





**DERECELENDİRİLMİŞ TEPKİ MODELİ TEMELLİ PARAMETRE  
KESTİRİMİNDE  
NORMALLİĞİN İHLALİNİN  
ÖLÇME KESİNLİĞİNE ETKİSİ**

**Yıldız Yıldırım**

**YÜKSEK LİSANS TEZİ  
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ  
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**EYLÜL, 2015**

## TELİF HAKKI ve TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren .....(.....) ay sonra tezden fotokopi çekilebilir.

### YAZARIN

Adı : Yıldız

Soyadı : Yıldırım

Bölümü : Eğitimde Ölçme ve Değerlendirme

İmza :

Teslim tarihi:

### TEZİN

Türkçe Adı : Derecelendirilmiş Tepki Modeli Temelli Parametre Kestiriminde Normalliğin İhlalinin Ölçme Kesinliğine Etkisi

İngilizce Adı : The Effect of Normality Violation in the Process of Parameter Estimation Based upon Graded Response Model on Measurement Precision

## 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: Yıldız Yıldırım

İmza :

## **Jüri onay sayfası**

Yıldız Yıldırım tarafından hazırlanan “Derecelendirilmiş Tepki Modeli Temelli Parametre Kestiriminde Normalliğin İhlalinin Ölçme Kesinliğine Etkisi” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

**Danışman:** Doç. Dr. Şeref TAN

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi) .....

**Başkan:** Prof. Dr. Mehtap ÇAKAN

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi)

.....

**Üye:** Yrd. Doç. Dr. Deniz Gülleroğlu

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Ankara Üniversitesi)

.....

Tez Savunma Tarihi:

Bu tezin Eğitimde Ölçme ve Değerlendirme Anabilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Servet KARABAĞ

Eğitim Bilimleri Enstitüsü Müdürü .....



*Anneme ve Babama*

## TEŞEKKÜR

Hem tez yazma sürecimde hem de aldığım derslerde derin akademik bilgisinden yararlanmamı sağladığı ve bana zaman ayırıp her zaman destek olduğu için örnek aldığım danışmanım sayın Doç Dr. Şeref TAN'a,

Yüksek lisans eğitimim ve çalışma hayatım süresince kendilerinden pek çok şey öğrendiğim saygıdeğer hocalarım Prof. Dr. Şener BÜYÜKÖZTÜRK, Prof. Dr. Mehtap ÇAKAN, Doç. Dr. Hakan Y. ATAR, Doç. Dr. İsmail KARAKAYA, Yrd. Doç. Dr. Emine ÖNEN'e ve lisans eğitimim sırasında beni ölçme ve değerlendirmeye tanıstırıp desteklerini yüksek lisans aşamasında da sürdüren değerli hocam Dr. M. Gülşah ŞAHİN'e

Tez jürimde bulunarak önemli katkılarını esirgemeyen Yrd. Doç. Dr. Deniz GÜLLEROĞLU'na ve elde ettikleri verileri tez çalışmam için benimle paylaşan Doç. Dr. Haluk MERGEN ve arkadaşlarına,

Bu süreçte beni yalnız bırakmayan, manevi desteklerini benden hiç bir zaman eksik etmeyen canım arkadaşlarım Arş. Gör Fazilet Gül İNCE, Arş. Gör. Tuba GÜNDÜZ ve Arş. Gör. Elif SEZER'e, ayrıca yine manevi desteklerini esirgemeyen ve tez çalışmamda da teknik bilgilerinden yararlandığım sevgili arkadaşlarım Arş. Gör Ayşenur ERDEMİR, Arş. Gör. Ertunç UKŞUL, Muhammet SEVİMKAN ve hocam Dr. Nagihan BOZTUNÇ ÖZTÜRK'e,

Hiç bir zaman yardımını ve desteğini esirgmeden tüm içtenliğiyle yanımda bulunan ve bu günlere gelmemde çok büyük katkıları olan müstakbel eşim Kurtuluş GÖRGÜLÜ'ye ve bugünlere gelmemi sağlayan, bana her zaman inanarak ve güvenerek yanımda olan biricik annem Hanife YILDIRIM'a, canım babam N. Rifat YILDIRIM'a ve kardeşim Cahit YILDIRIM'a,

Yüksek lisans eğitimimde yurt içi yüksek lisans burs olanağı sağlayan TÜBİTAK'a en içten teşekkürlerimi sunarım.



**DERECELENDİRİLMİŞ TEPKİ MODELİ TEMELLİ PARAMETRE  
KESTİRİMİNDE  
NORMALLİĞİN İHLALİNİN  
ÖLÇME KESİNLİĞİNE ETKİSİ  
(Yüksek Lisans Tezi)**

**Yıldız Yıldırım  
GAZİ ÜNİVERSİTESİ  
EĞİTİM BİLİMLERİ ENSTİTÜSÜ  
Eylül, 2015**

**ÖZ**

Likert tipi ölçekler arařtırmalarda sıklıkla kullanılan ölçek türlerinden biridir. Madde tepki kuramı ile analizinde kullanılan eşitlikler orijinalde normallik sayılıtısının sağlanmasını gerektirmektedir ancak bazı çalışma gruplarında bu sayılıtının sağlanması oldukça zordur. Bu arařtırmanın amacı, çok kategorili puanlanan maddelerden oluşan bir ölçeğin Samejima'nın derecelendirilmiş tepki modeli (DTM) ile analizinde normallik sayılıtısı ihlalinin ölçme kesinliğine etkisini arařtırmaktır. Bu arařtırma, likert tipi ölçeklerden elde edilen verilerin normal dağılımdan farklılaşmasının ölçme kesinliğini nasıl etkilendiğini ortaya koyacağı için önemlidir. Arařtırmanın amacı doğrultusunda çarpıklık katsayıları 0,00; 0,50; 1,00; 1,50 ve 2,00 olan her dağılım için 25 replikasyon yapılarak 3000 kişiden oluşan simülasyon veri setleri, "Beck Depresyon Envanteri" ile elde edilen madde parametreleri kullanılarak, Wingen 3.1 programıyla üretilmiştir. Üretilen veri setleri için madde parametresi kestirimleri MULTILOG 7.03'te marginal maximum likelihood (MML) kestirim yöntemiyle yapılmıştır ve ölçme kesinliği RMSE ve AAD istatistikleri ile incelenmiştir. Arařtırmanın bulgularında çarpıklık katsayısı arttığında a ve b parametreleri için RMSE ve AAD değerlerinin arttığı ve normalliğin ihlal edildiği durumlarda bu değerlerin oldukça yüksek olduğu görülmüştür, bu ölçme kesinliğinin azaldığının bir göstergesidir. Bu nedenle normallik sağlanmadığında DTM'ye dayalı olarak MML yöntemi ile madde parametresi kestirimlerinin uygun olmayacağı sonucuna varılmıştır.

Bilim Kodu : 6015

Anahtar Kelimeler: Derecelendirilmiş Tepki Modeli (DTM), Ölçme Kesinliği, Madde Tepki Kuramı, Normallik, Normalliğin İhlali.

Sayfa Adedi : xv + 96

Danışman : Doç. Dr. Şeref TAN

**THE EFFECT OF NORMALITY VIOLATION IN THE PROCESS OF  
PARAMETER ESTIMATION BASED UPON GRADED RESPONSE  
MODEL ON MEASUREMENT PRECISION**

**(M.S Thesis)**

**Yıldız Yıldırım**

**GAZI UNIVERSITY**

**GRADUATE SCHOOL OF EDUCATIONAL SCIENCES**

**September, 2015**

**ABSTRACT**

Likert-type scale is one of the scale types that are used in research commonly. Originally, normality assumption should be taken into consideration in this scale's analysis with item response theory equations but the handling of normality is difficult in some samples. The purpose of this research was to examine the effect of normality violation in the process of parameter estimation based upon Samejima's graded response model (GRM) on measurement precision in the scale consisting of polytomous item. This study is important because it demonstrated how data that were obtained from a Likert-type scale and that differed from the normal distribution affected measurement precision. For the purpose of this research, data sets whose skewness coefficients were 0,00; 0,50; 1,00; 1,50 and 2,00 were simulated by using item parameters obtained from "Beck Depression Inventory" by using Wingen 3.1, and these data sets were consisted of 3000 examinees and were replicated 25 times for each distribution. Item parameters for simulated data sets were estimated by marginal maximum likelihood estimation method (MML) by using MULTILOG 7.03 and measurement precision was examined by RMSE and AAD. In the findings of the research it was observed that RMSE and AAD values for a and b parameters increased when skewness was increased. In other words, when normality of score distribution was violated, RMSE and AAD were quite high and measurement precision was low. Therefore, it has been concluded that item parameters estimated by MML estimation method based on GRM may not be appropriate when normality violated.

Science Code :

Key Words : Graded Response Model (GRM), Measurement Precision, Item Response Theory, Normality, Violation of Normality, Samejima.

Page Number : xv + 96

Supervisor : Associated Professor Şeref TAN

## İÇİNDEKİLER

TELİF HAKKI ve TEZ FOTOKOPİ İZİN FORMU.....	i
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
Jüri onay sayfası.....	iii
TEŞEKKÜR .....	v
ÖZ.....	vi
ABSTRACT.....	vii
İÇİNDEKİLER.....	viii
TABLolar LİSTESİ .....	xii
ŞEKİLLER LİSTESİ .....	xiii
SİMGE VE KISALTMALAR LİSTESİ.....	xiv
TANIMLAR .....	xv
BÖLÜM I.....	1
GİRİŞ .....	1
Problem Durumu .....	1
Klasik Test Kuramı .....	2
Madde Tepki Kuramı.....	3
Madde Tepki Kuramı Modelleri (Lojistik).....	8
Çok Kategorili MTK Modelleri.....	9
Samejima'nın Derecelendirilmiş Tepki Modeli.....	10
Ölçme Kesinliği.....	13
Araştırmanın Amacı .....	14

Araştırmanın Önemi.....	15
Sayıtlar.....	15
Sınırlılıklar .....	16
<b>BÖLÜM II .....</b>	<b>17</b>
<b>İLGİLİ ARAŞTIRMALAR .....</b>	<b>17</b>
<b>BÖLÜM III.....</b>	<b>35</b>
<b>YÖNTEM.....</b>	<b>35</b>
Araştırmanın Modeli .....	35
Çalışma Grubu.....	35
Veri Toplama Aracı.....	36
Beck Depresyon Envanteri .....	36
Ölçüm Güvenirliği.....	38
Verilerin Üretilmesi .....	40
MTK Varsayımlarının Kontrolü .....	45
Verilerin Çözümlemesi ve Yorumlanması.....	45
<b>1. Alt Problem İçin Verilerin Analizi .....</b>	<b>46</b>
<b>1.1. Alt Problem İçin Verilerin Analizi.....</b>	<b>47</b>
<b>1.2. Alt Problem İçin Verilerin Analizi.....</b>	<b>47</b>
<b>2. Alt Problem İçin Verilerin Analizi .....</b>	<b>48</b>
<b>2.1. Alt Problem İçin Verilerin Analizi.....</b>	<b>49</b>
<b>2.2. Alt Problem İçin Verilerin Analizi.....</b>	<b>49</b>
<b>3. Alt Problem İçin Verilerin Analizi .....</b>	<b>50</b>
<b>4. Alt Problem İçin Verilerin Analizi .....</b>	<b>50</b>
<b>5. Alt Problem İçin Verilerin Analizi .....</b>	<b>51</b>
<b>BÖLÜM IV .....</b>	<b>53</b>
<b>BULGULAR VE YORUM.....</b>	<b>53</b>
<b>1. Alt Probleme İlişkin Bulgular.....</b>	<b>53</b>

1.1. Alt Probleme İlişkin Bulgular .....	53
1.2. Alt Probleme İlişkin Bulgular .....	55
2. Alt Probleme İlişkin Bulgular .....	58
2.1. Alt Probleme İlişkin Bulgular .....	58
2.2. Alt Probleme İlişkin Bulgular .....	60
3. Alt Probleme İlişkin Bulgular .....	63
4. Alt Probleme İlişkin Bulgular .....	64
5. Alt Probleme İlişkin Bulgular .....	66
BÖLÜM V.....	69
SONUÇ ve TARTIŞMA .....	69
1. Alt Probleme İlişkin Sonuçlar .....	69
1.1. Alt Probleme İlişkin Sonuçlar .....	69
1.2. Alt Probleme İlişkin Sonuçlar .....	70
2. Alt Probleme İlişkin Sonuçlar .....	71
2.1. Alt Probleme İlişkin Sonuçlar .....	71
2.2. Alt Probleme İlişkin Sonuçlar .....	71
3. Alt Probleme İlişkin Sonuçlar .....	71
4. Alt Probleme İlişkin Sonuçlar .....	72
5. Alt Probleme İlişkin Sonuçlar .....	72
BÖLÜM VI.....	75
ÖNERİLER .....	75
Uygulayıcılar İçin Öneriler .....	75
Araştırmacılar İçin Öneriler.....	75
KAYNAKLAR .....	77
EKLER .....	83
EK 1. Beck Depresyon Envanteri.....	84
EK 2. Dağılım Türlerinin Histogram Dağılım Diyagramları .....	86

<b>EK 3. Dağılım Türlerinin Kümülatif Dağılım Diyagramları.....</b>	<b>88</b>
<b>EK 4. Bazı Maddelere Ait İşlevsel Karakteristik Eğrileri .....</b>	<b>90</b>
<b>EK 5. Bazı Maddelere Ait Madde Bilgi Fonksiyonları.....</b>	<b>92</b>
<b>EK 6. Test Cevap Fonksiyonu .....</b>	<b>94</b>
<b>EK 7. Test Bilgi Fonksiyonu .....</b>	<b>95</b>
<b>EK 8. Örnek MULTILOG Komut Dosyası .....</b>	<b>96</b>



## TABLolar LİSTESİ

Tablo 1. BDE Madde ve Ölçek İstatistikleri: Madde Ortalamaları, Madde Standart Sapmaları, Madde-Toplam Korelasyonları ( $r_{it}$ ), Örneklem Büyüklüğü, Ham Cronbach $\alpha$ , Ölçek Ortalaması ve Ölçek Standart Sapması. ....	39
Tablo 2. Verilerin Üretilmesinde Kullanılan Madde Parametreleri.....	40
Tablo 3. Verilerin Üretilmesinde Kullanılan Madde Parametrelerine İlişkin Betimleyici İstatistikler.....	41
Tablo 4. Dağılımların Çarpıklık Düzeylerine Göre Verilerin Üretilmesinde Kullanılan Parametreler .....	41
Tablo 5. Her Dağılım Türü İçin Betimleyici İstatistikler .....	45
Tablo 6. Her Dağılım Türü İçin a Parametresi RMSE Değerleri .....	53
Tablo 7. Her Dağılım Türü İçin $b_1, b_2, b_3$ Parametresi RMSE Değerleri.....	55
Tablo 8. Her Dağılım Türü İçin a Parametresi AAD Değerleri.....	59
Tablo 9. Her Dağılım Türü İçin $b_1, b_2, b_3$ Parametresi AAD Değerleri .....	60
Tablo 10. Her Dağılım Türü İçin Marjinal Güvenirlik Katsayısı .....	63
Tablo 11. Her Dağılım Türü ve Replikasyon İçin Maximum Intercycle Parameter Change Değerleri .....	65
Tablo 12. Her Dağılım İçin Kriteri Sağlamayan Replikasyonların Sayısı ve Yüzdeleri .....	66
Tablo 13. Her Dağılım Türü İçin Tüm Replikasyonlar MULTLOG Programının Verdiği Hata Sayıları.....	67
Tablo 14. Her Dağılım Türü İçin Replikasyonlardaki Ortalama Hata Sayısı.....	68
Tablo 15. Her Dağılım Türü İçin Hata Veren ve Vermeyen Replikasyon Sayısı ve Yüzdesi .....	68

## ŞEKİLLER LİSTESİ

Şekil 1. Madde karakteristik eğrisi .....	7
Şekil 2. Beş kategorili bir madde için işlevsel karakteristik eğrisi .....	11
Şekil 3. 5 kategorili bir madde için kategori tepki fonksiyonları (0-4 puanlanan madde için belli bir kategoriyi ve ondan daha yüksek kategorileri seçme olasılığı).....	12
Şekil 4. $\text{ÇK} = 0,00$ dağılımına sahip verilerin üretilmesinde WinGen 3.1. arayüzü .....	42
Şekil 5. $\text{ÇK} = 1,00$ dağılımına sahip verilerin üretilmesinde WinGen 3.1 arayüzü .....	43
Şekil 6. $\text{ÇK} = 2,00$ dağılımına sahip verilerin üretilmesinde WinGen 3.1 arayüzü .....	44
Şekil 7. a parametresi için dağılım türü ile RMSE arasındaki ilişki.....	54
Şekil 8. DTM $b_1, b_2, b_3$ parametreleri için dağılım türü ile RMSE arasındaki ilişki.....	57
Şekil 9. DTM $b_1, b_2, b_3$ parametreleri için dağılım türü ile RMSE arasındaki ilişki ( $\text{ÇK} = 2,00$ olan dağılım hariç) .....	58
Şekil 10. a parametresi için dağılım türü ile AAD arasındaki ilişki .....	59
Şekil 11. $b_1, b_2, b_3$ parametreleri için dağılım türü ile AAD arasındaki ilişki .....	62
Şekil 12. $b_1, b_2, b_3$ parametreleri için dağılım türü ile AAD arasındaki ilişki ( $\text{ÇK} = 2,00$ olan dağılım hariç) .....	62
Şekil 13. Dağılım türleri ile marjinal güvenilirlik katsayısı arasındaki ilişki .....	64



## SİMGE VE KISALTMALAR LİSTESİ

DTM	Derecelendirilmiş Tepki Modeli
MTK	Madde Tepki Kuramı
MKE	Madde Karakteristik Eğrisi
İKE	İşlevsel Karakteristik Eğrisi
KTK	Klasik Test Kuramı
BDE	Beck Depresyon Envanteri
RMSE	Root Mean Squared Error
AAD	Ortalama Mutlak Farklılık
ÇK	Çarpıklık Katsayısı
BK	Basıklık Katsayısı
1PLM	Bir Parametrelili Lojistik Model
2PLM	İki Parametrelili Lojistik Model
3PLM	Üç Parametrelili Lojistik Model
Bias	İstatistiksel Yanlılık
$\hat{a}_i$	Kestirilen a parametresi
$a_i$	Gerçek a parametresi
I	Madde sayısı
$\widehat{b}_{1i}$ , $\widehat{b}_{2i}$ ve $\widehat{b}_{3i}$	Sırasıyla i maddesi için kestirilen $b_1$ , $b_2$ ve $b_3$ madde güçlük parametreleri
$b_{1i}$ , $b_{2i}$ ve $b_{3i}$	Sırasıyla i maddesi için gerçek $b_1$ , $b_2$ ve $b_3$ madde güçlük parametreleri

## TANIMLAR

Derecelendirilmiş Tepki Modeli: Çok kategorili puanlanan maddelerde belirli bir kategorinin ve bu kategoriden daha yüksek kategorilerin seçilme olasılığına dayanan, sıralı cevap kategorilerine sahip maddeler için kullanılan iki parametrelili MTK modelidir.

Çok Kategorili Madde (Polytomous madde): İki kategoriden daha fazla kategoriye sahip olan maddelere polytomous maddeler denir (De Mars, 2010).

Normallik: Ölçümlerin z puanı cinsinden ortalamasının 0, standart sapmasının 1 olduğu, çarpıklık ve basıklığın 0 olduğu simetrik dağılım durumu normal dağılım durumu olarak tanımlanmaktadır, ayrıca bu araştırma da çarpıklığın -1– +1 aralığında olduğu durumlarda normallik durumu olarak tanımlanmıştır.

Çarpıklık Katsayısı: Ölçüm değerlerinin verinin değer aralığının hangi alt aralığında daha fazla gözlendiği konusunda bilgi veren ölçü olarak tanımlanmaktadır (Ünver, Gamgam ve Altunkaynak, 2013).

# BÖLÜM I

## GİRİŞ

Bu bölümde problem durumuna, problem durumuna ilişkin kuramsal bilgilere, araştırmanın önemine, amacına, sayıltılarına ve sınırlılıklarına yer verilmiştir.

### **Problem Durumu**

Bilim Russell'e göre gözlem ve gözleme dayalı akıl yürütme yoluyla evrendeki olguları ve bu olguları birbirine bağlayan yasaları bulmaya çalışmaktadır (Russell, 2008). Bilim temel olarak Fen bilimleri, sosyal bilimler, din bilimleri olmak üzere üç kısma ayrılmaktadır ve bunlar pek çok ana bilim dalı ve bilim dalı içeren şemsiyelerdir. Psikoloji, sosyal bilimler şemsiyesi altında bir ana bilim dalıdır ve sosyal bilimler zaman ve mekân boyutları içerisinde ihtiyaçlara paralel olarak, insanda ve insan topluluklarında görülen içe ve dışa dönük ilişkileri inceler. Psikoloji de insanda görülen içe ve dışa dönük değişiklikleri ve davranış şekillerini inceleyen ana bilim dalı olarak tanımlanmaktadır (Kömürcü, 1992). İnsan davranışlarını inceleme konusu yapan psikolojide ölçme sorunlarının fen bilimlerine göre çok daha zor ve karmaşık olacağı açıktır. Çünkü psikolojide ölçülen değişkenler soyut değişkenlerdir.

