilir.
- Bu tezdeki simülasyon çalışmasında her bir koşul için 100 replikasyon gerçekleştirilmiştir. Araştırmacılar replikasyon sayısını değiştirerek analiz sonuçlarını yorumlayabilir.
- Bu araştırmada kayıp veri atama yöntemleri kullanılmadan FIML yöntemi ile Mplus yazılımı kullanılarak kayıp verilerin analizleri gerçekleştirilmiştir. Araştırmacılar kayıp veri atama yöntemlerini kullanarak Karma MTK modellerinde bu yöntemlerin

etkilerini inceleyebilirler. Ayrıca parametrelerin kestirimi için Mplus programındaki MLR kestirim yöntemi kullanılmıştır. Araştırmacılar Bayes gibi farklı yöntemler kullanabilir ve bu yöntemlerin karşılaştırılması yapılabilir.

- Bu araştırmada iki kategorili puanlanan maddeler ele alınarak veriler üretilmiştir. Araştırmacılar çok kategorili puanlanan maddelerle Karma MTK modelleri analizlerini gerçekleştirebilirler.

Çalışma kapsamında elde edilen bulgular doğrultusunda uygulayıcılara aşağıdaki önerilerde bulunulmuştur:

- Karma MTK modellerinde rastgele ve rastgele olmayan kayıp veri mekanizmasına sahip verilerin parametre kestiriminde RMSE ve yanlılık değerlerinin artmasına sebep olmasından dolayı daha çok veriye ulaşılması önerilmektedir.
- Madde sayısı 10 ve 30 olduğunda Karma Rasch model kullanılarak yapılan madde parametre kestirimleri daha az hataya sahip olduğu için madde sayısı 10 ve daha fazla olduğunda iki sınıflı Karma Rasch model kullanılabilir.
- Örneklem büyüklüğü arttıkça RMSE ve yanlılık değerleri düştüğü ve daha yüksek sınıflama doğruluğu yüzdeleri elde edildiği için Karma MTK modelleri için büyük örneklem tercih edilmelidir. Bu araştırmada 600 ve 1000 alınan örneklem büyüklüğü değerlerinde sınıf sayısı arttıkça RMSE değerleri arttığı gözlenmiştir. Daha büyük örneklemelerde sınıf başına düşen kişi sayısının fazla olması sebebiyle daha fazla bilgi elde edilebilmesi için büyük örneklem tercih edilmelidir.
- Örtük sınıfların varlığında MTK modelinin varsayımlarından biri olan parametrelerin değişmezliği varsayımı ihlal edilmektedir. Karma MTK modelleri ise madde parametrelerinin örtük sınıflar arasında değişmesine izin vermektedir. Öğrencilerin yeteneklerini örtük sınıflara ayırmak, araştırmacıların madde ve grup özellikleri hakkında daha güvenilir ve daha geçerli bilgiler edinmelerini sağlamaktadır. Sonuç olarak TIMSS uygulaması alt gruplar bazında kalibrasyon yapmak için Karma MTK modellerine ihtiyaç duymaktadır. Bu doğrultuda geniş ölçekli değerlendirmelerden elde edilen sonuçların geçerliğine dair doğru kanıtların toplanması elde edilen sonuçlara dayanarak eğitim programlarının yeniden düzenlenmesi, geliştirilmesi ve değerlendirilmesi için gereklidir.