Psikolojik yapılar doğrudan gözlenememesine ve ölçülememesine rağmen davranışlarla kendini gösteren olgulardır. Psikolojik yapılar soyut olduğu ve dolaylı yollarla ölçülebildikleri için bu özellikleri ölçmek amacıyla ölçme araçları tasarlanmaktadır (Crocker ve Algina, 1986). Günümüzde psikolojik yapıları ölçen farklı yöntemler vardır ve buna bağlı olarak ölçme araçları bulunmaktadır. Bu ölçme araçlarından biri de farklı cevap kategorilerine sahip ölçeklerdir. "Bogardus'un Toplumsal Uzaklık Ölçeği", "Thurstone Eşit Görünümlü Aralıklar Ölçeği", "Likert'in Dereceleme Toplamlarıyla Ölçekleme Tekniği" ve "Gutmann Yığılımlı Ölçekleme Tekniği" ve "Osgood Duygusal Anlam Ölçeği" bu ölçek türlerindedir (Tavşancıl, 2014). Alan yazında en sık kullanılan ölçek türü Likert tipi

ölçektir. Bu tip ölçekle sık karşılaşılmamasının nedeni, ölçek oluşturma süreci bakımından diğer ölçek türlerinden daha ekonomik olmasıdır (Tezbaşaran, 1997). Likert tarafından geliştirilen bu yöntem “dereceleme toplamlarıyla ölçekleme” olarak da bilinir (Turgut ve Baykul, 1992; Tezbaşaran, 1997; Erkuş, 2012; Tavşancıl, 2014). Likert tipi ölçeklerde verilen tepkiler sıralı kategorilere yöneliktir. Alan yazında Likert tipi ölçeklerden elde edilen verilerde hem Klasik Test Kuramı (KTK)’na dayalı olarak hem de Modern Test Kuramına dayalı olarak testin psikometrik özellikleri belirlenmekte ve madde analizleri yapılmaktadır. KTK bağlamında genellikle bu ölçeklerdeki maddeler için madde güçlüğü ve madde ayırt ediciliği alt-üst grup analizi ile hesaplanmakta ve ölçümlerin güvenilirliği ile geçerliği belirlenmektedir. Likert tipi ölçeklerde ölçümlerin güvenilirliğini belirlemek amacıyla genelde Cronbach- $\alpha$  katsayısı hesaplanmakta, yapı geçerliği için açımlayıcı ve/veya doğrulayıcı faktör analizi yapılmakta ve psikolojik yapıya bağlı olarak ölçüt geçerliği çalışmaları yapılmaktadır. MTK’de ise genel olarak DTM bağlamında a parametresi ve kategori sayısının bir eksiği kadar b parametresi hesaplanmakta ve marjinal güvenilirlik katsayısı hesaplanmaktadır.

Aşağıda KTK’nin temeli ve sınırlılıkları kısaca açıklanmıştır.

### **Klasik Test Kuramı**

KTK’nin temeli; bazı model sayıtları sağlandığında gözlenen test puanının, gerçek puanın ve randum hata puanının bileşimi olduğuna dayanır ve bu

$$X = T + E$$

formülüyle ifade edilir (Crocker ve Algina, 1986). Burada X gözlenen test puanını, T gerçek puanı ve E random hatayı temsil etmektedir. Örtük Özellikler Kuramı olarak da bilinen Madde Tepki Kuramı (MTK)’nda ise test puanlarında istatistiksel kestirimler yapmak amacıyla bazı matematiksel modeller geliştirilmiştir (van der Linden ve Hambleton, 1996). Bu nedenle KTK, MTK’ye göre daha kolay kullanılabilir olsa da birçok sınırlılığa sahiptir. Bu sınırlılıklardan en önemlisi birey özelliklerinin ve test özelliklerinin birbirinden ayrılmamış olmasıdır; yani test özellikleri gruba, grup özellikleri teste bağımlıdır. Grup özelliklerinin teste bağımlı olduğunu, bireylere zor bir test uygulanıyorsa yeteneklerinin (gerçek puanları) düşük, kolay bir test uygulanıyorsa yeteneklerinin yüksek olarak belirlenmesinden anlayabiliriz. Ayrıca madde güçlüğü maddeyi doğru cevaplayanların gruptaki tüm cevaplayıcılara oranı olduğu ve madde ayırt ediciliği, ölçüm

güvenirligi ve geçerligi de gruptan gruba deđiřtiđi için test özelliklerinin de grup özelliklerine bađımlı olduđunu söyleyebiliriz. Farklı testler uygulanan bireylerin karşılaştırılmasının zor olması da KTK'nin bir sınırlılıđıdır. Sadece farklı testlerde deđil aynı ya da paralel testleri alan bireyleri karşılařtırmakta problemlı bir durumdur. Güçlüđü birbirinden büyük oranda farklı olan iki testi ayrı ayrı yanıtlayan ve maddelerin yarısını dođru yapan iki bireyin birbirine denk olduđu düşünülemez. KTK'ye göre ölçme kesinliđi güvenirlilik katsayısı ya da ölçmenin standart hatası olarak belirlenir. KTK çerçevesinde güvenirlilik katsayısı kestirim yollarından biri, bir testin paralel formlarından elde edilen test puanları arasındaki korelasyonun hesaplanmasıdır; ancak tam olarak paralel formların oluşturulması imkânsız olmasa da zordur. Paralel formların oluşturulmasının zorluđu da KTK'de güvenirlilik belirlemeye yönelik bir sınırlılıktır. Ölçümlerin güvenirliliđinin ve varyansının bir fonksiyonu olan ölçmenin standart hatası ile ilgili bir problemde her puan için özgün bir kestirim yapılmayıp tüm test puanlarına ait tek bir hata dađılımı biriminin kestirilmesi. KTK'nin son sınırlılıđı ise madde yerine test odaklı olmasıdır. Klasik gerçek puan modeli belirli bir maddeyi belirli bir bireyin nasıl cevapladıđı hakkında bir bilgi sağlamaz (Hambleton, Swaminathan ve Rogers, 1991). Tüm bu sınırlılıklardan ötürü ölçme uzmanları alternatif kuramlar ve ölçme modelleri için arayıřa girdiklerini ifade etmişlerdir. Alternatif test kuramında aranan özellikler ařađıdaki gibi tanımlanmıştır:

1. Madde özelliklerinin gruptan bađımsız olması
2. Bireyin yeterliliđini gösteren puanların testten bađımsız olması
3. Test düzeyi yerine madde düzeyinde ifade edilebilen bir model olması
4. Güvenirliliđi hesaplamak için paralel testleri gerektirmeyecek bir model olması
5. Her bir yetenek puanı için hassas ölçüm sađlayan bir model olmasıdır.

Bu özellikler madde tepki kuramı olarak bilinen test kuramında bulunmaktadır (Lord, 1980; Hambleton ve Swaminathan, 1985; Hambleton vd., 1991).

Madde tepki kuramının temelleri, varsayımları ve parametreleri ařađıda kısaca açıklanmıştır.

### **Madde Tepki Kuramı**

Madde tepki kuramının KTK'ye göre avantajları olduđu açıktır ancak sađlanması genellikle güç olan sayılıtları vardır. Bu sayılıtlar tek boyutlu MTK modellerinde tek boyutluluk, yerel bađımsızlık, testin hız testi olmaması ve MTK normal ogive

modellerinde normalliktir. Tek boyutluluk ve yerel bağımsızlık ilişkili kavramlar olup tek boyutluluk varsayımı test maddelerinin tek bir yeteneği ölçtüğü anlamına gelmektedir (Hambleton vd., 1991). Tek boyutluluk varsayımı maddeler arası istatistiksel bağımlılığa dayanmaktadır (Crocker ve Algina, 1986). Yerel bağımsızlık ise aynı yetenek düzeylerindeki bireylerin herhangi bir madde çiftine verdikleri cevapların birbirinden istatistiksel olarak bağımsız olmasıdır. Başka bir deyişle, yetenek düzeyi sabit olduğunda testteki maddelerden birine verilen cevabın başka bir maddeye bağımlı olmaması gerekmektedir. Yani bireyin bir maddedeki performansı başka bir maddeye verdiği cevabı etkilememelidir (Hambleton vd., 1991). Tek boyutluluk ve yerel bağımsızlık arasındaki ilişkiden bahsedecek olursak tek boyutluluk ve yerel bağımsızlık kavramları aynı anlama gelen kavramlar değildir. Ancak genellikle testin boyutluluğu yerel bağımsız olabilen örtük özellik sayısına eşittir. Buna rağmen maddeler yerel bağımsız olduğunda tek boyutluluk varsayımının sağlandığı söylenemez (Crocker ve Algina, 1986). Hambleton vd. (1991)'ne göre ise tek boyutluluk varsayımı sağlandığında yerel bağımsızlık sayılı da sağlanmış olur. Dolayısıyla MTK modelleri ile yapılacak madde ve test analizlerine geçilmeden önce sayılıların sağlanıp sağlanmadığı test edilirken sadece tek boyutluluk sayılısını inceleyerek yerel bağımsızlık sayılı hakkında da bilgi sahibi olunabilir.

MTK'nin diğer bir sayılı olan normallik ise ölçümlerin normal dağılım göstermesi gerektiği anlamına gelmektedir. Normallik sayılısının sağlanıp sağlanmadığını test etmenin çeşitli yöntemleri vardır. Bu yöntemlerden grafiksel olanları; normal dağılım eğrisinin çizdirildiği histogramları, gövde yaprak diyagramlarını, boxplot diyagramlarını, normal Q-Q, Detrended Q-Q grafiğini (Büyüköztürk, Çokluk ve Köklü, 2012) ve P-P grafiğini (Field, 2009) incelemektir. İstatistiksel yöntemlerden ise örneklem büyüklüğünün 50'den fazla olması durumunda Kolmogorov-Smirnov hipotez testini, 50'den az olması durumunda Shapiro-Wilk hipotez testini kullanmaktır. Normallik sayılısının sağlanıp sağlanmadığı çarpıklık katsayısının incelenmesiyle de tespit edilebilir. Tam simetrik dağılımda (standart normal dağılım) çarpıklık katsayısı 0'dır ve çarpıklık katsayısının  $\pm 1$  arasında olması, dağılımın normalden aşırı bir sapma göstermediği anlamına gelmektedir (Büyüköztürk, 2012). SPSS paket programı kullanılarak elde edilen çarpıklık katsayısının standart hatasına bölünmesiyle elde edilen z-istatistiğini incelemekte normallik sayılısını kontrol etmenin bir yoludur.

Çarpıklık katsayısı, bir dağılımda verilerin daha çok solda, sağda ya da merkezde yığılmış olduğunu göstermektedir. Dağılımdaki verilerin bir yerde yığılması özelliğine dağılımın kayışıklığı ya da çarpıklığı denir. Çarpıklık katsayısı

$$\text{ÇK} = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n \cdot S^3}$$

eşitliğiyle tanımlanır (Baykul, 2010). Burada

$X_i$ : i. öğrencinin gözlenen test puanı,

$\bar{X}$ : gözlenen test puanlarının aritmetik ortalamasını,

n: testi alan birey sayısını,

S: standart sapmayı ifade etmektedir.

Baykul ve Güzeller (2013) yukarıdaki eşitlik dışında, çarpıklık ölçüsünün belirlenmesinde, ortalama ve ortancaya, üçüncü momente ve çeyreklere dayanan üç yaklaşımdan yararlanabileceğine değinmişlerdir. Bu yöntemlerin ilki Pearson çarpıklık katsayısı olarak bilinen “çarpıklığın ortalama ve ortancaya dayanılarak hesaplanması”dır ve çarpıklığın büyüklüğü ortalama ve ortanca arasındaki farka dayalı olarak

$$P_{\text{Ç.K.}} = \frac{3 \cdot (\bar{X} - X_{\text{ortanca}})}{S}$$

eşitliğiyle hesaplanabilir. Bu eşitlikte

$\bar{X}$ : gözlenen test puanlarının aritmetik ortalamasını,

$X_{\text{ortanca}}$ : gözlenen test puanlarının medyanı

S: standart sapmayı ifade etmektedir.

Bir diğer yöntem “çarpıklığın üçüncü moment yardımıyla hesaplanması” ise

$$\text{ÇK} = \sqrt{\frac{n}{n-1}} \cdot \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{(n-1) \cdot S^3}$$

eşitliğiyle yapılır ve bu eşitlik ortalama ve ortanca farkına dayanan yöntemdeki eşitlikten daha iyidir. Bu eşitlikte

$X_i$ : i. öğrencinin gözlenen test puanı,

$\bar{X}$ : gözlenen test puanlarının aritmetik ortalamasını,

n: testi alan birey sayısını,

S: standart sapmayı ifade etmektedir.

Aşağıda sunulan son yöntem ise “çarpıklığın çeyrekler yardımıyla hesaplanması”dır ve bu yöntemde ortanca, birinci ve üçüncü çeyrek kullanılır. Bu hesaplama

$$\text{ÇK} = \frac{Y_{75} + Y_{25} - 2X_{ortanca}}{Y_{75} + Y_{25}}$$

eşitliğiyle yapılır. Burada

$Y_{75}$ : üçüncü çeyreği,

$Y_{25}$ : birinci çeyreği,

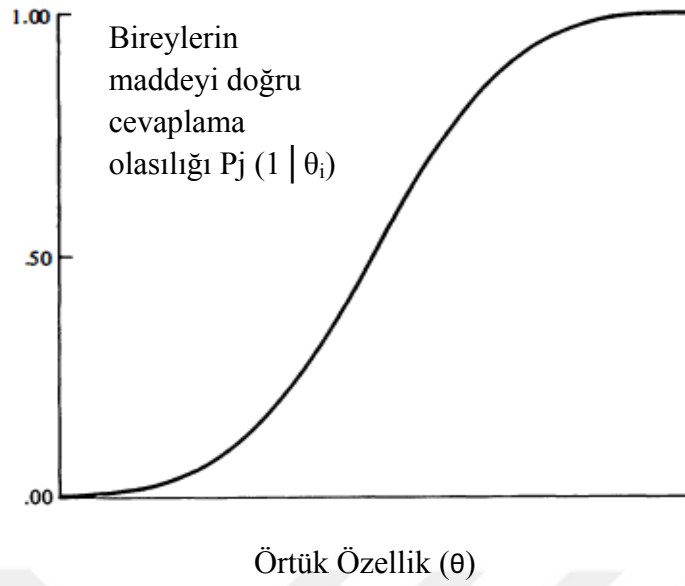
$X_{ortanca}$ : gözlenen test puanlarının medyanı temsil etmektedir.

Tüm bu eşitliklerden elde edilen çarpıklık katsayısı pozitif, negatif değerler olabilir veya sıfır olabilir. Katsayı negatif olduğunda dağılım sola çarpık ya da negatif kayışlı, pozitif olduğunda ise sağa çarpık ya da pozitif kayışlıdır (Baykul, 2010).

KTK’deki matematiksel modellerin Lord ve Novick (1968) tarafından yeniden formüle edilmesiyle MTK’nin temel kavramları oluşmuştur (Embretson ve Reise, 2000; Baker ve Kim, 2004; Erkuş, 2012). MTK iki temel sayılıya dayanan bir kuramdır. Bunlardan birincisi “Test maddesini yanıtlayan bireyin performansı özellik, örtük özellik veya yetenek olarak adlandırılan bir dizi faktör aracılığıyla kestirilebilir.” olarak ifade edilir. İkincisi ise “Bireylerin madde performansı ve bu performansın altında yatan özellikler arasındaki ilişki, madde karakteristik fonksiyonu (MKF) veya madde karakteristik eğrisi (MKE) olarak adlandırılan ve monoton şekilde artan bir fonksiyon ile gösterilebilir.” olarak ifade edilir (Hambleton vd., 1991).

MKE’de maddeyi doğru cevaplama olasılığı performansın altında yatan örtük özelliğin ( $\theta$ ) ve madde parametrelerinin bir fonksiyonudur ve çoğu MTK uygulamasında MKE’nin Şekil 1’deki gibi S şeklinde olduğu varsayılır (Crocker ve Algina, 1986).





Şekil 1. Madde karakteristik eğrisi

MKE'ye göre bir maddeyi doğru cevaplama olasılığı ile öğrencilerin yetenek düzeyi arasındaki ilişki yorumlanabilmektedir. Şekil 1'de de görüldüğü gibi örtük özelliğin düzeyi arttıkça maddenin doğru cevaplanma olasılığı da artmaktadır (Crocker ve Algina,1986; Hambleton vd., 1991).

MTK'de madde karakteristik fonksiyonundan faydalanarak kestirilen üç madde parametresi (örneğin üç parametrelili lojistik modelde) vardır. Bunlardan biri KTK'deki madde güçlük indeksine karşılık gelen b parametresidir ve örtük özellik ( $\theta$ ) ile aynı ölçekte kestirilir. b parametresi; c parametresi 0'a eşit olduğunda MKE'de bir sorunun doğru cevaplama olasılığının 0,5 olduğu noktanın karşılık geldiği yetenek düzeyi ( $\theta$ ) dir (Hambleton vd., 1991; De Mars, 2010). Grubun yetenek düzeyleri ortalama 0 standart sapma 1 olan göreceli bir dağılıma dönüştürüldüğünde b parametresi teorik olarak  $-\infty$  ile  $+\infty$  arasında değer alırken, pratikte genellikle -2 ve +2 arasında değer almaktadır ve bu değer -2'den +2'ye gittikçe madde zorlaşmaktadır (Hambleton vd., 1991).

MTK'deki diğer bir parametre de a parametresidir ve KTK'deki madde ayırt ediciliğine karşılık gelmektedir. MTK a parametresi MKE'nin b parametresindeki eğimidir ve eğim yetenek düzeyi arttığında maddeyi doğru yanıtlanma olasılığının nasıl değiştiğini gösterir (De Mars, 2010). Bu parametre MKE'de yetenek ( $\theta$ ) ekseninde b parametresinin denk geldiği noktadaki eğim hesaplanarak kestirilir (Hambleton vd., 1991; De Mars, 2010). Sınırlı bir yetenek ranjı için MKE'nin eğiminin dik olduğu maddeler farklı yetenek

düzeyindeki bireyleri daha iyi ayırt etmektedir. Madde ayırt ediciliğini temsil eden a parametresi teorik olarak  $-\infty$  ile  $+\infty$  arasında tanımlanmıştır. a parametresinin negatif olduğu maddeler testlerden çıkarılır, çünkü bu maddelerde yetenek düzeyi arttıkça doğru cevaplama olasılığı azalmaktadır. a parametresinin pozitif olduğu maddelerde ise bu değer genel olarak 2'yi geçmediği için pratikte 0 ile 2 aralığındadır (Hambleton vd., 1991).

Kuramın diğer bir parametresi en düşük yetenek düzeyindekilerin soruyu doğru cevaplama olasılığını belirten ve şans parametresi (psuedo guessing parameter) anlamına da gelen c parametresidir. Bu parametre MKE'de  $P(\theta)$  ekseninin sıfırdan farklı en düşük asimptotudur (Hambleton vd., 1991; De Mars,2010). MTK c parametresi teorik olarak 0 ile 1 arasında tanımlanmıştır ancak pratikte 0,35'i geçmediği için 0 ile 0,35 arasında olduğu söylenebilir (Baker, 2001).

Aşağıda iki kategorili maddeler için madde tepki kuramının lojistik modelleri kısaca açıklanmıştır.

### **Madde Tepki Kuramı Modelleri (Lojistik)**

Lojistik modeller madde karakteristik fonksiyonunun tanımında kullanılan madde parametrelerinin sayısına göre genelde bir, iki, üç ve dört parametrelili lojistik model olmak üzere dört grupta toplanmıştır. Bir parametrelili lojistik model (1PLM) b (madde güçlük) parametresinden oluşmaktadır. Bu modelde bütün maddelerin a parametresinin (ayırt edicilik gücünün) aynı olduğu ve c parametresinin (şans parametresi) 0 olduğu varsayılır. 1PLM'nin bu varsayımlarını sağlamak oldukça zor olduğu için bu model çok kullanışlı değildir. 1PLM için MKE eşitliği aşağıdaki gibidir.

$$P_j(\theta_i) = \frac{e^{(\theta_i - b_j)}}{1 + e^{(\theta_i - b_j)}}$$

İki parametrelili lojistik model (2PLM) b parametresi ile a parametresini kullanmaktadır. Birnbaum tarafından geliştirilen 2PLM için MKE eşitliği aşağıdaki gibidir.

$$P_j(\theta_i) = \frac{e^{Da_j(\theta_i - b_j)}}{1 + e^{Da_j(\theta_i - b_j)}}$$

Üç parametrelili lojistik modelde (3PLM) ise iki parametrelili modele c parametresi eklenmektedir. 3PLM için matematiksel ifade ise

$$P_j(\theta_i) = c_j + (1 - c_j) \cdot \frac{e^{Da_j(\theta_i - b_j)}}{1 + e^{Da_j(\theta_i - b_j)}}$$

şeklindedir (Hambleton vd.,1991). Bu eşitliklerde

$P_j(\theta_i)$ :  $\Theta_i$  yetenek düzeyindeki bir bireyin j maddesini doğru yanıtlama olasılığını,

$a_j$ : j maddesi için madde ayırt edicilik parametresini,

$b_j$ : j maddesi için madde güçlük parametresini,

$c_j$ : j maddesi için şans parametresini,

D: ölçekleme faktörünü (D=1,7) ifade etmektedir.

Bu modeller iki kategorili puanlanan (dichotomus) maddeler için geliştirilmiş modeller olmakla beraber çok kategorili puanlanan (polytomous) maddeler için farklı MTK modelleri bulunmaktadır.

Aşağıda çok kategorili puanlanan maddeler için madde tepki kuramı modelleri açıklanmaktadır.

### **Çok Kategorili MTK Modelleri**

Çok kategorili puanlanan maddeler için kullanılan modeller sıralı cevap kategorilerine sahip ve sahip olmayan modeller olmak üzere ikiye ayrılır. Sıralı cevap kategorileri için kullanılan modeller, Samejima'nın (1969) derecelendirilmiş tepki modeli (DTM), DTM'nin varyasyonu olan Master ve Wright'ın (1982) kısmi puanlama modeli (KPM) ve Muraki'nin (1992) genelleştirilmiş kısmi puanlama modeli (GKPM)'dir. Bock'un sınıflamalı tepki modeli (STM) ve Thissen ve Stenberg'in çok seçenekli maddeler için tepki modeli ise sıralı cevap kategorilerine sahip olmayan maddelerde kullanılmaktadır (van der Linden ve Hambleton, 1996). Sıralı cevap kategorilerine sahip olan modeller rubrik aracılığıyla puanlanan sunum, portfolyo, kompozisyon gibi ürünler ve likert tipi maddelere uygundur (De Mars, 2010).

Aşağıda çok kategorili puanlanan maddeler için madde tepki kuramı temellerine dayanarak Samejima (1969) tarafından geliştirilen "Derecelendirilmiş Tepki Modeli" kısaca açıklanmıştır.