KAYNAKLAR

- Adams, R.J. & Wu, M.L. (2007). The mixed-coefficients multinomial logit model: A generalized form of the Rasch model. In M. von Davier & C.H. Carstensen (Eds.), *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*. (pp.57–75). New York: Springer- Verlag.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Alexeev, N., Templin, J., & Cohen, A. S. (2011). Spurious latent classes in the mixture Rasch model. *Journal of Educational Measurement*, 48, 313–332.
- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38, 123-140.
- Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage.
- Baghaei, P., & Carstensen, C. H. (2013). Fitting the mixed Rasch model to a reading comprehension test:Identifying reader types. *Practical Assessment, Rese*, 18(5), 1–13.
- Baker, F. B., and Kim, S.-H. (2004). *Item Response Theory: Parameter estimation techniques*, 2nd Edn. New York, NY: Marcel Dekker.
- Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2001). A mixture item response for multiple choice data. *Journal of Educational and Behavioral Statistics*, 26, 381–409.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick, *Statistical theories of mental test scores* (pp.392-479). Reading, MA: Addison-Wesley.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443-459.

- Bock, R.D., & Zimowski, M.F. (1997). Multiple group IRT. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 433–448). New York, NY: Springer.
- Bolt, D. M., Cohen, A. S. & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, 39, 331-348.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345–370
- Bulut, O., & Sunbul, O. (2017). Monte Carlo simulation studies in item response theory with the R programming language. *Journal of Measurement and Evaluation in Education and Psychology*, 8(3), 266-287.
- Chen, Y., & Jiao H. (2014) Exploring the utility of background and cognitive variables in explaining latent differential item functioning: An example of the PISA 2009 reading assessment, *Educational Assessment*, 19(2), 77-96.
- Cho, S.-J., Cohen, A. S., & Kim, S.-H. (2013). Markov chain Monte Carlo estimation of a mixture item response theory model. *Journal of Statistical Computation and Simulation*, 83, 278–306.
- Cho, H. J., Lee, J., & Kingston, N. (2012). Examining the effectiveness of test accommodation using DIF and a mixture IRT model. *Applied Measurement in Education*, 25(4), 281–304.
- Cho, S.-J., Cohen, A. S., & Kim, S.-H. (2013). Markov chain Monte Carlo estimation of a mixture Rasch model. *Journal of Statistical Computation and Simulation*, 83, 278–306.
- Choi, Y. J., Alexeev, N., & Cohen, A. S. (2015). Differential item functioning analysis using a mixture 3-parameter logistic model with a covariate on the TIMSS 2007 mathematics test. *International Journal of Testing*, 15(3), 239–253.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York: Academic.
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis*. Hoboken, NJ: John Wiley & Sons.

- Christensen, K., Engelhard, G., & Salzberger, T. (2012). Ask the experts: Rasch vs Factor Analysis. *Rasch Measurement Transactions*, 26(3), 1373–1378.
- Cohen, A. S., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement*, 42, 133–148.
- Cook, L. (2006). *Practical considerations in linking scores on adapted tests*. Keynote address at the 5th International Meeting of the International Test Commission, Brussels, Belgium.
- Crocker, L., & Algina, J. (1986). *Introduction to modern test theory*. California: Thomson Learning.
- Çokluk, Ö., Şekercioğlu, G. & Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları*. Ankara: Pegem Akademi.
- De Ayala, R. J., Plake, B. S. & Impara, J. C. (2001). The impact of omitted responses on the accuracy of ability estimation in item response theory. *Journal of Educational Measurement*, 38, 213–234.
- De Ayala, R. J., Kim, S.H., Stapleton, L. M. & Dayton, C. M. (2002). Differential item functioning: A mixture distribution conceptualization. *International Journal of Testing*, 2, 243- 276.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford.
- De Ayala, R. J. & Santiago, S. Y. (2017). An introduction to mixture item response theory models. *Journal of School Psychology*, 60, 25-40.
- DeMars, C. (2010). *Item response theory*. New York: Oxford University.
- DeMars, C. E., & Lau, A. (2011). Differential item functioning detection with latent classes: how accurately detect who is responding differentially? *Educational and Psychological Measurement*, 71, 597–616.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38. Retrieved from <http://www.jstor.org/stable/2984875>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. *Quality of Life Research*. New Jersey: Lawrence Erlbaum Associates, Inc.

- Embretson, S. E. (2007). Mixed Rasch models for measurement in cognitive psychology. In M. von Davier, & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: extensions and applications* (pp. 235-253). New York: Springer Verlag.
- Enders, C. K. & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(3), 430-457.
- Enders, C. K. (2004). The Impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and Psychological Measurement*, 64(3), 419-436.
- Feinberg, R. R., & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice*, 35(2), 36–49.
- Finch, W. H., & Pierson, E. E. (2011). A mixture IRT analysis of risky youth behavior. *Frontiers in Quantitative Psychology*, 2, 1-10.
- Finch, W. H., & French, B. F. (2012). Parameter estimation with mixture item response theory models: A monte carlo comparison of maximum likelihood and bayesian methods. *Journal of Modern Applied Statistical Methods*, 11(1), 167-178.
- Frick H, Strobl C., Leisch, F. & Zeileis, A. (2012). Flexible Rasch mixture models with package psychomix. *Journal of Statistical Software*, 48(7), 1–25.
- Gamst, G., Meyers, L. S., & Guarino, A. J. (2008). *Analysis of variance designs: A conceptual and computational approach with SPSS and SAS*. Cambridge University, New York.
- Gilbert, N. (2004). *Agent-based social simulation: dealing with complexity*. <http://www.agsm.edu.au/bobm/teaching/SimSS/ABSS-dealingwithcomplexity-1-1.pdf> sayfasından ulaşılmıştır.
- Glück, J., & Spiel, C. (2007). Studying development via Item Response Model: A wide range of potential uses. In M. von Davier, & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 281-292). New York: Springer Verlag.

- Gollini, I., & Murphy, T. B. (2014). Mixture of latent trait analyzers for model-based clustering of categorical data. *Statistics and Computing*, 24(4), 569–588.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2), 215-231.
- Haberman, S. J. (1979). *Analysis of qualitative data. vol. 2, new developments*. London:Academic.
- Hagenaars, J. A. (1990). *Categorical longitudinal data: Log-linear panel, trend, and cohort analysis*. Thousand Oaks, CA: Sage.
- Hair, J. F., Anderson, R. E., Tatham, R., L., & Black, W. C. (1998). *Multivariate data analysis* (Fifth edition). United States: Prentice-Hall, Inc.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. New York: Springer Science & Business Media, LLC.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Harris, D. (1989). Comparison of 1-, 2-, and 3-Parameter IRT Models. *Educational Measurement: Issues and Practice*, 8(1), 35–41.
- Hong, S. (2007). Mixed Rasch modeling of the self-rating depression scale. *Educational and Psychological Measurement*, 67, 280-299.
- Howell, D. C. (2014). *Fundamental statistics for the behavioral sciences*. Belmont, CA:Cengage Learning.
- Jeon, M., & De Boeck, P. (2018). An analysis of an item-response strategy based on knowledge retrieval. *Behavior Research Methods*, 1–23.
- Jilke, S., Meuleman, B., & Van de Walle, S. (2015). We need to compare, but how? Measurement equivalence in comparative public administration. *Public Administration Review*, 75(1), 36–48.
- Kelderman, H., & Macready, G. B. (1990). The use of loglinear models for assessing differential item functioning across manifest and latent examinee groups. *Journal of Educational Measurement*, 27, 307-327.