### Samejima'nın Derecelendirilmiş Tepki Modeli

Rasch tipi modellerden farklı olan çok kategorili MTK modellemesi için tek büyük yaklaşım Samejima'nın çalışmasıdır. Samejima'nın DTM'si sıralı cevap kategorilerine sahip olan modellerden en çok bilineni ve kullanılanıdır. Alan yazında da likert tipi maddelerin MTK ile analizlerinde DTM'nin kullanıldığı sıklıkla görülmektedir. Bu çalışmanın temelleri Thurstone'un kümülatif sınır ölçümüne dayanmaktadır (Ostini ve Nering, 2006). Ayrıca Ostini ve Nering (2010) DTM'nin gelişiminin sırayla Lord ve Novick (1968) tarafından psikolojik tepki süreci için geliştirilen normal ogive modelin teorik gelişimine ve Berkson'un (1944, 1953) normal ogive modelin yerine lojistik fonksiyonların daha kullanışlı olabildiğini önermesiyle Birnbaum'un (1968) test kuramı için yeniden yaptığı tekrarlara dayandığını ifade etmiştir. Bu süreç Lord'un (1952) kümülatif normal dağılıma dayanarak oluşturduğu iki parametrelili normal ogive madde tepki modeli ile başlamıştır. Birnbaum (1968) ise bu modeli matematiksel olarak iki parametrelili normal ogive fonksiyonundan daha uygulanabilir olan iki parametrelili lojistik fonksiyonuna dönüştürmüştür (Hambleton ve Swaminathan, 1985; Hambleton vd., 1991; Ostini ve Nering, 2010). Samejima (1969) ise çok kategorili maddeler için iki parametrelili normal ogive modeli ve iki parametrelili lojistik modeli geliştirmiştir. Çok kategorili maddeler için geliştirilen iki parametrelili normal ogive modelde  $P^*_{x_g}(\theta)$

$$P^*_{x_g}(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_g(\theta - b_{x_g})} e^{-t^2/2} dt$$

formülüyle hesaplanır. Formülde  $a_g$  madde ayırt edicilik gücünü, her kategori sınırı ( $x_g$ ) için  $b_{x_g}$  madde güçlüğüne ifade etmektedir. Lojistik modelde ise  $P^*_{x_g}$

$$P^*_{x_g} = \{1 + e^{-Da_g(\theta - b_{x_g})}\}^{-1}$$

formülüyle hesaplanır. Bu formülde de  $a_g$  madde ayırt edicilik gücünü, her kategori sınırı ( $x_g$ ) için  $b_{x_g}$  madde güçlüğüne ifade etmektedir ve D ölçekleme faktörüdür. Lojistik modelde  $x_{g+1}$  kategori sınırı için ise  $P^*_{x_{g+1}}$

$$P^*_{x_{g+1}} = \{1 + e^{-Da_g(\theta - b_{x_{g+1}})}\}^{-1}$$

formülüyle hesaplanır ve  $x_{g+1}$  ve  $x_g$  kategori sınırları için madde güçlüğü parametreleri arasında  $b_{x_{g+1}} > b_{x_g}$  şeklinde bir ilişki vardır (Samejima, 1969).

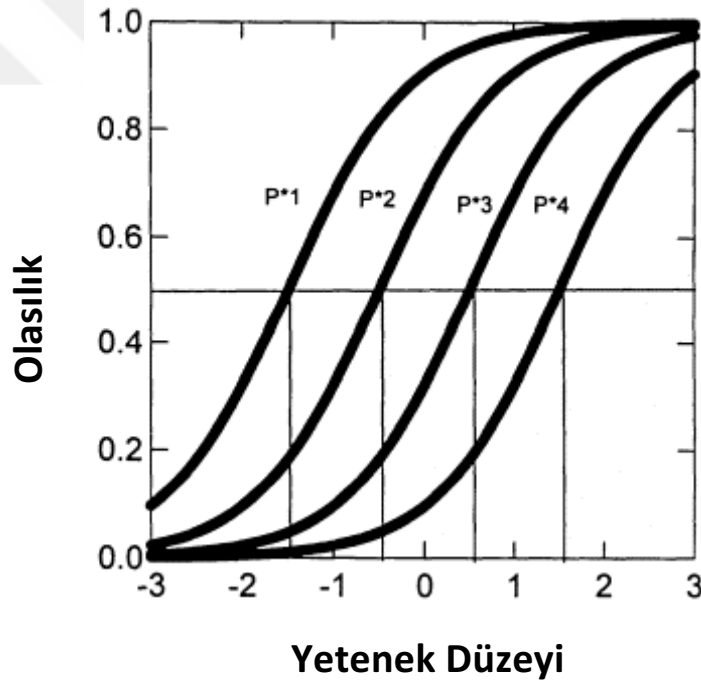
DTM, bir maddenin  $m+1$  kategorisini  $m$  parçaya bölmektedir. Bu parçalardan her biri iki kategorili maddelerde geçerli olan madde tepki fonksiyonları biçiminde ele alınmaktadır.

Bir i bireyinin j maddesine ait bir k kategorisini seçme olasılığı aşağıdaki eşitlikle hesaplanmaktadır;

$$P_{kj}(\theta_i) = P(k_{ij}) - P(k_{ij} + 1) .$$

Burada  $P(k_{ij})$  bir i bireyinin k ve k'nın üzerindeki bir kategoriye seçme olasılığı olarak ifade edilmektedir. DTM'de en düşük kategoriye ve bunun üzerindeki bir kategoriye seçme olasılığı 1 olarak tanımlanmıştır. DTM'nin önemli bir özelliği de kullanılan ölçeğin her bir maddesi için a parametresinin kestirilmesine imkân verilmesidir (Park'tan aktaran Nartgün, 2002).

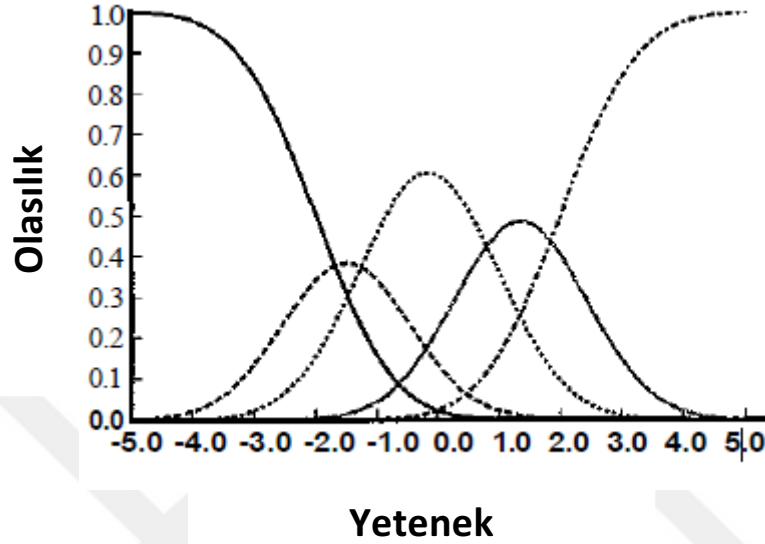
DTM'de her bir kategori için, o kategoriye veya daha üst bir kategoriye seçme olasılığı modellenmiştir ve yine her bir kategori için MKE oluşturulmuştur. Bunun bir örneği Şekil 2.'de verilmiştir;



Şekil 2. Beş kategorili bir madde için işlevsel karakteristik eğrisi (Embretson ve Reise, 2000)

Böyle bir grafik 0-4 arasında puanlanan bir maddeye aittir. Maddede 0 veya daha yüksek bir kategoriye seçme olasılığı "1" olduğu için grafikte 1'den 4'e kadar olan kategorilerin eğrileri verilmiştir. P\*1 eğrisi 1 veya daha yüksek kategoriye seçme olasılığını, P\*2 eğrisi 2 veya daha yüksek bir kategoriye seçme olasılığını, P\*3 eğrisi 3. kategoriye seçme olasılığını, P\*4 eğrisi ise 4. kategoriye seçme olasılığını kestirmede kullanılır. Bu eğriler

birbirine paraleldir ve bu eğrilerden (k veya daha yüksek bir kategoriye seçme olasılığında), k kategorisini seçme olasılığı hesaplanabilir. Bu olasılıklar Şekil 3'teki eğrilerle gösterilebilir. Bu şekil 0'da dâhil beş ayrı kategorinin fonksiyonlarını içermektedir (De Mars, 2010).



Şekil 3. 5 kategorili bir madde için kategori tepki fonksiyonları (0-4 puanlanan madde için belli bir kategoriye ve ondan daha yüksek kategorileri seçme olasılığı) (Samejima, 1996)

Samejima (1969) bu fonksiyonlara “işlevsel karakteristik (operating characteristics)” adını vermiştir. Ayala (2009)'da aynı biçimde Şekil 3'e “işlevsel karakteristik eğrisi (operating characteristics curve (OCC))” demiştir, Şekil 2 için ise “kategori karakteristik eğrisi (category characteristic curve (CCC))” terimini kullanmıştır. Samejima (1996), Embretson ve Reise (2000) ve Ostini ve Nering (2010) Şekil 2'yi “işlevsel karakteristik eğrisi” terimiyle adlandırırken, Şekil 3'ü Embretson ve Reise (2000) ve Ostini ve Nering (2010) “kategori tepki eğrisi (category response curve (CRC))” terimiyle Samejima (1996) ise “kategori tepki fonksiyonu (category response functions (CRF))” terimiyle adlandırmıştır. Şekil 2 için “işlevsel karakteristik eğrisi (İKE)”, Şekil 3 için ise “kategori tepki fonksiyonu (KTF)” terimleri kullanılabilir.

Aşağıda ölçme türlerinden ve ölçme kesinliğini belirlemek için kullanılan istatistiklerden bahsedilmiştir.

## Ölçme Kesinliği

Ölçmenin doğrudan ya da dolaylı olarak yapılmasına göre ölçme türleri temel ölçme ve dolaylı ölçme olmak üzere ikiye ayrılmıştır. Ölçülecek olan değişkenin doğrudan gözlenemediği durumlarda dolaylı ölçme yapılmaktadır (Croker ve Algina, 1986). Psikolojik yapılar doğrudan ölçülemediği için dolaylı olarak ölçülür. Tüm ölçme türlerinde ölçme sonuçlarına çeşitli hatalar karışmaktadır. En duyarlı araçlarla ve en doğru sanılan yöntemlerle yapılan ölçmelerde bile bir miktar hata bulunmaktadır (Turgut ve Baykul, 2010). Buna dayanarak "Her bir maddede veya testte ölçme kesinliğinin derecesi vardır" denilebilir. MTK'da parametre iyileştirme çalışmalarında bu ölçme kesinliği dereceleri "hata kareleri ortalamasının karekökü (root mean squared error (RMSE))" ve " ortalama mutlak farklılık (average absolute difference (AAD))" değerlerine göre yorumlanabilir. Bu değerler arttıkça ölçme kesinliği azalmakta, bu değerler azaldıkça ise ölçme kesinliği artmaktadır. RMSE her madde için kestirilen parametreler ile gerçek parametreler arasındaki farklılığın karelerinin toplamının madde sayısına bölünerek karekökünün alınmasıyla elde edilen bir istatistiktir. AAD ise kestirilen madde parametresi ile gerçek madde parametresi arasındaki farklılığın mutlak değerlerinin toplamının madde sayısına bölünmesiyle elde edilebilir.

Tüm MTK modellerinde sayıtlar ihlal edildiğinde ölçme kesinliğinin azalacağı düşünülmektedir. Günümüzde likert tipi ölçeklerin ve MTK'nin kullanımının artması hangi koşullarda DTM çerçevesinde analizler yapılabileceği sorusunu meydana getirmiştir. Buna ilişkin alan yazın taraması yapıldığında genellikle örneklem büyüklüğünün, test uzunluğunun, kestirim yönteminin, paket programların ve normalliğin ölçme kesinliğine etkisi incelenmiştir. DTM'de farklı dağılım türlerinin ölçme kesinliğini nasıl etkilediğini inceleyen araştırmalara bakıldığında dağılım türleri; Reise ve Yu (1990), Lautenschlager, Meade ve Kim (2006) ve Kieftenbeld ve Natesan (2012)'ın çalışmasında normal,  $\text{ÇK} = 1,25$  ve tekdüze, Ankenmann ve Stone (1992)'un çalışmasında normal ve  $\text{ÇK} = 0,75$  olan dağılım, Bahry (2012)'nin çalışmasında ise normal,  $\text{ÇK} = 0,5$  ve  $\text{ÇK} = 1,00$  olan dağılım şeklindedir. Bu araştırmada ise diğer çalışmaların kullandığı dağılım türlerinden farklı bir şekilde normalliğin ihlal edildiği durumlar olarak çarpıklık katsayısı 1,50 ve 2,00 olan dağılım türlerinde ölçme kesinliğinin nasıl etkilendiği araştırılmıştır. Ayrıca diğer araştırmalara bakıldığında araştırmacıların normalliğin ihlalinin marjinal güvenilirlik katsayısını, MULTILOG programıyla yapılan kestirimlerde programın verdiği hata sayısını nasıl etkilediğini incelemedikleri görülmüştür ve bunlar da bu araştırmanın problemi olarak

incelenmiştir. Bu doğrultuda “Derecelendirilmiş Tepki Modeline dayalı parametre kestiriminde normalliğin ihlali ( $\text{ÇK} = 1,50$  ve  $2,00$  olan dağılımlar) ölçme kesinliğini nasıl etkiler?” sorusuna cevap aranmıştır.

### **Araştırmanın Amacı**

Bu araştırmanın amacı, çok kategorili puanlanan maddelerden oluşan Beck Depresyon Envanterinin Samejima'nın (1969) derecelendirilmiş tepki modeli (DTM) ile analizinde normallik sayılıtısının ihlalinin ölçme kesinliğine etkisini belirlemektir. Bu amaç doğrultusunda aşağıdaki alt problemlere yanıt aranmıştır:

1. Alt Problem: Samejima'nın Derecelendirilmiş Tepki Modeli (DTM) ile madde parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı ( $\text{ÇK}$ ) =  $1,50$  ve  $2,00$  olan dağılımlar) hata kareleri ortalamasının karekökünü (RMSE) nasıl etkiler?

1.1. Samejima'nın Derecelendirilmiş Tepki Modeli ile a parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı ( $\text{ÇK}$ ) =  $1,50$  ve  $2,00$  olan dağılımlar) hata kareleri ortalamasının karekökünü (RMSE) nasıl etkiler?

1.2. Samejima'nın Derecelendirilmiş Tepki Modeli ile  $b_1$ ,  $b_2$ ,  $b_3$  parametrelerinin kestiriminde normalliğin ihlali (çarpıklık katsayısı ( $\text{ÇK}$ ) =  $1,50$  ve  $2,00$  olan dağılımlar) hata kareleri ortalamasının karekökünü (RMSE) nasıl etkiler?

2. Alt Problem: Samejima'nın Derecelendirilmiş Tepki Modeli ile madde parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı ( $\text{ÇK}$ ) =  $1,50$  ve  $2,00$  olan dağılımlar) ortalama mutlak farklılığı (AAD) nasıl etkiler?

2.1. Samejima'nın Derecelendirilmiş Tepki Modeli ile a parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı ( $\text{ÇK}$ ) =  $1,50$  ve  $2,00$  olan dağılımlar) ortalama mutlak farklılığı (AAD) nasıl etkiler?

2.2. Samejima'nın Derecelendirilmiş Tepki Modeli ile  $b_1$ ,  $b_2$ ,  $b_3$  parametrelerinin kestiriminde normalliğin ihlali (çarpıklık katsayısı ( $\text{ÇK}$ ) =  $1,50$  ve  $2,00$  olan dağılımlar) ortalama mutlak farklılığı (AAD) nasıl etkiler?

3. Alt Problem: Samejima'nın Derecelendirilmiş Tepki Modeli ile madde parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı ( $\text{ÇK}$ ) =  $1,50$  ve  $2,00$  olan dağılımlar) marjinal güvenirliliği nasıl etkiler?



4. Alt Problem: Samejima'nın Derecelendirilmiş Tepki Modeli ile madde parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı (ÇK) = 1,50 ve 2,00 olan dağılımlar) EM cycle yakınsaklık kriterini sağlamama yüzdesini nasıl etkiler?

5. Alt Problem: Samejima'nın Derecelendirilmiş Tepki Modeli ile madde parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı (ÇK) = 1,50 ve 2,00 olan dağılımlar) MULTILOG programının verdiği hata sayısını nasıl etkiler?

### **Araştırmanın Önemi**

Likert tipi ölçekler araştırmalarda sıklıkla kullanılan ölçek türlerindedir ve son zamanlarda bu ölçeklerle toplanan verilerin madde tepki kuramının modelleri ile analiz edildiği çalışmalar artmaktadır. Ancak MTK'nin sayıtlarının sağlanması ve örneklemin büyük olması gerekmektedir. Normal ogive madde tepki modellerinde MTK'nin sağlanması gereken sayıtlarından biri de normallik varsayımdır ve bazı çalışma gruplarında normallik varsayımını sağlamak oldukça zordur. Bu araştırma likert tipi ölçeklerde normallikten farklı düzeylerde sapıldığında ölçme kesinliğinin bundan nasıl etkilendiğine cevap olacağı için önemlidir. Özellikle normalliğin sayıtlı olarak öngörülmediği lojistik MTK modellerinde normallik ihlalinin ölçme kesinliğini nasıl etkilediğini belirlemek önemlidir. Çünkü pek çok araştırmacı ve ölçme uzmanı lojistik modellerde normallik sayıtlısını kontrol etmeye ihtiyaç duymamaktadır. Bu araştırma lojistik derecelendirilmiş tepki modelinde normalliğin bir sayıtlı olup olmayacağına sorusuna yönelik bir cevap niteliğindedir. Bunun yanı sıra başka araştırmacılar likert tipi bir ölçek ile toplanan verilerini DTM ile analiz etmeden önce bu çalışmanın sonuçlarına ve kendi dağılımlarının çarpıklık katsayısına bakarak yaptıkları ölçümlerin kesinliği hakkında tahmin yürütebilirler. Ayrıca araştırmacılar çalışmalarını için kabul edilebilir olan ölçme kesinlik düzeylerine göre örneklemlerinin dağılımlarının çarpıklık katsayısının yaklaşık olarak ne olması gerektiğini tahmin edebilirler. Son olarak bu araştırma DTM'ye göre yapılan kestirimler için normallik ihlal edildiğinde AAD'nin, marjinal güvenilirlik katsayısının, programın verdiği hata sayısının incelemesi açısından bir ilktir.

### **Sayıtlılar**

Bireyin BDE'ye verdikleri yanıtlar gerçek duygu ve düşüncelerini yansıttığı varsayılmıştır.

### **Sınırlılıklar**

1. Arařtırmada sadece 0,00; 0,50; 1,00; 1,50 ve 2,00 olan dađılımları incelenmiřtir.
2. Arařtırmada sadece normalliđin ihlali durumlarıyla sınırlıdır.
3. Arařtırma Samejima'nın (1969) derecelendirilmiř tepki modeline gre MULTILOG 7.03 programı ile yapılan kestirimlerle sınırlıdır.
4. Arařtırma MML kestirim yntemi ile sınırlıdır.
5. Arařtırmada sadece madde parametreleri kestirimi iin lme kesinliđinin incelenmesiyle sınırlıdır.



## BÖLÜM II

### İLGİLİ ARAŞTIRMALAR

Bu bölümde bu çalışma ile ilgili yurt içinde ve yurt dışında yapılan araştırmalara özet olarak yer verilmiştir.

Swaminathan ve Gifford (1979) yaptığı araştırmada 3PLM'de Urry'nin yöntemi ile MLE kestirim yöntemini karşılaştırmak ve farklı test uzunluklarının, örneklem büyüklüklerinin, dağılım türlerinin birey ve madde parametresi kestirimlerinin doğruluğuna etkisini incelemek amacıyla DATGEN programında simülasyon veri üretmiştir. Araştırmada 10, 15, 20 ve 80 madde olmak üzere dört test uzunluğunda, 50, 200 ve 1000 kişi olmak üzere üç örneklem büyüklüğünde, normal (0,1) , tekdüze (-1,73; +1,73 arası) ve negatif çarpık (5; 1,5) olmak üzere üç dağılım türünde olmak üzere toplam 36 koşul bulunmaktadır. Urry 'nin yöntemiyle kestirim yapabilmek için ANCILLES programını, MLE kestirimleri için ise LOGIST programını kullanmıştır. Kestirimlerin doğruluğunu incelemek için ise bias ve kestirim sonuçlarının tutarlılığı gibi istatistiksel özellikleri hesaplamıştır. Araştırmanın sonucunda a ve b parametresi kestirimlerinin çarpık dağılımda tekdüze ya da normal dağılıma göre daha problemli olduğunu saptamıştır.

Ree (1979) yaptığı simülasyon çalışmasında madde karakteristik eğrisi kestirimi yapan dört programının etkililiğini araştırmayı amaçlayarak her biri 2000 birey ve 80 maddeden oluşan farklı dağılımlarda veri üretmiştir. Bu dağılımlar tekdüze (-2,5; +2,5 arası), çarpık ( $\text{ÇK} = 0,64$ ,  $\text{BK} = 2,73$ ) ve normal dağılım (0,01; 1,01) olmak üzere 3 tanedir. Her bir dağılım için ANCILLES, LOGIST, OGIVIA programlarında ve dönüştürme prosedürüyle 3PLM'e göre elde ettiği madde karakteristik eğrilerini karşılaştırmanın yanı sıra her dağılımın a, b, c ve  $\theta$  parametrelerinin gerçek ve kestirilen değeri arasındaki korelasyonları incelemiştir. Araştırmanın sonucunda dört madde karakteristik eğrisi elde etme yolunda da a, b ve c parametreleri için korelasyonların çarpık dağılımda en düşük olduğunu ve a ve c

parametreleri için tekdüze dağılıma ait korelasyon normal dağılıma ait korelasyondan daha yüksekken b parametrelerinde normal dağılıma ait korelasyonların daha yüksek olduğunu bulmuştur.  $\theta$  parametresi içinde en düşük korelasyon diğer dağılımlardan çok farklı olmamasına karşın yine çarpık dağılıma aittir, en yüksek korelasyon ise tekdüze dağılım için gözlenmiştir. Çarpık dağılıma ilişkin korelasyonların normal ve tekdüze dağılıma ilişkin korelasyonlardan daha düşük (özellikle a ve c parametreleri için diğer dağılımlardan çok fazla düşük) olması nedeniyle çarpık dağılımdaki kestirimlerin daha zayıf olduğu gözlenmiştir.

Reise ve Yu (1990) yaptıkları parametre iyileştirme çalışmasında MML temelli MULTILOG programının kapasitesini araştırmak için Samejima'nın derecelendirilmiş tepki modelinde örneklem büyüklüğünün, gerçek  $\theta$  dağılımı türünün ve gerçek madde ayırt ediciliği dağılımının madde ve birey parametrelerinin kestirilen ve gerçek değerleri arasındaki korelasyona, RMSE'ye ve bias'a etkisini incelemişlerdir. Bu doğrultuda örneklem büyüklüğü (N) = 250, 500, 1000 ve 2000; dağılım türü = normal(0,1), tekdüze(-3,+3), çarpık(ÇK = 1,25, BK = 1,50) ve gerçek  $\alpha$  dağılımı (0,44 - 0,75; 0,58 - 0,98; 0,75 - 1,33) olmak üzere üç faktöre dayalı toplam 36 koşul için Monte Carlo simülasyon verisi üretmişlerdir. Tüm koşullarda test beş kategoriden ve 25 maddeden oluşmaktadır ve tüm maddelerin b parametreleri  $\beta_1 = -2$  ile  $-1$  aralığında,  $\beta_2 = -1$  ile  $0$  aralığında,  $\beta_3 = 0$  ile  $1$  aralığında,  $\beta_4 = 1$  ile  $2$  aralığında değişmektedir. Tek düze dağılımın ortalaması  $0$ , standart sapması  $1,7$  olduğu için bu dağılım türünde yeniden ölçekleme işlemi yapılmıştır; normal ve çarpık dağılımın ortalaması  $0$ , standart sapması  $1$  olduğu için bu dağılım türlerinde yeniden ölçeklendirme yapmaya ihtiyaç duymamışlardır. Araştırmanın sonucunda a parametresinde en yüksek korelasyonun tekdüze ( $0,92$ ), en düşük korelasyonun ise çarpık ( $0,85$ ) dağılımda olduğunu; korelasyonla benzer olarak en düşük RMSE'nin tekdüze ( $0,07$ ) dağılımda, en yüksek RMSE'nin ise normal ( $0,09$ ) ve çarpık ( $0,09$ ) dağılımda olduğunu bularak tekdüze dağılımın normal ve çarpık dağılımdan biraz daha iyi olduğunu ifade etmişlerdir. MTK b parametreleri için,  $\beta_2$  ve  $\beta_3$ 'ün gerçek ve kestirilen değerlerinin arasındaki korelasyonlarının  $\beta_1$  ve  $\beta_4$ 'ten daha yüksek RMSE'lerinin ise daha düşük olduğunu gözlemlemişlerdir. Ayrıca b parametreleri için en yüksek korelasyonun tekdüze ( $0,92$ ), en düşük korelasyonun ise normal ( $0,85$ ) ve çarpık ( $0,85$ ) dağılımda olduğunu; korelasyonla benzer olarak en düşük RMSE'nin tekdüze ( $0,17$ ) dağılımda, en yüksek RMSE'nin ise normal ( $0,20$ ) ve çarpık ( $0,20$ ) dağılımda olduğunu bulmuşlardır. Birey parametresine bakıldığında ise a ve b parametrelerinin aksine en yüksek RMSE'nin tekdüze

(0,47) dağılımda en düşük RMSE'nin ise normal (0,44) dağılımda olduğunu bulmuşlardır, korelasyona bakıldığında RMSE'nin aksine en yüksek korelasyonun tekdüze (0,96) dağılımda olduğunu ve en düşük korelasyonun normal ile çarpık (0,90) dağılımda olup eşit olduğunu gözlemlemişlerdir.

Seong (1990) araştırmasında önsel (prior)  $\theta$  dağılımlar temel  $\theta$  dağılımları ile uyuşmadığında madde ve yetenek ( $\theta$ ) parametrelerinin marjinal maksimum olabilirlik kestiriminin (MMLE) duyarlılığını incelemek için temel  $\theta$  dağılımının üç türünün her biri için 30'ar tane 45 maddelik veri seti üretmiştir. Daha sonra bu veri setlerini 2PLM'ye dayanarak PC-BILOG 1.1 ile analiz etmiştir. Araştırmanın sonucunda örneklem büyüklüğü arttığında madde ve yetenek parametreleri kestirimlerinin doğruluğunun arttığı bulunmuştur. Hem temel  $\theta$  dağılımını hem de önsel dağılımı çeşitlendiren Seong (1990) önsel dağılımda temel  $\theta$  dağılımına göre a ve b parametrelerinin kestiriminin daha düşük bias ve RMSE'ye sahip olduğunu yani daha iyi olduğunu gözlemlemiştir ayrıca yetenek kestiriminin daha fazla etkilendiğini bulmuştur ve bunun birey parametresi kestirim yöntemi olarak EAP kullanmasından kaynaklandığı düşünülmüştür. Çünkü EAP yönteminde yetenek parametreleri önsel dağılımdan doğrudan etkilenmektedir.

Ankenmann ve Stone (1992) yaptıkları çok replikasyonlu Monte Carlo çalışmasında QUASAR Cognitive Assessment Instrument (QCAI)'in 10 maddelik 5 puan düzeyli matematik testini kullanarak test uzunluğunun, örneklem büyüklüğünün ve dağılım türünün 1PL derecelendirilmiş model ve 2PL derecelendirilmiş modelde madde ve birey parametresi kestiriminin doğruluğuna etkisini incelemişlerdir. Bu doğrultuda 1PL derecelendirilmiş model için test uzunluğu 5, 10 ve 20; örneklem büyüklüğü 125, 250 ve 500; dağılım türü normal ( $\bar{X} = 0$ ,  $S_x = 1$ ) ve çarpık ( $\text{ÇK} = 0,75$ ,  $\text{BK} = 0,00$ ) olmak üzere 18 koşul; 2PL derecelendirilmiş model için de test uzunluğu 5,10 ve 20; örneklem büyüklüğü 250, 500 ve 1000; dağılım türü normal ve çarpık ( $\text{ÇK} = 0,75$ ,  $\text{BK} = 0,00$ ) olmak üzere 18 koşul oluşturulmuş ve her iki modelde de 100 replikasyon yapılmıştır. Normal yetenek dağılımını IMSL'nin RNNOA fonksiyonuyla, çarpık yetenek dağılımını ise normal dağılımdan Fleishman'in (1978) güç yöntemini kullanarak türetmişlerdir. Araştırmanın sonucunda 2PL derecelendirilmiş modelde  $a_j$  parametresinin kestiriminde dağılımın normal ya da çarpık olmasının RMSE üzerinde etkisi yokken, bias üzerinde önemsenmeyecek kadar az bir etkisi olduğu;  $b_{ij}$  parametresinin kestiriminde dağılım türünün RMSE üzerinde etkisi yokken, çarpık dağılımda bias'ın normal dağılıma göre daha yüksek olduğu; birey

parametrelerinin kestiriminde ise dağılım türünün RMSE ve bias üzerinde etkisi olmadığını bulmuşlardır.

Stone (1992) 2PLM'de MML kestirimlerini iyileştirmek amacıyla yaptığı çalışmada analizleri MULTILOG programıyla yapmıştır. Ölçme kesinliğine örneklem büyüklüğünün, test uzunluğunun ve dağılım türünün etkisini incelemek amacıyla GENIRV'de 100 replikasyon yaparak veri üretmiştir. Veri üretme aşamasında 2PLM'ye dayanarak 20 maddeli matematik başarı testinin kalibre edilmiş madde parametrelerini kullanmışlardır ve üç örneklem büyüklüğünde (250, 500, 1000), üç test uzunluğunda (10, 20, 40), üç dağılım türünde ( $N(0,1)$ , çarpık(0,75;0), basık (0;-1,0)) olmak üzere 27 koşulda veri üretmişlerdir. Farklı test uzunluğunda veri üretirken 20 maddeden çift maddeler seçilerek 10 maddelik test, 20 madde iki defa kullanarak 40 maddelik test oluşturulmuştur. Çarpık ve basık dağılımdaki verileri üretirken ise normal dağılımdan Fleishman'in (1978) güç yöntemini kullanarak üretmiştir. Koşulların ölçme kesinliğine etkisini yorumlayabilmek için madde ve test düzeyinde RMSE, bias hesaplamıştır ve iterasyon sayısını raporlamıştır. Araştırmanın sonucunda normal olmayan dağılımların iterasyon sayısına küçük bir etkisi olduğu görülmüştür. MTK a parametresi için test düzeyinde bias'ın dağılım türünden etkilendiği gözlenirken, RMSE'nin nispeten daha az etkilendiği gözlenmiştir. 40 maddeli test hariç test düzeyinde bias ve RMSE'nin a ve b parametresi için en yüksek değerlerin çarpık, en düşük değerlerin ise normal dağılımda olduğu bulunmuştur. Sonuç olarak gerçek çarpık ve basık  $\Theta$  dağılımlarda a ve b parametrelerinin kestiriminin daha hatalı olduğu görülmüştür.

Gelbal (1994) yaptığı çalışmada MTK'nin Rasch modeliyle ve KTK ile kestirilen madde ve yetenek parametrelerinin benzer yönlerini ortaya koymayı amaçlamıştır. Bu amaç doğrultusunda "İlköğretimde temel öğrenme ihtiyaçlarının karşılanması" konulu çalışmada 2072 beşinci sınıf öğrencisine uygulanan Türkçe testi ve 2077 beşinci sınıf öğrencisine uygulanan Matematik testine ait verileri kullanılarak Rasch modelinin  $b_g$  parametresi ile KTK'nin birim standart değerlere dönüştürülen madde güçlük indeksi ( $z_g$ ) ve MTK  $\theta$  parametresi ile KTK test puanları arasındaki korelasyonu incelemiştir. Test puanları normal dağılım göstermediği için Türkçe testinden 506 öğrenciyi, Matematik testinden 521 öğrenciyi çıkarmış ve normal dağılım gösteren iki yeni veri seti elde ederek test puanlarının hem normal dağıldığı hem de normal dağılmadığı durumlarda  $b_g$  ve  $z_g$  arasındaki korelasyonu ve standart hatayı karşılaştırarak normallik sayıltısı ihlalinin MTK

ve KTK ile kestirilen madde parametrelerinin benzerliğine etkisi olup olmadığını gözlemlemiştir. Araştırmanın sonucunda, Türkçe testi için hem normal hem de normal olmayan dağılımda  $b_g$  ve  $z_g$  arasındaki korelasyonu 1,00 bulmuş, standart hatayı ise normal dağılımda 0,033, normal olmayan dağılımda 0,029 bulmuştur. Matematik testi için de hem normal hem normal olmayan dağılımda  $b_g$  ve  $z_g$  arasındaki korelasyonu 1,00 bulmuş, standart hatayı ise normal dağılımda 0,035, normal olmayan dağılımda 0,030 bulmuştur. Her iki dağılımda da  $z_g$  ve  $b_g$  arasındaki ilişkinin aynı bulunmasının, dağılımın normal olup olmamasının bu benzerliği etkilemediğini gösterdiğini ve standart hatanın normal olmayan dağılımda daha az olmasının bu dağılımda örneklem büyüklüğünün daha fazla olması ile açıklanabileceğini belirtmiştir.

Lane, Stone, Ankenmann ve Liu (1995) yaptıkları araştırmada DTM'de testin boyutluluğunun ve sınav süresinin (hızlılığın) madde parametrelerinin sabitliğine etkisini incelemişlerdir ve bunun için matematik performansını değerlendirmek amacıyla geliştirilen QUASAR Cognitive Assessment Instrument (QCAI) alt testlerini kullanmışlardır. Toplam 36 açık uçlu görevden oluşan 4 form (her formda 9 görev) 6. ve 8. Sınıf düzeyindedir ve bu formlar kapsam, içerik ve güçlük açısından birbirine benzemektedir. Bu 4 form dört farklı dönemde (1990-Sonbahar, 1991-İlkbahar, 1991-Sonbahar ve 1992-İlkbahar) uygulanmıştır. 1991-Sonbahar ve 1992-İlkbaharda uygulama yapılan gruplar için süre 45 dakikadır ve formda 9 görev bulunduğu için her görev 5 dakikadır, diğer gruplar için ise 2 görev için süre 20 dakikadır ve her görev 10 dakikadır. Görevleri 5 kategoride puanlayarak, parametre kestirimlerini MULTILOG programı ve MML kestirim yöntemi ile iki parametrelili DTM'ye ve a parametreleri eşit olan iki parametrelili DTM ile yapmışlardır ve dört dönemde elde edilen verilerin bu modellerle model veri uyumunu incelemişlerdir. Testlerin boyutluluğunu LISREL 7 programında doğrulayıcı faktör analizi yaparak incelemişler ve test durumlarını tek boyutlu bulmuşlardır. Araştırmanın sonucunda farklı sürede yapılan testler (5 dakika, 10 dakika) arasında iki görev için a ve b parametrelerinin sabit olmadığını, başka iki görev için ise sadece a parametresinin sabit olmadığını gözlemlemiştir. Ayrıca iki dönemde uygulanan testlerde parametre kestirimleri sabitken, diğer iki dönemde iki görev için hem a hem de b parametrelerinin değişken olduğunu, bunlar dışındaki iki görev için de sadece b parametresinin değişken olduğunu bulmuştur.

Seong, Kim ve Cohen (1997) yaptığı araştırmada yaygın olarak kullanılan üç yetenek kestirim yönteminin (ML, MAP, EAP) ölçme kesinliğine etkisini DTM altında karşılaştırmayı amaçlamıştır. Bu amacın yanı sıra test uzunluğunun, dağılım türünün ve örneklem büyüklüğünün de ölçme kesinliğine etkisini incelemiştir. Bu doğrultuda GENIRV programıyla DTM'ye dayalı olarak iki örneklem büyüklüğü (300, 1000), iki dağılım türü (  $N(0,1)$ ,  $N(1,1)$  ) ve üç test uzunluğu (5, 10, 30) olmak üzere 12 koşul için 10 replikasyon yaparak veri üretmişlerdir. Verilerin üretimi için madde parametrelerini Wisconsin Student Assessment System'in bir kısmı olarak geliştirilen 36 maddeli matematik testinden elde etmişlerdir. Bu 36 maddenin b parametrelerinin ortalamasına ve ortalama standart sapmasına yakın 5, 10 ve 30 maddelik testler oluşturup veri üretiminde bu maddelere ait parametreleri kullanmışlardır. Ayrıca üretilen dağılım türlerinden  $N(1,1)$  dağılımı ortalama test gücüyle eşleştirilmiştir ve bu dağılımın elde edildiği testler daha zordur. Verilerin analiz edilmesi aşamasında birey parametresi kestiriminde ML, MAP, EAP kestirim yöntemleri kullanılırken madde parametresi kestiriminde MMLE kestirim yöntemini kullanmışlardır ve analizleri MULTILOG programıyla yapmışlardır. Ölçme kesinliğini RMSE, bias ve korelasyon hesaplayarak incelemiştir. Modeldeki a ve b parametresi için RMSE'yi eşleştirilmiş dağılımda  $N(0,1)$  dağılımına göre daha düşük, korelasyonu daha yüksek bulmuşlardır. Madde b parametreleri için RMSE'yi  $N(0,1)$  dağılımında  $b_1 < b_2 < b_3 < b_4$  şeklinde sıralanırken  $N(1,1)$  dağılımında  $b_2$  ve  $b_3$ 'te  $b_1$  ve  $b_4$ 'tekine göre daha düşük bulmuşlardır. Bias sonuçlarını incelediklerinde tüm koşullarda düşük olduğunu ve eşleştirilmiş dağılımda  $N(0,1)$  dağılımına göre genellikle daha düşük olmasına rağmen aradaki farkın çok az olduğunu ifade etmişlerdir. Ayrıca RMSE, bias ve korelasyon sonuçlarına göre eşleştirilmiş dağılımda madde parametresi kestiriminin eşleştirilmemiş dağılıma göre daha güçlü olduğunu vurgulamışlardır.

De Ayala ve Bolesta (1999) STM'de parametre iyileştirmeyi çeşitli faktörler açısından incelemiştir. Bu faktörler örneklem büyüklüğü, dağılım türü, kategori sayısı ve madde bilgi düzeyidir ( $I_{max}^s$ ). Araştırmacılar amaçları doğrultusunda iki kategori sayısında (3 seçenek ve 4 seçenek), üç madde bilgi düzeyinde (yüksek (0,25), orta (0,16), düşük (0,09)), üç dağılım türünde ( $N(0,1)$ , pozitif çarpık(1,25;10), tekdüze(-4 ile +4 aralığında)) ve dört örneklem büyüklüğünde olmak üzere 72 koşulda veri üretmiş ve 25 replikasyon yapmışlardır. Burada örneklem büyüklüklerini kategori sayılarına göre belirlemişlerdir ve 3 seçeneikli maddelerin bulunduğu test için örneklem büyüklükleri 420, 840, 1680, 3360 olarak; 4 seçeneikli maddelerin bulunduğu test için ise 560, 1120, 2240, 4480'dir ayrıca tüm



veri setlerinde madde sayısı 28'dir. Yüksek, orta ve düşük madde bilgi düzeyleri sırasıyla 1,0; 0,8 ve 0,6  $\alpha$  ayırt edicilik katsayısına tekabül etmektedir, pozitif çarpık dağılımların çarpıklık katsayıları ise 1,277-1,326 arasında değişmektedir. Verilerin analizi MML madde parametresi kestirimi yapan MULTILOG 5.1 programıyla maksimum iterasyon ve cycle sayısı 999 olarak yapılmıştır. Ölçmenin doğruluğunu yorumlamak için RMSE, bias ve korelasyon hesaplayan araştırmacılar dört yönlü ANOVA yaparak ve  $\omega^2$  değeri hesaplanarak belirlenen faktörlerin etkileri de incelenmiştir. MTK a parametresi için korelasyon katsayıları en yüksek tekdüze dağılımdayken en düşük değer çarpık dağılımda olduğunu, c parametresi için ise tüm dağılım türleri arasındaki farklılık büyük olmamakla beraber genellikle en yüksek tekdüze dağılımda en düşük çarpık dağılımda olduğunu bulmuşlardır. Modeldeki a parametresi için RMSE en düşük değerlerini tekdüze dağılımda en yüksek değerlerini ise çarpık dağılımda alırken c parametresinde yine dağılım türleri arasındaki farklılık büyük olmamakla beraber genel olarak en düşük tekdüze dağılımda en yüksek çarpık dağılımda olduğu görülmüştür. Ayrıca a parametresi için en düşük bias değerleri normal dağılımdayken c parametrelerinde çarpık dağılımın bias değerlerinin örneklem büyüdükçe azaldığını gözlemlemişlerdir. ANOVA sonuçlarını incelediklerinde ise a parametresi için dağılım türlerindeki farklılığı manidar bulmuş ve  $\omega^2$  değerini 0,425; c parametresi için de farklılığı manidar bulup  $\omega^2$  değerini 0,012 bulmuşlardır. Sonuç olarak dağılım türünün a parametresi kestirimlerini önemli derecede etkilerken, c parametrelerinin kestirimlerini etkilemediğini ifade etmişlerdir ve a parametrelerinde en zayıf kestirimlerin çarpık en güçlü kestirimlerin ise tekdüze dağılımda olduğu görülmüştür.

Boughton, Klinger ve Gierl (2001) yaptıkları çalışmada GKPM ve DTM'de randum puanlayıcı hatasının, test uzunluğunun, kategori sayısının ve kullanılan modellerin parametre iyileştirmeye etkisini incelemeyi amaçlamışlardır. Bu amaç doğrultusunda RESGEN 3 programıyla her iki model için üç çeşit kategori sayısında (4, 6, 8), üç test uzunluğunda (4, 8, 16) olmak üzere 18 koşulda 30 replikasyon yaparak veri üretmişlerdir ve bu koşulları puanlayıcı hatasının olduğu ve olmadığı durumlarda kullanmışlardır. Veri üretimi aşamasında a parametresini 0,75 - 1,50 arasında tekdüze dağılımla, b parametrelerini -2,25 ile +2,25 aralığında randum olarak belirlemişlerdir. Parametre iyileştirmesini incelemek için gerçek ve kestirilen parametreler arasındaki RMSE değerlerini hesaplamışlardır. MTK a, b ve  $\Theta$  parametreleri için hem DTM hem GKPM'de test uzunluğu ve kategori sayısının artmasının RMSE'yi düşürdüğünü bulmuşlardır. B

rey ( $\Theta$ ) parametresi için RMSE değerleri hem randum puanlayıcı hatası bulunan hem de bulunmayan durumlarda DTM'de GKPM'ye göre daha yüksek, a parametresi için hata bulunan durumda DTM için RMSE değerinin daha düşük olduğunu, hata olmayan durumda ise farklılık olmadığını, hata olan durumda uç kategorilerin b parametrelerinde DTM için RMSE'nin daha yüksek olduğunu ve hata olmayan durumda b parametresi için farklılık çok olmamakla birlikte GKPM için daha düşük olduğunu gözlemlemişlerdir. Ayrıca uçlardaki b parametrelerinin ortadakilere göre daha büyük RMSE değerlerine sahip olduğunu ifade etmişlerdir. Sonuç olarak random puanlayıcı hatasının bulunmasının ölçme kesinliğini büyük ölçüde olumsuz etkilediği ve kestirimin bu durumda daha zayıf olduğu, test uzunluğunun, kategori sayısının artmasının ölçme kesinliğini olumlu olarak etkilediğini ve kestirimlerin bu durumlarda daha güçlü olduğu kanısına varılmıştır.