- Kirk, R. E. (2008). *Statistics an introduction* (Fifth edition). United States: Thomson Higher Education.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices*. New York: Springer.
- Kreiner, S., & Christensen, K. B. (2007). Validity and objectivity in health-related scales: Analysis by graphical loglinear Rasch models. In M. von Davier, & C. H. Carstensen (Eds.). *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 329-346). New York: Springer Verlag.
- Kreiner, S., & Christensen, K. B. (2014). Analyses of model fit and robustness . A new look at the PISA. *Psychometrika*, 79(2), 210–231.
- Kutscher, T., Eid, M., & Crayen, C. (2019). Sample size requirements for applying mixed polytomous item response models: Results of a Monte Carlo simulation study. *Frontiers in Psychology*, 10, 2494.
- LaRoche, S., Joncas, M., & Foy, P. (2015). Sample design in TIMSS 2015. *Methods and Procedures in TIMSS*.
- Lazarsfeld, P. F. (1950). Logical and mathematical foundation of latent structure analysis. In *Measurement and Prediction* (SA Stouffer *et al.*, eds.) 4, 362-412. Princeton, NJ:Princeton University.
- Lee, S. (2012). *The Impact of Missing Data On The Dichotomous Mixture IRT Models* Doctoral Dissertation, The University of Georgia, USA.
- Li, F., Cohen, A. S., Kim, S.-H., & Cho, S.-J. (2009). Model selection methods for mixture dichotomous MTK models. *Applied Psychological Measurement*, 33, 353-373.
- Linden, W. J & Glas, G. A. W. (2002.) *Computerized Adaptive Testing: theory and practice*. USA: Kluwer Academic.
- Linden, W.J. & Hambleton, R.K. (1997). *Handbook of modern response theory*, New York, USA.
- Liu, H., Liu, Y., & Li, M. (2018). Analysis of process data of PISA 2012 computer-based problem solving: Application of the modified multilevel mixture IRT model. *Frontiers in Psychology*, 9.

- Lord, F. M. (1980). *Applications of item response theory to practical testing problems. Journal of Chemical Information and Modeling* (Vol. 53). New Jersey: Lawrence Erlbaum Associates.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Maij-de Meij, A. M., Kelderman, H., & van der Flier, H. (2008). Fitting a mixture item response theory model to personality questionnaire data: Characterizing latent classes and investigating possibilities for improving prediction. *Applied Psychological Measurement*, 32, 611-631.
- Martin, M. O., Mullis, I. V. S., & Hooper, M. (2016). *Methods and procedures in TIMSS 2015*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center. Zugriff am (Vol. 21).
- McCutcheon, A.L (1987). *Latent class analysis*, Beverly Hills: Sage.
- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Meiser, T., Stern, E., & Langeheine, R. (1998). Latent change in discrete data: Unidimensional, multidimensional, and mixture distribution Rasch models for the analysis of repeated observations. *Methods of Psychological Research Online*, 3, 75-93.
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational and Behavioral Statistics*, 7(2), 105–118.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23,13–23.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.
- Mislevy, R., & Huang, C., W. (2007). Measurement models as narrative structures. In M. von Davier, & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 15-35). New York: Springer Verlag.

- Mislevy, R. J., & Verhelst, N. D. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, 55, 195-215.
- Muthén, B. O. (2004). *Mplus technical appendices* (3rd ver.). Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (Eighth Edition). Los Angeles, CA: Muthén & Muthén.
- Muthén, B., & Asparouhov, T. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors*, 31, 1050–1066.
- Nagin, D. (2005). *Group-based modeling of development*. Harvard University.
- Norouzian, R., & Plonsky, L. (2018). Eta- and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. *Second Language Research*, 34, 257–271.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569.
- Oliveri, M. E., & von Davier, M. (2011). Investigation of model fit and score scale comparability in international assessments. *Psychological Test and Assessment Modeling*, 53(3), 315–333.
- Oliveri, M.E., & von Davier, M. (2014) Toward increasing fairness in score scale calibrations employed in international large-scale assessments, *International Journal of Testing*, 14(1), 1-21.
- Oliveri, M. E., Ercikan, K., Zumbo, B. D., & Lawless, R. (2014). Uncovering substantive patterns in student responses in international large-scale assessments—Comparing a latent class to a manifest DIF approach. *International Journal of Testing*, 14(3), 265-287.
- Park, Y. S., Lee, Y.-S., & Xing, K. (2016). Investigating the Impact of Item Parameter Drift for Item Response Theory Models with Mixture Distributions. *Frontiers in Psychology*, <https://doi.org/10.3389/fpsyg.2016.00255>
- Perline, R., Wright, B.D., & Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement*, 3, 237–255.