Kirisci, Hsu ve Yu (2001) yaptıkları araştırma da test boyutluluğunun, dağılım türünün ve kestirim programının madde ve birey parametreleri kestirimlerinin doğruluğunu nasıl etkilediklerini incelemişlerdir. İncelemek amacıyla üç dağılım türünde (Normal( $\text{ÇK} = 0,0$ ;  $\text{BK} = 0,0$ ), pozitif çarpık( $\text{ÇK} = 0,75$ ,  $\text{BK} = 0,0$ ) ve basık( $\text{ÇK} = 0,0$ ,  $\text{BK} = -1,0$ )) ve iki boyut sayısında (tek boyutlu ve üç boyutlu) olmak üzere altı koşul oluşturmuşlardır. Basık ve pozitif çarpık dağılımları Flesihman'ın güç yöntemiyle oluşturan araştırmacılar parametre kestirimlerini 3PLM'ye dayanarak yapmışlardır. 10 replikasyon yapan araştırmacılar veri setlerini a parametresi 0,4 ile 2,00 aralığında, b parametresi -2,0 ile +2,0 aralığında ve c parametresi 0,0 ile 0,3 aralığında tekdüze olarak dağılacak şekilde üretmişlerdir. Tüm veri setlerinde örneklem büyüklüğünü 1000, madde sayısını 40 olarak belirlemişler ve analizleri MML kestirim yöntemine dayalı olarak BILOG 3, MULTILOG 6.0 ve XCALIBRE 1.10 programlarına göre yapmışlardır. Tüm programlarda maksimum cycle sayısı 50 olarak belirlenirken, maksimum iterasyon sayısı XCALIBRE 1.10'da belirlenemediği için sadece BILOG 3 ve MULTILOG 6.0'da 10 olarak belirlenmiştir. Kestirimlerin doğruluğunu incelemek amacıyla RMSE hesaplamış, faktörlerin etkisini incelemek amacıyla ise logRMSE değerleri için üç yönlü ANOVA yapmışlardır. ANOVA sonuçlarına baktıklarında dağılım türünün ve dağılım türünün diğer faktörlerle etkileşiminin hiçbir parametrenin (a, b, c,  $\theta$ ) kestirim doğruluğuna etkisinin  $p < 0,001$ ,  $\eta^2 > 0,09$  düzeyinde manidar olmadığını gözlemlemişlerdir. Dağılım türünün manidar bir etkisinin olmamasının örneklem büyüklüğünün ve test uzunluğunun büyük olmasından kaynaklanabileceğini ifade etmişlerdir.

De Mars (2002) yaptığı araştırmada Samejima'nın DTM'ye ve GKPM'ye göre birey ve madde parametreleri kestirimlerinin iyileştirilmesinde MULTILOG ve PARSCALE'in doğruluğunu karşılaştırmayı amaçlamıştır. Bunun yanı sıra parametre kestiriminin doğruluğu üzerinde dağılım türünün ve örneklem büyüklüğünün etkisini de incelemiştir. Bu amaçla dağılım türü normal(0, 1), beta(2; 5,5), tekdüze(-1,73; +1,73); örneklem büyüklüğü (N) 250 ve 500 olmak üzere iki faktöre dayalı altı koşulda veri üretmiştir, her bir koşul için 100 replikasyon yapmıştır ve bu replikasyonların her birinin farklı madde parametrelerine sahip olduğunu belirtmiştir. Üretilen testlerin her birini beş kategori ve 10 maddeden oluşturmuştur ve maddelerin  $\beta_1$  parametresi -2 ile -1 aralığında tekdüze bir dağılım göstermekte olup aynı maddedeki ardışık b parametrelerinin arasında 0,33 birim fark bulunduğunu belirtmiştir. Verilerin üretilmesi aşamasında tekdüze ve normal dağılımların ortalaması 0, standart sapması 1 olduğu için yeniden ölçeklendirme yapılmazken, çarpık dağılım 0,267 ile çarpılıp 6,59 çıkarılarak yeniden ölçeklendirilmiştir. Araştırmada GKPM ve DTM olmak üzere iki model, MML kestirimi yapan MULTILOG ve PARSCALE olmak üzere iki program kullanıldığı için toplam 2400 analiz yapılmıştır. De Mars (2002) parametre iyileştirmenin doğruluğunu bias ve RMSE hesaplayarak, MULTILOG ve PARSCALE arasındaki farkı ise RMSD hesaplayarak kontrol etmiştir. Araştırmanın sonucunda MULTILOG ve DTM ile kestirilen a parametresi için 500 kişilik örneklem büyüklüğünde bias'ın en düşük değerini çarpık dağılımda (0,005) en yüksek değerini tekdüze dağılımda (0,031) aldığı, RMSE'nin en düşük değerini tekdüze dağılımda (0,093) en yüksek değerini çarpık dağılımda (0,097) aldığı; 250 kişilik örneklem büyüklüğünde ise bias 500 kişilik örnekleme benzer sonuçlar verirken RMSE en düşük değerini çarpık dağılımda (0,133) en yüksek değerini ise normal dağılımda (0,137) aldığı görülmüştür. Modeldeki b parametreleri için 500 kişilik örnekleme bias incelendiğinde en düşük değeri normal dağılımda (0,001), en yüksek değeri çarpık dağılımda (-0,017) aldığı; RMSE'ninde bias ile paralel olduğu; 250 kişilik örnekleme bias ve RMSE'nin en düşük değeri tekdüze dağılımda en yüksek değeri çarpık dağılımda aldığı bulunmuştur. Ayrıca b parametreleri için genel olarak  $\beta_2$  ve  $\beta_3$ 'ün gerçek ve kestirilen değerleri arasındaki RMSE'lerin  $\beta_1$  ve  $\beta_4$ 'ten daha düşük olduğunu gözlemlemiştir. Yetenek parametreleri için her iki örneklem büyüklüğünde de bias ve RMSE en düşük tekdüze dağılımda bulunurken genel olarak en yüksek çarpık dağılımda bulunmuştur. Genel itibariyle a parametresi için en doğru kestirimler çarpık dağılımda en zayıf kestirimler tekdüze dağılımda yapılırken, b ve  $\theta$  parametreleri için en doğru kestirimler tekdüze dağılımda en zayıf kestirimler çarpık

dağılımda yapılmıştır ve dağılımların arasındaki farklılıklar küçük olup dağılım türünün parametre iyileştirmeye çok büyük bir etkisi olmadığı görülmüştür.

Nartgün (2002) araştırmasında farklı örneklemelerden elde edilen veriler üzerinde KTK'ye ve DTM'ye dayalı olarak kestirilen madde parametrelerinin ne derece değişmezlik özelliği gösterdiğini incelemiştir. Bu amaç doğrultusunda likert ve metrik ölçekler ile 1299 kişiden veri toplamış ve değişik özellikler dikkate alınarak beş farklı alt örneklem oluşturmuştur, daha sonra tüm gruptan kestirilen madde parametreleri ile alt örneklemelerden kestirilen madde parametreleri arasındaki ilişkiyi test ederek, ölçeğin likert formunda tüm gruptan ve alt gruplardan MTK'ye göre kestirilen a parametresi değerleri arasındaki korelasyonları 0,98; 0,97; 0,98; 0,86 ve 0,81 bulmuştur. Elde edilen bu korelasyon katsayılarına dayanarak farklı örneklemelerden kestirilen madde parametrelerinin benzer olduğu sonucuna varılmıştır.

De Mars (2003) yaptığı araştırmada STM'de kategori sayısının, test uzunluğunun, örneklem büyüklüğünün, dağılım türünün ve ayırt edicilik katsayısının parametre iyileştirmedeki etkisini incelemiştir. Bunun için dört farklı test oluşturmuştur ve bunlar üç kategorili 20 madde ile üç kategorili 40 madde olmak üzere madde başı 6 parametre içeren testler ve altı kategorili 10 madde ile altı kategorili 20 madde olmak üzere madde başı 12 parametre içeren testlerdir. Her test için üç dağılım türünden ( $N(0,1)$ , çarpık( $\alpha = 1,25$   $\beta = 10$ ), tekdüze(-1,7 ile +1,7 aralığında) ), iki ayırt edicilik düzeyinden (düşük ve yüksek) ve iki örneklem büyüklüğünden (600 ve 2400) oluşan koşullarda 100 replikasyonla veri üretmiştir. Madde parametresi kestirimini maksimum iterasyon ve cycle sayısını 200'e çıkararak MULTLOG 6.0 ile yapan De Mars (2003) parametre iyileştirmenin doğruluğunu RMSE, bias ve korelasyon hesaplayarak yorumlamıştır. Ayrıca belirlenen koşulların etkisini incelemek amacıyla ANOVA yapmış ve  $\omega^2$  değerini hesaplamıştır. Araştırmada a parametresi için tüm örneklem büyüklüğü, madde parametresi sayısı ve ayırt edicilik düzeylerinde RMSE'nin düşük değerleri normal ve tekdüze dağılımda yüksek değerleri ise çarpık dağılımda aldığı bulunmuştur. Bunun aksine en düşük bias değerlerinin de çarpık dağılımda olduğu, daha yüksek ve aralarında büyük bir farklılık olmayan değerlerin ise normal ve tekdüze dağılımda olduğu ayrıca 12 parametrelilikte normal dağılımın RMSE'si tekdüze dağılımından daha düşükken 6 parametrelilikte normal ve tek düze dağılımın RMSE'leri arasında büyük bir farklılık olmadığı görülmüştür. MTK c parametrelerinde ise 6 parametrelilikte koşullar için tüm örneklem büyüklüğü, ayırt

edicilik düzeylerinde bias'ın her dağılımda aynı olduğu, 12 parametrelili koşullar için de en düşük bias'ın çarpık dağılımda olduğu, yüksek ve aralarında büyük bir farklılık olmayan bias'ında normal ve tekdüze dağılımda olduğu bulunmuştur. Modeldeki c parametresi için RMSE değerlerine gelindiğinde hem 6 parametrelili hem de 12 parametrelili koşullarda en yüksek değerleri çarpık dağılım alırken 6 parametrelili koşullarda her durumda normal ve tek düze dağılım arasındaki farklılık çok az, 12 parametrelili koşulda ise her durumda normal dağılımın RMSE'si tek düze dağılımından düşüktür. Son olarak hem a parametresi hem c parametresi için her koşulun RMSE değerlerini incelediğinde dağılım türlerinin  $p < 0,0001$  düzeyinde manidar bir farklılık yarattığı ve  $\omega^2$  değerinin a parametresinde 0,07, c parametresinde 0,04 olduğunu bulmuştur. Tüm faktörlere bakıldığında en büyük  $\omega^2$  değerinin madde başına parametre sayısında (6 parametrelili, 12 parametrelili) olduğu, bu faktörü örneklem büyüklüğünün takip ettiği görülmüştür. Araştırmacı hata varyansının çarpık dağılımda daha büyük olduğunu ifade etmiştir.

Doğan ve Tezbaşaran (2003) yaptıkları araştırmada klasik test kuramını ve madde tepki kuramını örneklem bağlamında gerçek veriler üzerinden karşılaştırmayı amaçlamıştır. Bu doğrultuda 4 seçenekli 100 sorudan oluşan ÖSYS 2001'e giren bireylerin oluşturduğu evrenden randum, sağa çarpık, sola çarpık, basık ve normal dağılım gösteren beş örneklem seçmiştir ve örneklem büyüklükleri sırasıyla 27501, 23531, 23531, 24691, 29244'tür. Örneklemelerin çarpıklık ve basıklık katsayıları ise sırasıyla 0,20 ve -0,73 (evren ile aynı çarpıklık ve basıklık katsayısına sahip randum dağılım); 1,00 ve 1,37; -1,00 ve 1,37; 0,00 ve -1,07; 0,00 ve 0,00'dır. Oluşturulan her bir örneklem için varsayımların kontrolü yapılmış olup örneklemelerin tek boyutluluk ve yerel bağımsızlık varsayımlarını sağlarken normallik varsayımını yeterince sağlamadığı görülmüştür. KTK ve MTK'de örneklemden kestirilen parametreleri evrenden kestirilen parametreler ile karşılaştırmak için Spearman'ın sıra farkları korelasyonunu hesaplamış ve korelasyon farklarını Fisher z'si ile test etmiştir ve MTK'ye dayalı b parametresinde tüm modellerde (1PLM, 2PLM ve 3PLM), a parametresinde de tüm modellerde (2PLM, 3PLM) örneklem ve evren arasındaki korelasyonu manidar bulmuştur. MTK b parametresi için 1PLM'de diğer dağılımlara göre düşük korelasyona sahip olanlar sağa çarpık(0,950) < sola çarpık(0,979) şeklinde sıralanırken 2PLM'de sola çarpık(0,877) < sağa çarpık(0,944) şeklinde, 3 PLM için ise basık(0,949) < sola çarpık(0,950) < sağa çarpık(0,953) şeklinde sıralanmıştır. MTK b parametresi için en yüksek korelasyon ise tüm modellerde randum dağılımdadır. Modellerdeki a parametresi incelendiğinde ise hem 2PLM hem de 3PLM için en düşük

korelasyon sağı çarpık dağılımda (sırasıyla: 0,731; 0,501), en yüksek korelasyon ise randum dağılımda (sırasıyla: 0,988; 0,989) gözlenmiştir. Sonuç olarak korelasyon değerleri incelendiğinde a parametrelerinin dağılımın çarpıklığından daha fazla etkilendiği bulunmuştur.

Özdemir (2004) araştırmasında çoktan seçmeli test maddelerinin iki kategorili (1, 0) ve ağırlıklı puanlanmasının (1, 2, 3, 4) testin geçerliğini ve güvenilirliğini nasıl etkilediğini KTK ve MTK bağlamında incelemeyi amaçlamıştır ve bu amaç doğrultusunda ilköğretim düzeyindeki (4., 5., 6. ve 7. sınıflar) 1608 bireye geliştirdiği 4 seçenekli 20 maddeden oluşan çoktan seçmeli Türkçe testini uygulamıştır. MTK DTM'ye dayalı yapılan analizlerde güvenilirlik katsayısı olarak ikili puanlanan maddeler için Lord'un güvenilirlik katsayısını, ağırlıklı puanlanan maddeler için ise marjinal güvenilirlik katsayısını hesaplamıştır;  $\theta$  ölçümlerinin geçerliğini incelemek amacıyla ise hem ikili puanlanan hem de ağırlıklı puanlanan test için 4, 6 ve 7. sınıflarındaki bireylerin yetenek puanları ölçüt kabul edilerek bu bireylerin ortalamaları ile 5. sınıftaki bireylerin yetenek puanları ortalamaları arasında manidar bir farklılık olup olmadığı t-testi ile test edilmiştir. Güvenirlik katsayısının ikili puanlanan testte 0,78, ağırlıklı puanlanan testte 0,68 olduğu görülmüştür. Geçerliği incelemek için yapılan t-testinde ise hem ikili hem de ağırlıklı puanlanan testler için 4.-5. sınıflar ve 5.-7. sınıflar arasında manidar bir farklılık olduğu, 5-6. sınıflar arasında ise manidar bir farklılık olmadığı gözlenmiştir. Araştırmanın sonucunda MTK bağlamında güvenilirlik açısından maddelerin ikili puanlanmasının ağırlıklı puanlanmasından daha iyi olduğu; geçerlik açısından ise iki puanlama yönteminin de 6. sınıf hariç tüm grupları birbirinden ayırdığı için geçerli olduğu ifade edilmiştir.

Lautenschlager, Meade ve Kim (2006) yaptıkları simülasyon çalışmasında test uzunluğu, örneklem büyüklüğü ve dağılım türünün DTM'de madde ve birey parametresi kestiriminin doğruluğuna etkisini incelemişlerdir. Simülasyonun gerçeğe yakın olması amacıyla 20 maddelik ve 5'li likert tipi Minnesota Satisfaction Questionnaire (MSQ) kısa formu ile 891 sanayi çalışanından elde edilen verileri kullanmışlardır. Bu verileri kullanarak GENIRV programıyla 100 replikasyon yapılarak oluşturulan 84 koşul; 7 örneklem büyüklüğünü (75, 150, 200, 300, 500, 1000 ve 2000), 4 test uzunluğunu (5, 10, 15, 20) ve 3 dağılım türünü (normal, çarpık(+1,25), tekdüze) içermektedir. Sonuç olarak Marginal Maximum Likelihood (MML) ile madde parametreleri kestirimi için mean squared error (MSE) değerlerinin çarpık dağılımda çok yüksek olduğunu, tekdüze dağılımda ise en düşük

olduğunu, örneklem büyüklüğü arttıkça genellikle düştüğünü, test uzunluğu arttıkça ise çok az miktarda azaldığını görmüşlerdir. Maximum a posteriori (MAP) ile birey parametreleri kestirimi için MSE değerlerine bakıldığında ise normal dağılımda en düşük olduğunu, tekdüze dağılımda en yüksek olduğunu, örneklem büyüklüğünün MSE'yi neredeyse hiç etkilemediğini bulmuşlardır.

Dolma (2009) yaptığı simülasyon çalışmasında MTK çerçevesinde geliştirilen çok kategorili Rasch modeli (Andrich'in Dereceleme Ölçeği Modeli ve Masters'ın KPM'si) ile DTM'yi parametre kestirme doğrulukları açısından karşılaştırma temel amacıyla WinGen2'de a parametresi 0,8 ile 2,00 aralığında bulunan veriler üretilmiştir. Üretilen standart normal dağılıma sahip veriden farklı iki örneklem büyüklüğünde (150 ve 300) rastgele veri setleri seçilmiştir, ölçek uzunluğunda da (6, 9 ve 12) en kapsamlı veri üretilerek rastgele örnekleme yapılmıştır. Kategori sayısı (4 ve 6 kategorili) için ise 6 kategoriye göre veri üretilmiş 4 kategorili maddeler içeren veri setleri yeniden kodlama ile oluşturulmuştur. Araştırmada bulunan bu 12 ( $2 \times 3 \times 2$ ) koşul dört kez tekrarlanarak 48 analiz MULTILOG 7 ve Winsteps programlarında yapılmıştır. Dolma (2009) araştırmanın problemleri doğrultusunda her koşul için kestirim doğruluğunu incelemiştir. Araştırmanın sonucunda çok kategorili Rasch modeli ile DTM arasında kestirim doğruluğu yönünden önemli bir fark olmadığını ve örneklem büyüklüğü, test uzunluğu ve kategori sayısı arttığında kestirim doğruluğunun arttığını bulmuştur.

Bahry (2012) araştırmasında DTM ve MULTILOG 7 kullanarak parametre kestiriminde örneklem büyüklüğünün ve çarpıklığın etkisini incelemiştir ve bu doğrultuda, çarpıklık katsayısı 0,00; 0,50 ve 1,00 olan üç tip dağılımda, 100, 250, 500, 750, 1000, 1500 ve 3000 olmak üzere yedi çeşit örneklem büyüklüğünde olmak üzere 21 koşulda simülasyon verisi üretmiştir. Tüm koşullarda 20 maddeli ve 5 kategorili olan veriler WinGen 3 programıyla elde etmiştir ve kestirimlerin doğruluğunu RMSE ve bias değerlerini inceleyerek yorumlamıştır. Araştırmanın sonucunda tüm örneklem büyüklükleri için en zayıf a ve b parametresi kestirimlerinin çarpıklık katsayısı 1,00 olan dağılımda yapıldığını ve genel olarak örneklem büyüklüğü arttıkça parametre kestirimi doğruluğunun arttığı sonucuna ulaşmıştır. Ayrıca EM cycle yakınsaklık kriterini sağlamama yüzdesini de inceleyen araştırmacı 3000 kişilik örneklem için  $\text{ÇK}=0,00$  ve  $\text{ÇK}=0,50$  olan dağılımlarda kriteri sağlamayan replikasyon bulunmazken  $\text{ÇK}=1,00$  olan dağılımda kriteri sağlamayan replikasyon yüzdesini 11,40 olarak bulmuştur. Bahry (2012) araştırmasının sonucunda

parametre kestiriminin doğruluğu için normal dağılımlarda minimum örneklem büyüklüğünün 750 olması gerektiğini de ifade etmiştir.

Kieftenbeld ve Natesan (2012) yaptıkları simülasyon çalışmasında farklı dağılımlar, test uzunlukları ve örneklem büyüklükleri altında DTM'ye dayalı olarak "Markov chain Monte Carlo (MCMC)" ve "Marginal Maximum Likelihood (MML)" yöntemlerinin parametre iyileştirme etkililiklerini karşılaştırmayı amaçlamışlardır. Bu amaç doğrultusunda Lautenschlager vd. (2006)'nın 20 maddelik ve 5'li likert tipi MSQ-kısa formu ile elde ettikleri verilerin parametre kestirimlerine dayanarak dağılım türü (normal( $\bar{X} = 0, S_x = 1$ ), çarpık( $\text{ÇK} = 1,25; BK = 1,50$ ), tekdüze(-3;+3) ), 4 farklı test uzunluğu (5, 10, 15, 20 madde) ve 5 farklı örneklem büyüklüğünden (75, 150, 300, 500, 1000 kişi) oluşan 60 koşul için 100 replikasyon yaparak 600 veri seti üretmişlerdir ve her veri seti için 2 parametre kestirim yönteminde de (MML, MCMC) gerçek ve kestirilen parametreler arasındaki korelasyonları, RMSE'leri ve farklı koşul özelliklerinin (örneklem büyüklüğü, dağılım türü, test uzunluğu×dağılım türü vb.) RMSE ve korelasyonu açıklama varyanslarını hesaplamışlardır. Araştırmacılar çarpık dağılımı Fleishman'in (1978) güç yöntemini kullanarak normal dağılımdan türetmişlerdir ve MML parametre kestirimi için MULTILOG 7, MCMC parametre kestirimi için OpenBUGS kullanmışlardır. Araştırmanın sonucunda, madde parametrelerinin kestiriminde varyansın büyük bir kısmını örneklem büyüklüğünün, birey parametrelerinin kestirilmesinde ise test uzunluğunun açıkladığını, 300 ve daha fazla bireyin bulunduğu örneklemelerde MML ile MCMC arasındaki farkın çok az olduğunu bulmuşlardır. Ayrıca dağılım türünün  $a(2,55)$ ,  $b_1(4,96)$ ,  $b_2(2,19)$ ,  $b_3(0,28)$ ,  $b_4(4,29)$  ve  $\theta(2,62)$  parametrelerinin kestiriminde RMSE'yi açıklama varyanslarının düşük olduğunu, hem birey hem madde parametrelerinin kestiriminde tekdüze dağılımın korelasyonları açıklama varyansının normal ve çarpık dağılıma göre daha yüksek olduğunu, özellikle  $\theta(22,22)$  kestiriminde tekdüze dağılımın korelasyonu açıklama varyansının yüksek olduğunu bulmuşlardır ve her iki kestirim yönteminde de tekdüze dağılımın normal ve çarpık dağılıma göre daha iyi olduğu sonucuna varmışlardır.

Bilican Demir (2014) yaptığı çalışmada çok kategorili puanlanan maddelerde madde işlev farklılığını (MIF) belirlemede "Mantel Testi"ni ve "MTK-Olabilirlik Oran Testi (MTK-OOT)"ni farklı koşullarda I. tip hata ve istatistiksel güç oranları açısından karşılaştırmayı amaçlamıştır. Bu koşullar I. tip hata için üç örneklem büyüklüğü (500, 1250 ve 2000), üç dağılım türü (Referans grup  $\sim N(0,1)$ , Odak grup  $\sim N(0,1)$ ;  $R \sim N(0,1)$ , O



$\sim N(-0,5; 1)$ ;  $R \sim N(0,1)$ ,  $O \sim N(-1,1)$ ), iki MİF belirleme testi (Mantel, MTK-OOT) olmak üzere 18 tanedir; istatistiksel güç oranı için ise üç örneklem büyüklüğü (500, 1250 ve 2000), üç dağılım türü (Referans grup  $\sim N(0,1)$ , Odak grup  $\sim N(0,1)$ ;  $R \sim N(0,1)$ ,  $O \sim N(-0,5; 1)$ ;  $R \sim N(0,1)$ ,  $O \sim N(-1,1)$ ), iki MİF miktarı (orta = 0,43; büyük = 0,64), iki MİF örüntüsü (düşük, yüksek), iki MİF belirleme testi (Mantel, MTK-OOT) olmak üzere 72 tanedir ve bu koşullar WinGen3 simülasyon programıyla KPM'ye dayalı olarak oluşturulmuştur. Ayrıca her koşul için 100 replikasyon yapılmıştır. Araştırmacı I. tip hata oranlarını 20 madde için, istatistiksel güç oranlarını ise MİF içeren üç madde için hesaplamıştır. Verilerin analizi MTK-OOT için MULTILOG programıyla, Mantel Test için DIFAS programıyla yapılmıştır. I. tip hata oranları tüm örneklem büyüklüklerinde MTK-OOT'de en düşük  $O \sim N(0,1)$  olan dağılımda, en yüksek ise  $O \sim N(-1,1)$  olan dağılımdadır, Mantel Testinde de en yüksek  $O \sim N(-1,1)$  olan dağılımdayken en düşük genel olarak  $O \sim N(-0,5; 1)$  olan dağılımdadır. İstatistiksel güç oranları düşük MİF örüntüsü için tüm örneklem büyüklüklerinde, MİF belirleme testlerinde ve tüm MİF miktarlarında incelendiğinde en yüksek genel olarak  $O \sim N(-1,1)$  olan dağılımda, en düşük ise genel olarak  $O \sim N(0,1)$  olan dağılımda olduğu, yüksek MİF örüntüsü için ise düşük örüntünün tam tersine en yüksek genel olarak  $O \sim N(0,1)$  olan dağılımda, en düşük  $O \sim N(-1,1)$  olan dağılımda olduğu görülmüştür. Sonuç olarak I. Tip Hata değerleri her iki MİF belirleme testi için de odak grup yetenek dağılım ortalamasının, referans grup yetenek dağılımına göre sapma gösterdiği koşullarda, "referans grup  $\sim N(0,1)$ , odak grup  $\sim N(0,1)$ " koşuluna göre daha yüksektir. İstatistiksel güç değerleri incelendiğinde ise düşük MİF örüntüsünde orta düzey MİF miktarı için birim normal dağılımdan saptıkça arttığı, büyük MİF miktarında değerlerin birbirine yakın olduğu görülmüştür. Ayrıca birim normal dağılımdan sapma gösteren odak gruba sahip koşullarda yüksek MİF örüntüsünde istatistiksel güç değerlerinin çok düşük olduğu ifade edilmiştir.

Preston ve Reise (2014) yaptıkları araştırmada dağılım türü normal, çarpık ve iki modlu olduğunda sıralı çok kategoriye sahip maddelerin STM'ye dayalı olarak yapılan madde parametre kestirimlerinin daha doğru olmasını sağlamak için alternatif bir kestirim yaklaşımı olan Ramsay Curve MTK'yi (RE-MTK) kullanmışlardır. Araştırma da STM'ye dayalı bir ön çalışma yapmış ve bu çalışma için üç dağılım türünde (normal, çarpık(ÇK = 1,75, BK = 6,75), iki modlu) ve CBD parametrelerini iki ayırt edicilik düzeyinde (yüksek (1,5), orta (0,75)) olmak üzere 6 koşulda veri üretmişlerdir. Verileri 4 kategorili 10 maddeli olarak üreterek ve kesme (intersection) parametrelerini simetrik olarak -1,0; 0; 1,0

olarak tüm koşullarda sabitleyerek 1000 replikasyon yapmışlardır. Kestirimleri MML yöntemiyle yapan araştırmacılar RMSE ve bias hesaplamışlardır. Ön çalışmanın sonucunda her iki ayırt edicilik düzeyinde de en düşük RMSE ve bias'ın genellikle normal dağılımda olduğu, en yüksek RMSE ve bias'ın ise çarpık dağılımda olduğunu gözlemlemişlerdir. Ön çalışma sonrasında iki örneklem büyüklüğünde (500 ve 2000), üç dağılım türünde (normal, çarpık ( $\text{ÇK} = 1,75$ ,  $\text{BK} = 6,75$ ), iki modlu) ve başka değişken faktörlerde olmak üzere toplam 144 koşulda STM'ye dayalı olarak veri üretmişlerdir. Çarpık dağılıma sahip veri setlerini Fleishman'ın güç yöntemini kullanarak üreten araştırmacılar 10 madde, 4 kategori özelliklerini sabit tutmuş ve 1000 replikasyon yapmışlardır. Bu veri setlerinin kestirimlerini RE-MTK uygulayan EQSIRT programında STM'ye dayalı olarak kestirmişlerdir ve bu programda maksimum EM cycle sayısını 999'a ayarlamışlardır. Bias ve TCC farklılık indeksi hesaplayan araştırmacılar tüm koşullarda en zayıf kestirimin genellikle çarpık dağılımda, en iyi kestirimlerin ise normal dağılımda olduğunu bulmuşlardır. İki modlu dağılımın bias ve TCC değerlerinin genel olarak çarpık dağılımla benzer olduğunu ancak az bir farkla iki modlu dağılımda daha düşük olduğunu gözlemlemişlerdir. Sonuçlara dayalı olarak normal olmayan dağılımlarla karşılaşan araştırmacıların RE-MTK kullanabileceklerini ifade etmişlerdir.

Sen (2014) yaptığı araştırmada normallik sayılıtları ihlalinin karma MTK modellerinden (mixture-IRT) karma Rasch modelde, karma 2PLM ve karma 3PLM'de parametreleri iyileştirmek için iki simülasyon çalışmasına yer vermiştir. Ayrıca normal olmayan dağılıma sahip geniş ölçekli değerlendirmelerde sonuçların nasıl yorumlanacağını göstermek amacıyla deneysel bir çalışma yapmıştır. Her iki monte carlo simülasyon çalışmasında da iki test uzunluğunda (28 ve 40), iki örneklem büyüklüğünde (600 ve 2000) 50 replikasyon yaparak veri üretmiştir ve MCMC'ye dayalı parametre kestirimlerini WinBUGS 1.4'te yapmıştır. Ayrıca verilerin simülasyonu için Florida Comprehensive Assessment Test (FCAT) 9. sınıf matematik testinin madde parametrelerini kullanmıştır. Birinci simülasyon çalışmasında dağılım türünün parametre iyileştirmeye etkisini incelemeyi amaçlayarak test uzunluğu ve örneklem büyüklüğü faktörlerinin yanı sıra beş dağılım türünde [normal ( $N(0,1)$ ), çarpık ( $\text{ÇK} = 0,75$ ,  $\text{BK} = 0,00$ ), basık ( $\text{ÇK} = 0,00$ ,  $\text{BK} = -0,75$ ), tekdüze (-2 ile +2 aralığında) ve iki modlu simetrik ( $N(-1,5;1,0)$ ,  $N(1,5;1,0)$ )] veri üreterek farklı koşullar oluşturmuştur. İkinci simülasyon çalışmasında ise çarpıklık ve basıklık katsayılarının etkisini incelemeyi amaçlamıştır ve bu doğrultuda altı dağılım türünde [1. koşul ( $\text{ÇK} = 1,5$ ,  $\text{BK} = 4,0$ ), 2. koşul ( $\text{ÇK} = 1,5$ ,  $\text{BK} = 3,5$ ), 3. koşul ( $\text{ÇK} = 1,5$ ,

BK = 3,0), 4. koşul (ÇK = 1,0, BK = 3,5), 5. koşul (ÇK = 0,5, BK = 3,5), 6. koşul (ÇK = 0,0, BK = 3,5)], iki test uzunluğunda, iki örneklem büyüklüğünde, üç modelde ve başka faktörlerde toplam 432 ( $6 \times 2 \times 2 \times 3 \times 3 \times 2$ ) koşulu kapsayan veri üretmiştir. Birinci simülasyon çalışmasında oluşturulan basık ve çarpık dağılımlar ile ikinci simülasyon çalışmasında oluşturulan altı koşul Fleishman'ın güç yöntemi ile oluşturulmuştur. İki simülasyon çalışmasının yanı sıra yapılan deneysel çalışma da ise TIMSS 2011 8. sınıf matematik testi kullanılmış ve negatif çarpık dağılıma sahip olan Güney Kore çalışma grubunu oluşturmuştur. Bu grubun çarpıklık ve basıklık katsayılarını RCLOG v.2 ile hesaplamış, verileri üç karma MTK modeline de dayalı olarak analiz ederek çarpıklığın parametre kestirimine etkisini incelemiştir. Parametre iyileştirmeyi RMSE ve bias ile inceleyen araştırmacı test uzunluğu, örneklem büyüklüğü, dağılım türleri ve modelin etkisini incelemek amacıyla dört yönlü ANOVA yapmıştır. Birinci simülasyon çalışmasının sonuçlarında b parametresi için RMSE'nin tüm koşullarda en yüksek iki modlu simetrik dağılımda iken diğer dağılımlar arasında farklılığa sahip olmadığını, bias değerlerinin karma Rasch modelde en yüksek iki modlu simetrik dağılımda olduğunu, en düşük basık dağılımda iken çarpık ve normal dağılımda benzer olduğunu, karma 2PLM'de en yüksek iki modlu simetrik dağılımda iken diğer dağılımlar arasında farklılık olmadığını ve karma 3PLM'de de en yüksek iki modlu simetrik dağılımda olduğunu bulmuştur. Ayrıca b parametresi için ANOVA sonuçlarında dağılım türünün hem RMSE'yi hem de bias'ı 0,001 düzeyinde manidar bir şekilde etkilediğini ve etki büyüklüğünün sırasıyla  $\eta^2 = 0,125$  orta etki,  $\eta^2 = 0,378$  büyük etki olduğunu gözlemlemiştir. MTK a parametresi için RMSE hem karma 2PLM hem de karma 3PLM için genel olarak en büyük iki modlu simetrik dağılımda daha sonra tekdüze dağılımdayken normal ve basık dağılımın RMSE değerleri birbirine benzer olduğunu, bias değerlerinde ise normal, çarpık ve basık dağılıma ait değerlerin 0,025'ten küçük olduğunu, en yüksek değerleri iki modlu simetrik dağılımda daha sonra ise genel olarak tekdüze dağılımda gösterdiğini bulmuştur. Bunun yanı sıra ANOVA sonuçlarında hem RMSE hem bias için dağılım türleri arasındaki farklılığı manidar bulmuştur, etki büyüklüklerini sırasıyla  $\eta^2 = 0,554$  (büyük etki),  $\eta^2 = 0,062$  (orta etki) olarak gözlemlemiştir. İkinci simülasyon çalışmasının sonuçlarında ise b parametresi için RMSE değerlerinin karma Rasch modelinde ve karma 2PLM'de çoğu koşulda basıklık azaldıkça azaldığını, çarpıklık katsayısından çok etkilenmediğini; karma 3PLM'de ise diğer modellerin tersine basıklık katsayısından çok etkilenmediğini, çarpıklık katsayısı azaldıkça azaldığını gözlemlemiştir. MTK b parametresi için bias değerlerinin karma Rasch

modelinde RMSE sonuçlarıyla paralel olduğunu, ancak karma 2PLM'de hem basıklıktan hem çarpıklıktan etkilenmediğini ve her koşulda en düşük bias'ın 6. koşulda olduğunu bulmuştur. ANOVA sonuçlarında dağılım türünün RMSE ve bias üzerinde manidar bir etkisi bulunduğunu ve etki büyüklüklerinin sırasıyla  $\eta^2 = 0,022$  ve  $\eta^2 = 0,082$  olduğunu gözlemlemiştir. MTK a parametresi için RMSE değerlerinin karma 2PLM'de hem çarpıklıktan hem basıklıktan etkilenmediğini, karma 3PLM'de çarpıklık ve basıklık arttıkça arttığını, bias değerleri için ise karma 2PLM'de en yüksek değer en yüksek basıklığa sahip 1. koşulda (BK = 4,00) olduğunu, 3PLM'de en yüksek değer 3. koşulda (ÇK = 1,5, BK = 2,5) olduğunu bulmuştur. Birinci simülasyon çalışması için en zayıf kestirimler iki modlu simetrik dağılımda olduğu, ikinci simülasyon çalışması için ise çarpıklık ve basıklık katsayılarının artması genel olarak daha zayıf kestirimlerin yapılmasına neden olduğu sonucuna ulaşmıştır.

Alan yazın incelendiğinde DTM ya da diğer modeller için genellikle örneklem büyüklüğünün, test uzunluğunun, kestirim yönteminin, paket programların ve normalliğin ölçme kesinliğine etkisi incelenmiştir. DTM'de farklı dağılım türlerinin ölçme kesinliğini nasıl etkilediğini inceleyen araştırmalara bakıldığında dağılım türleri; Reise ve Yu (1990), Lautenschlager, Meade ve Kim (2006) ve Kieftenbeld ve Natesan (2012)'in çalışmasında normal, ÇK = 1,25 ve tekdüze, Ankenmann ve Stone (1992)'un çalışmasında normal ve ÇK = 0,75 olan dağılım, Bahry (2012)'nin çalışmasında ise normal, ÇK = 0,5 ve ÇK = 1,00 olan dağılım şeklindedir. Bu çalışmada ise diğer çalışmaların kullandığı dağılım türlerinden farklı bir şekilde normallik durumları olarak (ÇK = -1 ile +1 aralığında dağılım normal olduğu için) çarpıklık katsayısı 0,00; 0,50 ve 1,00 olan dağılımlara karşılık, normalliğin ihlal edildiği durumlar olarak çarpıklık katsayısı 1,50 ve 2,00 olan dağılım türlerinde ölçme kesinliğinin nasıl etkilendiği araştırılmıştır. Ayrıca diğer araştırmalara bakıldığında araştırmacıların normalliğin ihlalinin marjinal güvenirlik katsayısını, MULTILOG programıyla yapılan kestirimlerde programın verdiği hata sayısını nasıl etkilediğini incelemedikleri görülmüştür ve bunlar da bu araştırmanın problemi olarak incelenmiştir.

## BÖLÜM III

### YÖNTEM

Bu bölümde araştırmanın modeli, çalışma grubu, veri toplama aracı tanıtılarak verilerin üretilmesi, madde tepki kuramının varsayımlarının kontrolü ile verilerin çözümlenmesi ve yorumlanması aşamalarından söz edilmiştir.

#### **Araştırmanın Modeli**

Bu çalışmada Samejima'nın (1969) derecelendirilmiş tepki modelinde normallik ihlali edildiği durumlarda ölçme kesinliğinin nasıl etkilendiği incelenmiştir. Normallik ihlali ve normallik durumlarını karşılaştırabilmek amacıyla veriler veri simülasyon programıyla üretildiği için araştırma bir simülasyon çalışmasıdır. Karasar (2013)'e göre temel araştırmaların amacı var olan bilgiye yenilerini katmaktır. Bu çalışmada normallik ihlali durumuna diğer çalışmalardan farklı olarak çarpıklık katsayısı 1,50 ve 2,00 olan dağılımları da katıldığı için araştırma temel araştırma niteliğindedir.

#### **Çalışma Grubu**

Bu araştırmanın çalışma grubu Mergen, Bernstein, Tavlı, Öngel, Tavlı ve Tan (2011) tarafından Beck Depresyon Envanterinin (BDE) uygulandığı Uludağ Üniversitesi Aile Sağlık Merkezi'ne başvuran ve genellikle öğrencilerden oluşan 626 hastadan oluşmaktadır. Bu hastalar üniversite polikliniğine art arda başvuran bireylerdir ve araştırmacıların kullandığı örnekleme yöntemi uygun örneklemedir. Bu bireylerden elde edilen veriler MULTİLOG 7.03'te (Thissen, Chen ve Bock, 2003) Samejima'nın (1969) DTM'sine dayanarak analiz edilmiştir. Mergen vd. (2011)'nin izniyle kestirilen madde parametreleri kullanılarak WinGen 3.1 (Han, 2007) simülasyon programıyla 5 tür dağılım için 25 replikasyon ile Samejima'nın (1969) DTM'sine dayalı olarak 3000 kişilik veri setleri

(25×5=125 veri seti) oluşturulmuştur. Bu veri setlerinde çarpıklık katsayıları sırasıyla 0,00; 0,50; 1,00; 1,50; 2,00 (sağa çarpık, pozitif kayışlı) olan dağılımlardır.

## **Veri Toplama Aracı**

### **Beck Depresyon Envanteri**

Beck Depresyon Envanteri (BDE) Beck, Ward, Mendelson, Mock ve Erbaugh (1961) tarafından geliştirilmiştir. Beck Depresyon Envanterinin (BDE) geliştirilme aşamasında ilk olarak Beck psikanalitik psikoterapi sürecindeki depresif hastaları sistematik olarak gözlemleyerek onların karakteristik davranışlarını, semptomlarını kayıt altına almıştır. Beck vd. (1961) Beck'in seçtiği bu depresif hastalarda ortaya çıkan davranış ve semptomların psikiyatri literatüründe yer alan depresyon tanımlarıyla uyumlu olduğunu ifade etmiştir. Bu sürece dayanarak gözlenen davranışlardan ve semptomlardan oluşan 21 maddeyi elde etmiş ve bu maddelerden envanteri oluşturmuştur. Klinik olarak oluşturulan bu envantere her bir madde depresyonun özel bir belirtisini betimlemektedir ve 4 ya da 5 öz değerlendirme durumunun derecelendirilmiş dizilerinden oluşmaktadır. Beck vd. (1961) maddelerdeki derecelendirilmiş ifadeleri semptomların nötr şiddetinden maksimum şiddetine doğru sıralamıştır. BDE'de 0-3 arasında puanlanmaktadır ve her bir ifadeye şiddet derecesini belirtmek amacıyla bu değerler atanmıştır. Beş ifade bulunan maddelerde iki ifadeye aynı puan verilmektedir, örneğin C maddesinde 2 ile puanlanan iki ifade vardır ve bunlar 2a, 2b olarak adlandırılmıştır. Beck vd. (1961) envanteri oluşturduktan sonra Pennsylvania Üniversitesi Hastanesi ve Philadelphia General Hastanesi psikiyatri bölümlerinde ve yataklı servislerindeki hastalardan elde ettiği ölçümlerin güvenilirliklerini incelemişlerdir. Yaşları 15-44 arasında değişen hastalardan 226'sı orijinal grubu, 183'ü ise replikasyon grubunu oluşturmuştur. Bu iki grubun ölçümleri ile geçerlik incelemesi yapan geliştiriciler yaptıkları tek yönlü Kruskal-Wallis nonparametrik testi sonuçlarını 0,001 düzeyinde manidar bulmuştur. İki yarıya bölme yöntemiyle iç tutarlık güvenilirlik katsayısını hesaplayan Beck vd. (1961) iki yarı arasındaki korelasyonu Pearson ve Spearman-Brown yöntemleri ile hesaplamış ve bu ölçümler için sırasıyla 0,86 ve 0,93 bulmuşlardır. Ayrıca iç tutarlığı incelemek için yaptıkları Kruskal-Wallis non-parametrik testinin sonuçlarını 0,01 düzeyinde manidar bulmuşlardır (Beck vd., 1961). BDE 1969-1972 yılları boyunca revize edilerek bu versiyonun telif hakkı 1978'de alınmıştır (Zengin, 1999). Beck Depresyon Envanterinin revize versiyonu Beck, Rush, Shaw ve Emery (1979,

s. 396) tarafından yayınlanmıştır. Bu envanterin son versiyonu olan BDE-II ise Beck, Steer ve Brown (1996) tarafından, DSM-IV majör depresif atak tanı ölçütlerine daha sıkı bağlı olması amacıyla geliştirilmiştir. Bahsedilen üç versiyon dışında, alan yazında BDE'nin kısa formu (BDI-SF, BDE-KF), birinci basamak için BDE (BDI-PC) gibi versiyonları da bulunmaktadır. Bunların yanı sıra Beck ve Steer tarafından 1993'te revize edilmiş versiyon olan BDI-IA' ye de alan yazında rastlanmaktadır (Beck, Steer, Ball ve Ranieri, 1996).

Türkiye'de BDE'nin Türkçeye uyarlanması için iki önemli çalışma yapılmıştır. Bu uyarlamaların birincisi Teğin (1980) tarafından yapılmıştır. Teğin (1980) yaptığı uyarlama çalışmasında, 30 hastadan ve 40 öğrenciden elde ettiği ölçümlerin güvenirlik katsayılarını yarıya bölme yöntemiyle sırasıyla 0,61 ve 0,78 olarak hesaplamış, 40 öğrenciden elde edilen ölçümlerin güvenirlik katsayısını ise test tekrar test yöntemiyle 0,65 olarak hesaplamıştır (Hisli, 1989). BDE'nin ikinci uyarlama çalışması ise Hisli tarafından yapılmıştır. Hisli 1984 yılında Türkçeyi ve İngilizceyi iyi bilen 58 kız öğrenciden 28'ine önce Türkçe bir hafta sonra da İngilizce formu uygulamış, 30'una ise önce İngilizce bir hafta sonra Türkçe formu uygulayarak bir uyarlama çalışması yapmış iki form arasındaki korelasyonu 0,81 ve 0,73 bulmuştur, 1988 yılında ise Türkçeye uyarladığı bu envanterle psikiyatri polikliniği hastalarından elde edilen ölçümlerin geçerliğini test etmiştir. Ölçümlerin geçerliğini test etmek amacıyla MMPI kişilik testinin Depresyon Skalasını (D) kriter olarak almış ve iki envanter arasındaki korelasyonu 0,63 ( $p<0,001$ ) bulmuştur. Hisli (1989) uyarladığı BDE'nin 259 üniversite öğrencisinden elde edilen ölçümler için de güvenirliğini ve geçerliğini test etmiştir. Ölçümlerin güvenirliği Cronbach- $\alpha$  ve yarıya bölme yöntemleriyle inceleyerek güvenirlik katsayılarını sırasıyla 0,80 ve 0,74 bulmuştur. Ölçüt geçerliği için yine MMPI-D'yi kullanarak iki envanter arasındaki korelasyonu 0,50 bulmuştur. Savaşır ve Hisli-Şahin (1997) bahsedilen bu iki uyarlamayı farklı adlandırmayı önererek, Teğin'in uyarlamasına Beck Depresyon Ölçeği (BDÖ), Hisli'nin uyarlamasına Beck Depresyon Envanteri (BDE) demişlerdir.