- Preinerstorfer, D., & Formann, A. K. (2012). Parameter recovery and model selection in mixed Rasch models. *British Journal of Mathematical and Statistical Psychology*, 65, 251–262.
- Rasch, G. (1960). Probabilistic models for some intelligence and achievement tests. *Copenhagen, Denmark: Danish Institute for Educational Research*, 1960.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer.
- Richardson, J.T. (2011) Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review* 6, 135–47.
- Rijkes, C. P. M., & Kelderman, H. (2006). Latent-response rasch models for strategy shifts in problem-solving processes. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (p. 311-329). New York: Springer.
- Rost, J. (1990). Rasch Models in Latent Classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14(3), 271–282.
- Rost, J., Carstensen, C., & von Davier, M. (1997). Applying the mixed Rasch model to personality questionnaires. In J. Rost, & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 324-332). Munster, Germany: Waxmann.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.
- Rutherford, A. (2011). *ANOVA and ANCOVA: A GLM approach*. Hoboken, N.J: Wiley.
- Rutkowski, L., & Rutkowski, D. (2016). A call for a more measured approach to reporting and interpreting PISA results. *Educational Researcher*, 45(4), 252-257.
- Rutkowski, L., & Rutkowski, D. (2018). Improving the comparability and local usefulness of international assessments: A look back and a way forward. *Scandinavian Journal of Educational Research*, 62(3), 354–367.
- Samuelsen, K. (2005). *Examining differential item functioning from a latent class perspective*. Doctoral Dissertation. Available from ProQuest Dissertations and Theses database. (UMI No. 3175148)

- Samuelsen, K. M., & Dayton, C. M. (2010). Latent class analysis. In G. Hancock & R. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 173-184). New York: Routledge.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman and Hall/CRC.
- Schwarz, G. (1978). *Estimating the dimension of a model*. *Annals of Statistics*, 6, 461-464.
- Sen, S. (2014). *Robustness of mixture IRT models to violations of latent normality*. Doctoral Dissertation, The University of Georgia, USA.
- Sen, S., Cohen, A. S., & Kim, S. H. (2016). The impact of non-normality on extraction of spurious latent classes in mixture IRT models. *Applied Psychological Measurement*, 40, 98- 113.
- Sen, S. (2018). Spurious latent class problem in the mixed Rasch model: A comparison of three maximum likelihood estimation methods under different ability distributions. *International Journal of Testing*, 18(1), 71-100.
- Sen, S., & Cohen, A. S. (2019). Applications of mixture IRT models: A literature review, *Measurement: Interdisciplinary Research and Perspectives*, 17(4), 177-191.
- Sigott, G. (2004). *Towards identifying the c-test construct*. Frankfurt/am: Peter Lang.
- Smith, E.V., Ying Y., & Brown, S.W. (2012). Using the mixed Rasch model to analyze data from the beliefs and attitudes about memory survey. *Journal of Applied Measurement*, 13, 23-40
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (Sixth edition). United States: Pearson Education.
- Toker, T. (2016). *A comparison latent class analysis and the mixture Rasch model: A cross-cultural comparison of 8th grade mathematics achievement in the fourth international mathematics and science study (TIMSS-2011)*. Doctoral Dissertation, The Faculty of the Morgridge College of Education University of Denver, USA.
- Vermunt, J. K. (1997). *EM: A general program for the analysis of categorical data*. WORC Research Paper, Tilburg University.