Bu araştırmada kullanılacak veriler Türkiye'de BDE'nin ikinci uyarlaması olan Hisli (1988) BDE uyarlaması ile toplanmıştır ve BDE EK-1'de verilmiştir. Araştırmada BDE ile elde edilen verilerin kullanılmasının nedeni BDE'nin hem dünyada hem de Türkiye'de sık kullanılan bir ölçme aracı olmasıdır. Beck, Steer ve Garbin (1988) yayınladıkları makalelerinde, BDE'nin psikiyatrik hastalarda ve normal örneklerde depresyon belirtilerinin şiddetini değerlendirmek amacıyla en sık kullanılan ölçme araçlarından biri

olduđuna ve ölçeđin geçerlik ve güvenilirliđinin klinik örneklemelerin geniş bir kısmında kökleştiđine deđinmiřlerdir.

Arařtırmada kullanılan BDE 4'lü likert tipi 21 maddeden oluřan ölçektir ve her madde 0-3 puan aralıđındadır. Ölçekten en az 0, en fazla 63 puan alınabilmektedir (Hisli, 1988; Hisli, 1989; Arkar ve řafak, 2004; Avřar, 2007). BDE'de puan aralıklarının hangi depresyon düzeylerini belirttiđine bakıldıđında;

4 ve altı = Depresyonu inkâr etme, uydurma iyilik; bu puan normal insanlar için düşüktür

05 – 09 Normal,

10 – 18 Hafif depresyona eğilimli,

19 – 29 řiddetli depresyona eğilimli,

30 – 63 řiddetli depresyon řeklinde sınıflandırılmıřtır (Avřar, 2007).

### **Ölçüm Güvenirliđi**

Bu çalışmada kullanılacak olan veri seti üzerinden Mergen vd. (2011) Cronbach  $\alpha$  katsayısını hesaplayarak 0,89 olarak hesaplamıřtır. Ölçüm güvenirliđinin yeterli olduđunun söylenebilmesi için Cronbach  $\alpha$  katsayısının 0,70'den daha yüksek olması beklenmektedir, bu nedenle çalışmada kullanılacak veriler için ölçümün güvenilir olduđu söylenir (Büyüköztürk, 2012).

Mergen vd. (2011)'in topladıkları veriler için hesaplamıř oldukları BDE'nin madde ve ölçek istatistikleri Tablo 1'de verilmiřtir.



Tablo 1. BDE Madde ve Ölçek İstatistikleri: Madde Ortalamaları, Madde Standart Sapmaları, Madde-Toplam Korelasyonları ( $r_{it}$ ), Örneklem Büyüklüğü, Ham Cronbach  $\alpha$ , Ölçek Ortalaması ve Ölçek Standart Sapması.\*

Madde	Ortalama	Std	$r_{it}$
1	.40	.68	.63
2	.34	.53	.49
3	.26	.60	.44
4	.50	.60	.62
5	.60	.61	.60
6	.42	.73	.50
7	.33	.59	.65
8	.49	.67	.57
9	.09	.38	.37
10	.50	.97	.45
11	.55	.80	.49
12	.46	.71	.62
13	.59	.78	.65
14	.23	.63	.45
15	.54	.64	.55
16	.47	.68	.49
17	.33	.64	.51
18	.24	.52	.42
19	.09	.33	.23
20	.28	.57	.32
21	.34	.73	.37
N		620	
Ham $\alpha$		.89	
St. $\alpha$		.89	
Ölçek Ortalaması		8.03	
Ölçek $S_x$		7.61	

\*: Bu tablo Mergen vd., 2011'den alınmıştır.

Tablo 1'de görüldüğü gibi madde toplam korelasyonları 0,23-0,65 arasında değişmektedir. Madde-toplam korelasyonları pozitif ve yüksekse maddeler benzer davranışları örneklemektedir ve iç tutarlık yüksektir denilebilir (Büyüköztürk, 2012).

KTK'ye göre yapılan bu analizlerin yanı sıra verilerin üretilmesinde kullanılan madde parametrelerinin elde edildiği ölçümler MULTILOG 7.03 programıyla MTK DTM'ye dayalı olarak analiz edildiğinde -2loglikelihood model veri uyumu değeri 8344,5 bulunmuştur ve bu değer pozitif olması model-veri uyumunun sağlandığını göstermektedir (Embretson ve Reise, 2000; Klein, 2005). Marjinal güvenilirlik katsayısı ise 0,9019 bulunmuştur. Buna dayanarak ölçümlerin güvenilirliğini yüksek olduğu söylenebilir.

## Verilerin Üretilmesi

Verilerin üretilmesi aşamasında, Uludağ Üniversitesi Aile Sağlık Merkezi'ne başvuran ve genellikle öğrencilerden oluşan 626 bireyden Mergen vd. (2011) tarafından Beck Depresyon Envanteri kullanılarak toplanan verilerin madde parametreleri kullanılmıştır. Madde parametreleri MULTLOG 7.03 (Thissen vd., 2003) programında derecelendirilmiş tepki modeline (DTM) dayalı olarak ve marjinal en çok olasılık (MML) kestirim yöntemi kullanılarak kestirilmiştir. Verilerin üretilmesinde kullanılan madde parametreleri Tablo 2'de verilmiştir.

Tablo 2. Verilerin Üretilmesinde Kullanılan Madde Parametreleri

Maddeler	a	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>
M1	2,93	0,2	1,21	1,79
M2	1,63	0,28	2,72	3,32
M3	1,78	0,81	1,57	2,96
M4	2,54	-0,18	1,55	2,59
M5	2,49	-0,45	1,44	2,63
M6	1,76	0,26	1,72	1,95
M7	3,15	0,25	1,51	1,94
M8	1,98	-0,07	1,55	2,33
M9	2,08	1,52	2,1	2,89
M10	1,45	0,6	1,33	1,49
M11	1,68	0,06	1,33	2,05
M12	2,34	0,09	1,24	1,99
M13	2,63	-0,11	0,76	2,17
M14	2,23	0,92	1,40	1,95
M15	1,88	-0,25	1,56	2,86
M16	1,67	0,04	1,97	2,23
M17	1,68	0,54	1,61	2,99
M18	1,36	0,94	2,45	4,05
M19	1,07	2,19	4,51	6,02
M20	1,1	0,95	3,00	3,96
M21	1,26	1,04	1,45	3,76

Verilerin üretilmesinde kullanılan madde parametrelerinin ortalamaları, minimum ve maksimum değerleri Tablo 3'te verilmiştir.

Tablo 3. Verilerin Üretilmesinde Kullanılan Madde Parametrelerine İlişkin Betimleyici İstatistikler

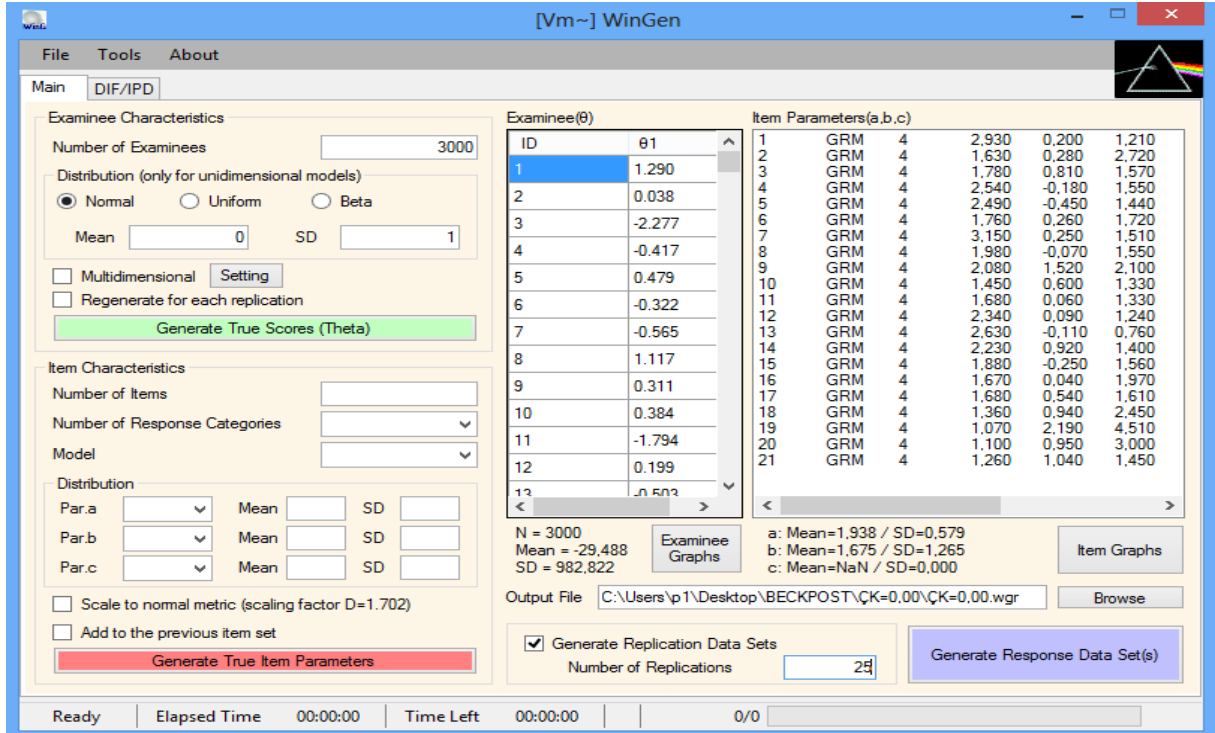
Madde Parametreleri	Ortalama	Minimum	Maksimum
a	1,94	1,07	3,15
b <sub>1</sub>	0,46	-0,45	2,19
b <sub>2</sub>	1,81	0,76	4,51
b <sub>3</sub>	2,76	1,49	6,02
b	1,40	-0,45	6,02

Gerçek verilerden elde edilen parametrelere dayalı olarak üretilen bu post-hoc simülasyon çalışmasında veri setlerinin üretilmesi amacıyla WinGen 3.1 (Han, 2007) simülasyon programı kullanılmıştır. Bu programla farklı modellerde, kategori sayılarında, dağılımlarda, örneklem büyüklüklerinde, test uzunluklarında ve parametre dağılımlarında veri üretmek mümkündür. Bu çalışmada sadece farklı dağılımlara sahip, 3000 kişilik ve 4 kategorili 21 maddeden oluşan veri setleri DTM'ne göre üretilmiştir. Araştırmada bulunan farklı dağılımlar ve bu dağılımların elde edilmesinde kullanılan parametreler Tablo 4'te sunulmuştur.

Tablo 4. Dağılımların Çarpıklık Düzeylerine Göre Verilerin Üretilmesinde Kullanılan Parametreler

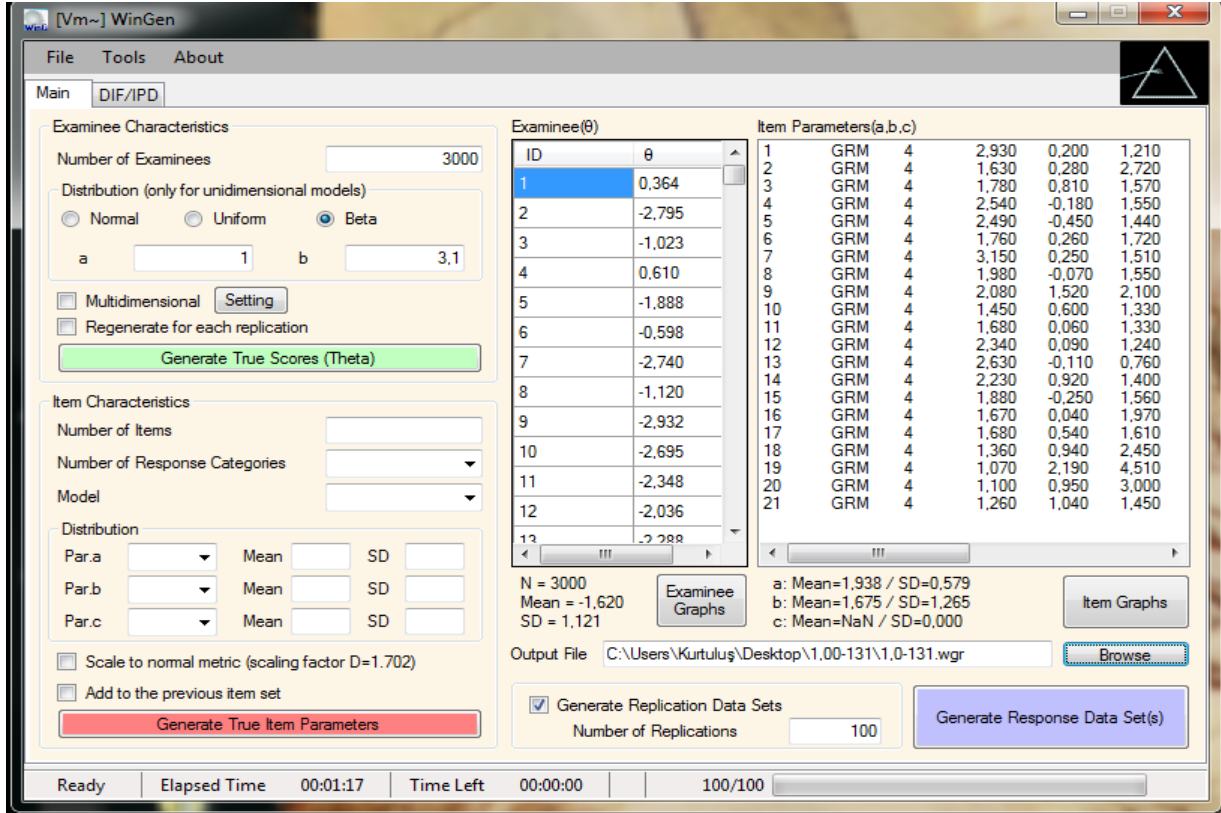
Çarpıklık Düzeyi	Kullanılan Parametreler	
	Ortalama	Standart Sapma
ÇK = 0,00	0	1
	$\alpha$	$\beta$
ÇK = 0,5	2	4
ÇK = 1,0	1	3,1
ÇK = 1,5	1	8
ÇK = 2,0	1	16

Araştırmada 5 dağılım türünde veri üretilmiştir ve her dağılım türü için 25 replikasyon kullanılmıştır. Toplam 125 veri setinin her birinde 3000 bireyin cevap örüntüleri ve gerçek birey parametreleri WinGen 3.1 programından elde edilmiştir. Bu programda 0,00 çarpıklık katsayısına sahip dağılım, 1,00 çarpıklık katsayısına sahip dağılım ve 2,00 çarpıklık katsayısına sahip dağılımın üretilmesi aşamasında görünen ara yüz sırasıyla Şekil 4, Şekil 5 ve Şekil 6'da verilmiştir.



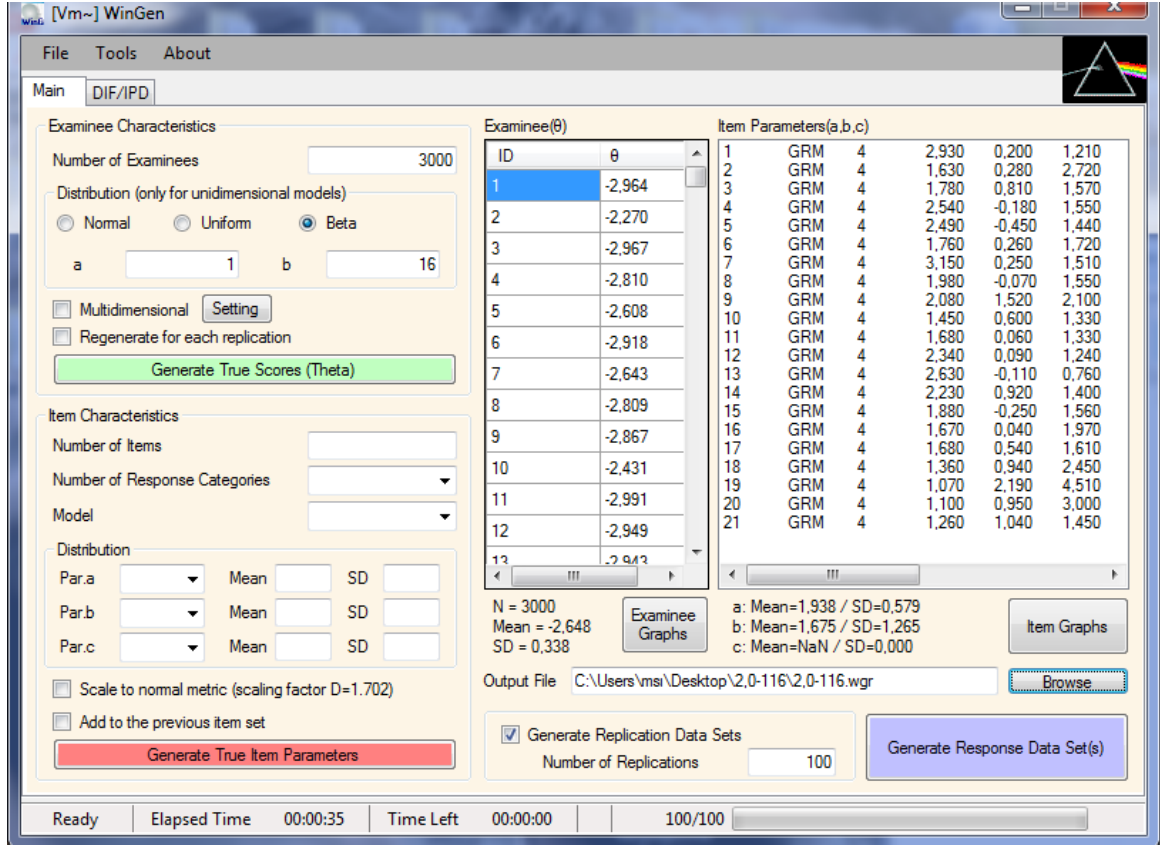
Şekil 4.  $\text{ÇK} = 0,00$  dağılımına sahip verilerin üretilmesinde WinGen 3.1. arayüzü

Şekil 4’te görüldüğü gibi  $\text{ÇK} = 0,00$  olan dağılımı Wingen3’te üretirken "distribution" bölümünde “normal” işaretlenmiştir. Daha sonra Tablo 4’te belirtildiği gibi programın istediği ortalama değeri 0, standart sapma değeri 1 olarak ilgili yerlere yazılmış ve  $N(0,1)$  dağılımı oluşturulmuştur. Bu işlemler birey parametresi üretmek amacıyla yapılmıştır ve sonrasında “generate true scores (theta)” butonu tıklanarak birey parametreleri (theta) üretilmiştir. Madde parametresi üretmek için ise “item characteristics” ve “distribution” bölümü yapılacak çalışmaya uygun olarak doldurulabilir. Ancak bu araştırma bir post hoc çalışması olduğu için madde parametreleri bellidir ve bu değerler Tablo 2’de verilmiştir. Bu değerlerle oluşturulan “.wgi” dosyası programda “file > open > item parameter data” seçilerek programa çekilmiştir ve bu parametreler arayüzde “Item Parameters (a, b, c)” bölümünde gözükmektedir. Bu bölümde birinci sütun madde numarasını, ikinci sütun modeli, üçüncü sütun kategori sayısını, dördüncü sütun a parametresini, beşinci ve altıncı sütun ise sırasıyla  $b_1$  ve  $b_2$  parametrelerini göstermektedir. Birey parametreleri üretilip madde parametreleri programa çekildikten sonra “.wgr” uzantılı ve “tools > options” butonlarında “Save a .csv (comma delimited) file” işaretlenerek oluşturulan “.csv” uzantılı veri dosyaları “Output File” bölümünde “Browse” tıklanarak belirlenen bir yere kaydedilmiştir. Ayrıca araştırma kapsamında 25 replikasyon yapıldığı için “Generate Replication Data Sets” işaretlenerek replikasyon sayısı yazılmıştır.



Şekil 5.  $\text{ÇK} = 1,00$  dağılımına sahip verilerin üretilmesinde WinGen 3.1 arayüzü

Şekil 5’te görüldüğü gibi  $\text{ÇK} = 1,00$  olan dağılımı Wingen3’te üretirken "distribution" bölümünde "beta" işaretlenmiştir. Beta dağılımının kullanımı çarpık dağılıma sahip veri simülasyonun kolaylaştırmaktadır ve beta dağılımını oluşturmak için ara yüzün "Examinee Characteristics" bölümündeki "distribution" kısmında a ve b parametreleri ( $\alpha$  ve  $\beta$ ) tanımlanmaktadır (Han ve Hambleton, 2007). Bu nedenle Tablo 4’te belirtildiği gibi programın istediği  $\alpha$  değeri 1,  $\beta$  değeri 3,1 olarak ilgili yerlere yazılmış ve "generate true scores (theta)" butonu tıklanarak birey parametreleri üretilmiştir. BDE’den elde edilen madde parametreleri kullanılarak oluşturulan ".wgi" dosyası programda "file > open > item parameter data" seçilerek programa çekilmiştir ve bu parametreler arayüzde "Item Parameters (a, b, c)" bölümünde gözükmektedir. Birey parametreleri üretilip madde parametreleri programa çekildikten sonra ".wgr" uzantılı ve "tools > options" butonlarında "Save a .csv (comma delimited) file" işaretlenerek oluşturulan ".csv" uzantılı veri dosyaları "Output File" bölümünde "Browse" tıklanarak belirlenen bir yere kaydedilmiştir. Ayrıca araştırma kapsamında 25 replikasyon yapıldığı için "Generate Replication Data Sets" işaretlenerek replikasyon sayısı yazılmıştır.