- Vermunt, J. K., & Magidson, J. (2004). Latent class analysis. *The Sage encyclopedia of social sciences research methods*, 549-553.
- Vermunt, J. K., & Magidson, J. (2005). *Technical guide for latent GOLD 4.0: Basic and advanced*. Belmont, MA: Statistical Innovations.
- von Davier, M., & Rost, J. (2006). Mixture distribution item response models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26): Psychometrics, (pp. 643-661). Amsterdam, Netherlands: Elsevier.
- von Davier, M. & Rost, J. (2017). Logistic mixture-distribution response models. In W. J. van der Linden (Ed.), *Handbook of item response theory, volume one: Models* (p. 393-406). Boca Raton: Chapman and Hall/CRC
- von Davier, M., Rost, J., & Carstensen, C. H. (2007). Introduction: Extending the Rasch model. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 1–12). New York, NY: Springer.
- von Davier, & Yamamoto, K. (2007). Mixture- distribution and hybrid Rasch models. In M. von Davier, & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 99-115). New York: Springer Verlag.
- Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, 17, 228-243.
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. Hoboken, NJ: John Wiley & Sons.
- Willse, J. (2009). *MixRasch: Mixture Rasch models with JMLE*. Retrieved from <http://cran.r-project.org/web/packages/mixRasch/mixRasch.pdf>
- Yamamoto, K. & Kulick, E. (2000). *Scaling methodology and procedures for the TIMSS mathematics and science scales*. In M.O. Martin, K.D. Gregory & S.E. Stemler (Eds.), TIMSS 1999 Technical Report. (pp. 237– 263).
- Yang, X., Shaftel, J., Glasnapp, D., & Poggio, J. (2005). Latent class analysis of mathematical ability for special education students. *Journal of Special Education*, 38, 194–207.

- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125–145.
- Zhang, D., Orrill, C., & Campbell, T. (2015). Using the mixture Rasch model to explore knowledge resources students invoke in mathematics and science assessments. *School Science and Mathematics*, 115(7), 356-365.
- Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods*, 7(2), 168–190.

EKLER

EK-1. Veri Üretimi İçin Kullanılan R Kodu

```
setwd("D:/R_2c_10_10_1000")

nItems <- 10 # number of items

nSubj <- 1000 # number of subjects

permissing <- 0.10

nClass <- 2 # number of classes

pClass <- c(0.5, 0.5)

meanTheta1 <- 0 # mean of normal for class 1

meanTheta2 <- 0 # mean of normal for class 2

sdTheta1 <- 1

sdTheta2 <- 1

beta1 <- c(-2.7, -2.1, -1.5, -0.9, -0.3, 0.3, 0.9, 1.5, 2.1, 2.7)

# difficulty parameters for class 1

beta2 <- c(2.7, 2.1, 1.5, 0.9, 0.3, -0.3, -0.9, -1.5, -2.1, -2.7)

# difficulty parameters for class 2

for (iter in 1:100) {

  ### generating complete data set ###

  response <- mat.or.vec(nSubj, nItems+1)

  for (row in 1: nSubj) {

    r <- runif(1, 0, 1)

    if (r <= 0.5) {

      class <- 1

      beta <- beta1

      theta <- rnorm(1, mean = meanTheta1, sd = sdTheta1)

    }

    else {
```

```

class <-2

beta <-beta2

theta <- rnorm (1, mean = meanTheta2 ,sd= sdTheta2 )

}

for (column in 1: nItems) {

p= exp (theta - beta[column])/(1+ exp (theta - beta[column]))

r<- runif (1 ,0 ,1)

if (r <=p) {

response [row , column]<-1

}

else {

response [row , column]<-0

}

}

response [row , nItems +1]<- class

}

#### end of generating complete data set ####

#### generating MAR data set ####

temp <- response [ ,1: nItems ]

sum <- mat.or.vec(nSubj ,1)

group <- mat.or.vec(nSubj ,1)

missing <-mat.or.vec(nSubj ,1)

mar <- cbind (response, sum, group, missing)

mnar <- cbind (response, sum, group, missing)

for (row in 1: nSubj ) {

c<- mar [row , nItems +1]

```

```

if (c ==1) {
s<- sum (mar[row ,1:6])
if (s <=1) {
mar [row , nItems +3] <-1
}
else if (s >=2 & s <=4) {
mar [row , nItems +3] <-2
}
else if (s >=5 & s <=7) {
mar [row , nItems +3] <-3
}
else if (s >=8) {
mar [row , nItems +3] <-4
}
}
else if (c ==2) {
s<- sum (mar[row ,5:10])
if (s <=1) {
mar [row , nItems +3] <-1
}
else if (s >=2 & s <=4) {
mar [row , nItems +3] <-2
}
else if (s >=5 & s <=7) {
mar [row , nItems +3] <-3
}
}

```

```

else if (s >=8) {
  mar [row , nItems +3] <-4
}
}
}

count <-1

while (count <= nSubj*nItems*permissing) {
  for ( row in 1: nSubj ) {
    if (count <= nSubj*nItems*permissing) {
      g<- mar [row , nItems +3]
      c<- mar [row , nItems +1]
      if (c ==1) {
        if (g ==1) {
          for (i in 1:4) {
            r<- runif (1 ,0 ,1)
            if (r <0.6) {
              if ( mar [row , nItems -9+i ]!=9) {
                mar [row ,nItems -9+ i]<-9
                mar [row , nItems +4] <-1
                # 1 = missing , 0 = not missing
                count <- count +1
              }
            }
          }
        }
      }
    }
  }
  else if (g ==2) {

```

```

for (i in 1:4) {
  r<- runif (1 ,0 ,1)

  if (r <0.4) {

    if ( mar [row , nItems -9+i ]!=9) {

      mar [row ,nItems -9+ i]<-9

      mar [row , nItems +4] <-1

      # 1 = missing , 0 = not missing

      count <- count +1

    }

  }

}

else if (g ==3) {

  for (i in 1:4) {

    r<- runif (1 ,0 ,1)

    if (r <0.2) {

      if ( mar [row , nItems -9+i ]!=9) {

        mar [row ,nItems -9+ i]<-9

        mar [row , nItems +4] <-1

        # 1 = missing , 0 = not missing

        count <- count +1

      }

    }

  }

}

}

```

```

else if (c ==2) {
  if (g ==1) {
    for (i in 1:4) {
      r<- runif (1 ,0 ,1)
      if (r <0.6) {
        if ( mar [row ,i ]!=0) {
          mar [row ,i]<-0
          mar [row , nItems +4] <-1
          # 1 = missing , 0 = not missing
          count <- count +1
        }
      }
    }
  }
  else if (g ==2) {
    for (i in 1:4) {
      r<- runif (1 ,0 ,1)
      if (r <0.4) {
        if ( mar [row ,i ]!=0) {
          mar [row ,i]<-0
          mar [row , nItems +4] <-1
          # 1 = missing , 0 = not missing
          count <- count +1
        }
      }
    }
  }
}

```

```

}

else if (g ==3) {

for (i in 1:4) {

r<- runif (1 ,0 ,1)

if (r <0.2) {

if ( mar [row ,i ]!=0) {

mar [row ,i]<-0

mar [row , nItems +4] <-1

# 1 = missing , 0 = not missing

count <- count +1

}

}

}

}

}

}

else break

}

}

### generating MNAR data set ###

count <-1

while (count <= nSubj * nItems * permissing ) {

for ( row in 1: nSubj ) {

if (count <= nSubj * nItems * permissing ) {

c<- mnar [row , nItems +1]

if (c ==1) {

```

```

for (i in 1:4) {
  if ( mnar [row , nItems -9+i ]!=9) {
    if ( mnar [row , nItems -9+i ]==1) {
      r<- runif (1 ,0 ,1)
      if (r <0.3) {
        mnar [row , nItems -9+i]<-9
        mnar [row , nItems +4] <-1
        # 1 = missing , 0 = not missing
        count <- count +1
      }
    }
    else if ( mnar [row ,nItems -9+ i ]==0) {
      if (r <0.1) {
        mnar [row , nItems -9+i]<-9
        mnar [row , nItems +4] <-1
        # 1 = missing , 0 = not missing
        count <- count +1
      }
    }
  }
}

else if (c ==2) {
  for (i in 1:4) {
    if ( mnar [row ,i ]!=9) {
      if ( mnar [row ,i ]==1) {

```

```

r<- runif (1 ,0 ,1)

if (r <0.3) {

mnar [row ,i]<-9

mnar [row , nItems +4] <-1

# 1 = missing , 0 = not missing

count <- count +1

}

}

else if ( mnar [row ,i ]==0) {

if (r <0.1) {

mnar [row ,i]<-9

mnar [row , nItems +4] <-1

# 1 = missing , 0 = not missing

count <- count +1

}

}

}

}

}

}

else break

}

}

# generating output file for complete data

# response : item + 1 = class

comp <- response [ ,1: nItems ]

```

```

filename1 <- paste ("mplus",iter ,".dat", sep = "")
# file for item calibration

filename2 <- paste ("mplus_o",iter ,".dat", sep = "")
# original file containing other information

write.table (comp , file = filename1 ,sep ='\t', row.names =FALSE ,col.names = FALSE )
write.table (response , file = filename2, sep ='\t', row.names =FALSE ,col.names = FALSE
)

# generating output file for mar data

# mar : item + 1 = class , item + 2 = sum , item + 3 = group ,item + 4 = missing

# 1 = missing , 0 = not missing

resmar <-mar [ ,1: nItems ]

filename3 <- paste ("marmplus",iter ,".dat",sep = "")
filename4 <- paste ("marmplus_o",iter ,".dat", sep = "")

write.table (resmar , file = filename3 ,sep ='\t',row.names =FALSE ,col.names = FALSE )
write.table (mar, file = filename4 ,sep ='\t', row.names =FALSE ,col.names = FALSE )

# generating output file for mnar data

# mnar : item + 1 = class , item + 4 = missing

# 1 = missing , 0 = not missing

resmnar <- mnar [ ,1: nItems ]

filename5 <- paste ("mnarmplus",iter ,".dat", sep = "")
filename6 <- paste ("mnarmplus_o",iter ,".dat", sep = "")

write.table (resmnar, file = filename5 , sep ='\t',row.names =FALSE , col.names = FALSE
)

write.table (mnar, file = filename6 ,sep ='\t', row.names =FALSE ,col.names = FALSE )

}

```

EK-2. Birden Çok Input Dosyası Üretmek İçin Kullanılan Mplus Kodu

```
[[init]]
iterators=numbers;
numbers=1:100;
dir="C:\\Users\\asus\\Desktop\\SIMULASYON\\1PL\\CO210P60010\\input";
filename = "mplus[[numbers]].inp";
outputDirectory = [[dir]];
[/init]]

TITLE:Rasch model
DATA: FILE IS mplus1.dat;
VARIABLE: NAMES ARE M1-M10;
USEVARIABLES = M1-M10;
CATEGORICAL = M1-M10;
CLASSES = c(2);
ANALYSIS: TYPE = MIXTURE;
ALGORITHM = INTEGRATION;
MODEL: %OVERALL%
f BY M1-M10@1;
[f@0];
%c#1%
f BY M1-M10@1;
f@1;
[M1$1 -M10$1];
%c#2%
f BY M1-M10@1;
f@1;
[M1$1 -M10$1];
OUTPUT: TECH1 TECH8 TECH11;
SAVEDATA: FORMAT = F10.5;
FILE = "respondents[[numbers]].cprob";
SAVE = CPROBABILITIES;
```



GAZİLİ OLMAK AYRICALIKTIR...