Şekil 6.  $\text{ÇK} = 2,00$  dağılımına sahip verilerin üretilmesinde WinGen 3.1 arayüzü

Şekil 6’te görüldüğü gibi  $\text{ÇK} = 2,00$  olan dağılımı Wingen3’te üretirken "distribution" bölümünde "beta" işaretlenmiştir. Daha sonra Tablo 4’te belirtildiği gibi programın istediği  $\alpha$  değeri 1,  $\beta$  değeri 16 olarak ilgili yerlere yazılmış ve "generate true scores (theta)" butonu tıklanarak birey parametreleri üretilmiştir. BDE’den elde edilen madde parametreleri kullanılarak oluşturulan ".wgi" dosyası programda "file > open > item parameter data" seçilerek programa çekilmiştir ve bu parametreler arayüzde "Item Parameters (a, b, c)" bölümünde gözükmektedir. Birey parametreleri üretilip madde parametreleri programa çekildikten sonra ".wgr" uzantılı ve "tools > options" butonlarında "Save a .csv (comma delimited) file" işaretlenerek oluşturulan ".csv" uzantılı veri dosyaları "Output File" bölümünde "Browse" tıklanarak belirlenen bir yere kaydedilmiştir. Ayrıca araştırma kapsamında 25 replikasyon yapıldığı için "Generate Replication Data Sets" işaretlenerek replikasyon sayısı yazılmıştır.

Araştırmada üretilen verilerin çarpıklık - basıklık katsayıları, ortalama  $\theta$  ve standart sapmaları Tablo 5’te verilmiştir.

Tablo 5. Her Dağılım Türü İçin Betimleyici İstatistikler

Dağılım Türü	ÇK	BK	$\bar{X}$	$S_x$
ÇK = 0,00	0,00	0,04	-0,02949	0,98282
ÇK = 0,50	0,50	-0,27	-1,00328	1,07377
ÇK = 1,00	1,00	0,48	-1,62023	1,12122
ÇK = 1,50	1,50	2,47	-2,32952	0,60998
ÇK = 2,00	2,00	6,19	-2,64832	0,33848

WinGen 3.1 programında Examinee Graphs > Plot Dialogs'dan elde edilen her bir dağılım türüne ait histogram dağılım diyagramları EK-2'de, kümülatif dağılım diyagramları ise EK-3'te verilmiştir.

Verilerin üretilmesinde kullanılan bazı maddelerin madde düzeyinde İKE'leri EK-4'te, yine bazı maddelerin madde bilgi fonksiyonları EK-5'te, test cevap fonksiyonu EK-6'da, test bilgi fonksiyonu ise EK-7'de verilmiştir.

### **MTK Varsayımlarının Kontrolü**

MTK varsayımlarından tek boyutluluk açımlayıcı faktör analizi ya da doğrulayıcı faktör analizi ile incelenebilir. Holgado-Tello, Chacón-Moscoso, Barbero-García ve Vila-Abad (2010) yaptıkları çalışmada likert tipi ölçekler için faktör analizinin pearson korelasyon katsayısı yerine polikorik (polychoric) korelasyon katsayısına dayanarak yapılmasının daha doğru sonuçlar verdiğini tespit etmişlerdir. Ancak bu araştırmada verilerin üretilmesi aşamasında Wingen 3.1 programında "Multidimensional" işaretlenmediği için tüm replikasyonlarda veri setleri tek boyutludur.

Veri setlerinin tek boyutlu olması aynı zamanda yerel bağımsızlık varsayımının da sağlandığını göstermektedir. Lord (1980)'e göre yerel bağımsızlık ek bir varsayım değildir ve tek boyutluluğun peşi sıra gelmektedir.

Normallik sayılıtısına değinilecek olursa; bu araştırma normallik sayılıtısının ihlalinin ölçme kesinliğini nasıl etkileyeceğini incelemek amacıyla yapıldığı için sayılıtının kontrolü yapılmamıştır.

Son varsayım olan testin hız testi olup olmadığı incelendiğinde çalışma verilerinin simülasyon ile üretilmesi nedeniyle ulaşılamayan ya da boş olan cevap bulunmadığından testin hız testi olmadığı söylenebilir (Hambleton ve Swaminathan, 1985).

### **Verilerin Çözümlemesi ve Yorumlanması**

Bu araştırmanın post-hoc simülasyon çalışması olması sebebiyle öncelikli olarak veri üretiminde kullanılacak madde parametreleri kestirilmiştir. Bu kestirimler MULTILOG

7.03 (Thissen vd., 2003) programıyla yapılmış ve BDE'deki 21 maddenin her biri için  $a$ ,  $b_1$ ,  $b_2$ ,  $b_3$  madde parametreleri elde edilmiştir. Bu gerçek parametre değerlerine dayanarak veri üretimi yapılmıştır.

Oluşturulan veri setlerinin çözümlenmesinde veriler çok kategorili puanlanan maddelerden oluşan ve likert tipi olan BDE ile toplandığı için derecelendirilmiş tepki modeli (DTM) kullanılmıştır (DeMars, 2010). DTM ile yapılacak bu kestirimler için maksimum M-step iterasyon sayısı 999'a ve maksimum EM cycle sayısı 99999'a çıkarılarak MULTILOG 7.03 (Thissen vd., 2003) programında MML kestirim yöntemi kullanılarak yapılmıştır. Bu program kullanılarak 5 dağılım türündeki 125 veri seti için marjinal güvenilirlik katsayısı, programın verdiği hata sayısı ve EM cycle yakınsaklık kriterini sağlamama durumunun kontrol edilmesi için "maximum intercycle parameter change" değeri ve veri setlerindeki her bir madde için madde parametreleri ( $a$ ,  $b_1$ ,  $b_2$ ,  $b_3$  parametreleri) elde edilmiştir. Bu parametreleri ve değerleri elde etmek için kullanılan örnek bir söz dizimi EK-8'de verilmiştir (Du Toit, 2003).

Ölçme kesinliğini inceleyebilmek adına MULTILOG 7.03'ten (Thissen vd., 2003) elde edilen "kestirilen madde parametreleri" nin yanı sıra "gerçek madde parametreleri" ne de ihtiyaç duyulmaktadır. Gerçek madde parametreleri; BDE'den elde edilerek verilerin üretilmesi aşamasında kullanılan parametrelerdir.

### 1. Alt Problem İçin Verilerin Analizi

Araştırmanın 1. alt probleminde ölçme kesinliği RMSE değerleri ile incelenmiştir ve RMSE her replikasyon için

$$RMSE = \sqrt{\frac{\sum_{i=1}^I (\hat{\delta}_i - \delta_i)^2}{I}}$$

genel formülüyle hesaplanmış ve farklı dağılımlara ait RMSE değerlerini birbirleriyle karşılaştırabilmek için ortalaması alınmıştır (Roberts, Donoghue ve Laughlin, 1999; Roberts, Donoghue ve Laughlin, 2002; Kieftenbeld ve Natesan (2012)). Formülde  $\hat{\delta}_i$  maddesi için kestirilen parametre değerini ifade ederken,  $\delta_i$  i maddesi için gerçek parametre değerini ifade eder. Formülde I ise madde sayısını,  $\delta$  ise ilgili madde ( $a$ ,  $b_1$ ,  $b_2$ ,  $b_3$ ) ve birey parametrelerinden ( $\theta$ ) birini ifade etmektedir.



### 1.1. Alt Problem İçin Verilerin Analizi

Araştırmanın 1.1. alt probleminde a parametresi için ölçme kesinliği RMSE değerleri ile incelenmiştir ve RMSE her replikasyon için

$$RMSE(a_i) = \sqrt{\frac{\sum_{i=1}^I (\hat{a}_i - a_i)^2}{I}}$$

formülüyle hesaplanmıştır. Burada,

$\hat{a}_i$  : i. madde için kestirilen a parametresini,

$a_i$  : i. madde için gerçek a parametresini,

I : Madde sayısını ifade etmektedir.

Tüm replikasyonlar için a parametrelerine ait RMSE değerleri hesaplandıktan sonra, bu değerlere dayanarak dağılım türleri karşılaştırılacağından ortalama RMSE hesaplanmıştır. RMSE değeri 0'da uzaklaştıkça kestirimlerin doğruluğu azalmaktadır. Diğer bir deyişle RMSE değerinin azalması ölçme kesinliğinin arttığının, artması ise ölçme kesinliğinin azaldığının bir göstergesidir ve bulgular buna göre yorumlanmıştır. Ayrıca Tate (2000), De Mars (2003) ve Sen (2014)'ün yaptıkları çalışmalarda olduğu gibi RMSE değerlerinin 0,1 değerinden büyük olup olmamasına göre yorumlanmıştır.

### 1.2. Alt Problem İçin Verilerin Analizi

Araştırmanın 1.2. alt probleminde b parametreleri için ölçme kesinliği RMSE değerleri ile incelenmiştir ve RMSE sırasıyla  $b_1$ ,  $b_2$  ve  $b_3$  parametrelerinde her replikasyon için

$$RMSE(b_{1i}) = \sqrt{\frac{\sum_{i=1}^I (\widehat{b}_{1i} - b_{1i})^2}{I}}$$

$$RMSE(b_{2i}) = \sqrt{\frac{\sum_{i=1}^I (\widehat{b}_{2i} - b_{2i})^2}{I}}$$

$$RMSE(b_{3i}) = \sqrt{\frac{\sum_{i=1}^I (\widehat{b}_{3i} - b_{3i})^2}{I}}$$

formülleriyle hesaplanmıştır. Burada,

$\widehat{b}_{1i}$ ,  $\widehat{b}_{2i}$ , ve  $\widehat{b}_{3i}$ : Sırasıyla i. madde için kestirilen  $b_1$ ,  $b_2$  ve  $b_3$  parametrelerini

$b_{1i}$ ,  $b_{2i}$  ve  $b_{3i}$ : Sırasıyla i. madde için gerçek  $b_1$ ,  $b_2$  ve  $b_3$  parametrelerini

I : Madde sayısını ifade etmektedir.

Tüm replikasyonlar için  $b_1$ ,  $b_2$  ve  $b_3$  parametrelerine ait RMSE değerleri hesaplandıktan sonra, bu değerlere dayanarak dağılım türleri karşılaştırılacağından ortalama RMSE hesaplanmıştır. RMSE değeri 0'da uzaklaştıkça kestirimlerin doğruluğu azalmaktadır. Diğer bir deyişle RMSE değerinin azalması ölçme kesinliğinin arttığının, artması ise ölçme kesinliğinin azaldığının bir göstergesidir ve bulgular buna göre yorumlanmıştır. Ayrıca Tate (2000), De Mars (2003) ve Sen (2014) yaptıkları çalışmalarda olduğu gibi RMSE değerlerinin 0,1 değerinden büyük olup olmamasına göre yorumlanmıştır.

## 2. Alt Problem İçin Verilerin Analizi

Willmott ve Matsuura (2005) bias'ın (mean bias error (MBE)) kullanışlı bilgi sağladığını ancak tipik hata büyüklüğüyle tutarsız bir şekilde ilişkili olduğu için dikkatli bir şekilde yorumlanması gerektiğini ve bias değerinin diğer istatistiklerden daha düşük olduğunu ( $MBE \leq MAE \leq RMSE$ ) ifade etmişlerdir. Bunun yanı sıra mean absolute error'un (MAE ya da literatürdeki diğer kısaltmalarıyla MAD, AAE ve AAD) RMSE'nin aksine ortalama hata büyüklüğünün en doğal ölçümü olduğunu göstermişlerdir. Bu nedenle 2. alt problemde ölçme kesinliği AAD değerleri ile incelenmiştir. De Ayala, Dodd ve Koch (1991) AAD'yi (ortalama mutlak farklılık) birey parametreleri için;

$$AAD = \frac{\sum_{i=1}^N |\widehat{\theta}_j - \theta_{T_j}|}{N}$$

olarak formüle etmişlerdir ve burada  $\widehat{\theta}_j$  j bireyi için kestirilen birey parametresini,  $\theta_{T_j}$  j. birey için gerçek birey parametresini, N ise birey sayısını ifade etmektedir. Madde parametresinde de her replikasyon için AAD kestirilen madde parametresi ile gerçek madde parametresi arasındaki farklılığın mutlak değerlerinin toplanması ve madde sayısına bölünmesiyle elde edilebilir ve

$$AAD = \frac{\sum_{i=1}^I |\widehat{\delta}_i - \delta_i|}{I}$$

şeklinde formüle edilebilir.  $\widehat{\delta}_i$  i maddesi için kestirilen parametre değerini,  $\delta_i$  i maddesi için gerçek parametre değerini, I ise madde sayısını ifade etmektedir ve  $\delta$  ilgili madde

parametrelerinden  $(a, b_1, b_2, b_3)$  birini simgelemektedir. AAD bu formüle dayanarak her replikasyon için hesaplanmış ve farklı dağılımlara ait AAD değerlerini birbirleriyle karşılaştırabilmek için ortalaması alınmıştır.

### 2.1. Alt Problem İçin Verilerin Analizi

Araştırmanın 2.1. alt probleminde  $a$  parametresi için ölçme kesinliği AAD değerleri ile incelenmiştir ve AAD her replikasyon için

$$AAD = \frac{\sum_{i=1}^I |\hat{a}_i - a_i|}{I}$$

formülüyle hesaplanmıştır. Burada,

$\hat{a}_i$  :  $i$ . madde için kestirilen  $a$  parametresini,

$a_i$  :  $i$ . madde için gerçek  $a$  parametresini,

$I$  : Madde sayısını ifade etmektedir.

Tüm replikasyonlar için  $a$  parametrelerine ait AAD değerleri hesaplandıktan sonra, bu değerlere dayanarak dağılım türleri karşılaştırılacağından ortalama AAD hesaplanmıştır. AAD değeri 0'dan uzaklaştıkça kestirimlerin doğruluğu azalmaktadır. Diğer bir deyişle AAD değerinin azalması ölçme kesinliğinin arttığına, artması ise ölçme kesinliğinin azaldığına bir göstergesidir ve bulgular buna göre yorumlanmıştır.

### 2.2. Alt Problem İçin Verilerin Analizi

Araştırmanın 2.2. alt probleminde  $b$  parametreleri için ölçme kesinliği AAD değerleri ile incelenmiştir ve AAD sırasıyla  $b_1, b_2$  ve  $b_3$  parametrelerinde her replikasyon için

$$AAD(b_{1i}) = \frac{\sum_{i=1}^I |\hat{b}_{1i} - b_{1i}|}{I}$$

$$AAD(b_{2i}) = \frac{\sum_{i=1}^I |\hat{b}_{2i} - b_{2i}|}{I}$$

$$AAD(b_{3i}) = \frac{\sum_{i=1}^I |\hat{b}_{3i} - b_{3i}|}{I}$$

formülleriyle hesaplanmıştır. Burada,

$\hat{b}_{1i}, \hat{b}_{2i},$  ve  $\hat{b}_{3i}$ : Sırasıyla  $i$ . madde için kestirilen  $b_1, b_2$  ve  $b_3$  parametrelerini

$b_{1i}, b_{2i}$  ve  $b_{3i}$ : Sırasıyla i. madde için gerçek  $b_1, b_2$  ve  $b_3$  parametrelerini

I : madde sayısını ifade etmektedir.

Tüm replikasyonlar için  $b_1, b_2$  ve  $b_3$  parametrelerine ait AAD değerleri hesaplandıktan sonra, bu değerlere dayanarak dağılım türleri karşılaştırılacağından ortalama AAD hesaplanmıştır. AAD değeri 0'dan uzaklaştıkça kestirimlerin doğruluğu azalmaktadır. Diğer bir deyişle AAD değerinin azalması ölçme kesinliğinin arttığı, artması ise ölçme kesinliğinin azaldığı bir göstergesidir ve bulgular buna göre yorumlanmıştır.

### 3. Alt Problem İçin Verilerin Analizi

Araştırmanın 3. alt probleminde marjinal güvenilirlik katsayıları her replikasyon için 7.03 (Thissen vd., 2003) programıyla elde edilmiştir ve dağılımları birbiriyle marjinal güvenilirlik açısından karşılaştırabilmek için her dağılıma ait 25 replikasyonun marjinal güvenilirliklerinin ortalaması hesaplanmıştır. Marjinal güvenilirlik katsayısı 0 ile 1 arasında değişmektedir ve bu değer 1'e yaklaştıkça güvenilirlik artmakta, 0'a yaklaştıkça ise güvenilirlik azalmaktadır. Bulguları yorumlama aşamasında farklı dağılımların güvenilirliğinin nasıl olduğu ve çarpıklık katsayısı ile normallik sayıltısının ihlalinin ölçüm güvenilirliğini nasıl etkilediği incelenmiştir.

### 4. Alt Problem İçin Verilerin Analizi

Araştırmanın 4. alt probleminde EMcycle yakınsaklık kriterini sağlamama yüzdesi MULTILog 7.03 programının sonuçlarından incelenmiştir. Programda EM cycle yakınsaklık kriteri 0,001 olarak belirlenmiştir ve "maximum intercycle parameter change" değeri bu değerden küçükse kriter sağlanmış, bu değerden büyükse kriter sağlanmamıştır. "EMcycle yakınsaklık kriterini sağlamama yüzdesi" kriteri sağlamayan replikasyon sayısının yüzdesi hesaplanarak elde edilmiştir ve

$$\text{EM cycle yakınsaklık kriterini sağlamama yüzdesi} = \frac{\text{Kriteri sağlamayan replikasyon sayısı}}{\text{Toplam replikasyon sayısı}} \times 100$$

şeklinde formüleleştirilebilir. Burada toplam replikasyon sayısı 25'tir ve bulguları yorumlama aşamasında EM cycle yakınsaklık kriterini sağlamama yüzdesinin çarpıklık katsayısından ve normallik sayıltısının ihlalinin nasıl etkilendiği incelenmiştir.

## 5. Alt Problem İin Verilerin Analizi

Samejima'nın (1969) DTM'sinde b parametrelerinin bklg arasındaki iliŐki  $b_1 < b_2 < b_3$  Őeklindedir. MULTILOG 7.03 programı kestirim yapamadıđı durumlarda "ERROR MESSAGE FROM SUBROUTINE INVSD: MATRIX IS SINGULAR" hatası vermektedir ve bu hata sayısı kadar maddede  $b_3$  parametresi  $b_2$ 'den kk olmaktadır ve 1,1 deđerinde gzkmektedir. Bu alt problem iin bulguları yorumlama aŐamasında arpıklık katsayısı arttıka hata veren replikasyon sayısının ve hata sayısının nasıl etkilendiđi incelenmiŐtir.





## BÖLÜM IV

### BULGULAR VE YORUM

Bu bölümde tüm araştırma problemlerine ilişkin bulgulara yer verilerek bu bulgular yorumlanmıştır.

#### 1. Alt Probleme İlişkin Bulgular

Araştırmanın 1. Alt problemi “Samejima’nın Derecelendirilmiş Tepki Modeli (DTM) ile madde parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı (ÇK) = 1,50 ve 2,00 olan dağılımlar) hata kareleri ortalamasının karekökünü (RMSE) nasıl etkiler?” olarak ifade edilmiştir. Bu alt problem 2 ayrı kademedede incelenecektir.

##### 1.1. Alt Probleme İlişkin Bulgular

“Samejima’nın Derecelendirilmiş Tepki Modeli ile a parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı (ÇK) = 1,50 ve 2,00 olan dağılımlar) hata kareleri ortalamasının karekökünü (RMSE) nasıl etkiler?” sorusuna yanıt aramak amacıyla yapılan analizlerin sonuçları Tablo 6’da verilmiştir.

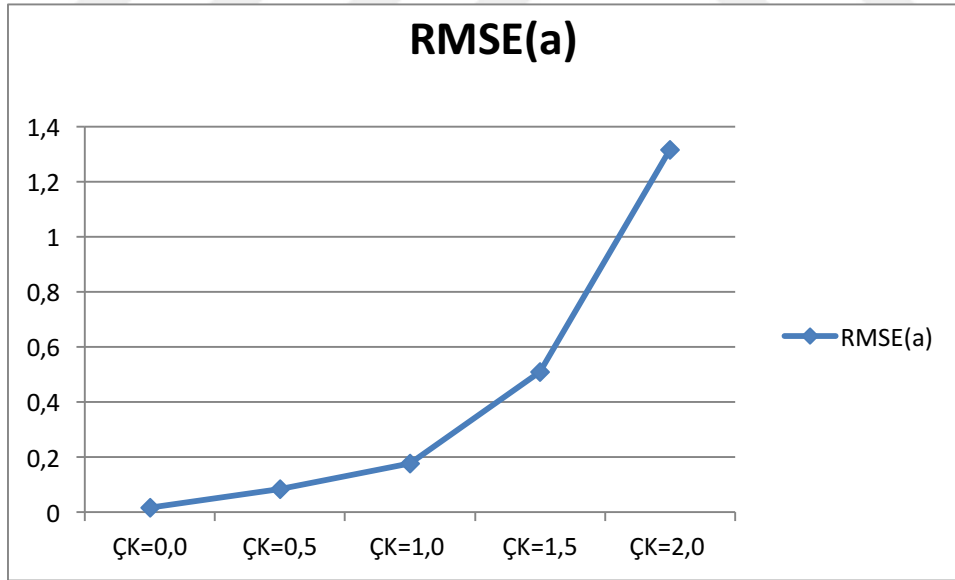
Tablo 6. Her Dağılım Türü İçin a Parametresi RMSE Değerleri

Dağılım Türü	a parametresi RMSE değerleri
ÇK = 0,00	0,017014
ÇK = 0,50	0,084615
ÇK = 1,00	0,178370
ÇK = 1,50	0,509082
ÇK = 2,00	1,317861

Tablo 6’daki RMSE ölçme kesinliği değerleri incelendiğinde en düşük RMSE’nin ÇK = 0,00 olan dağılımda (0,017014), en yüksek RMSE’nin ise ÇK = 2,00 olan dağılımda

(1,317861) olduğu görülmüştür. Ayrıca  $a$  parametresi için dağılımların çarpıklık katsayısı arttıkça RMSE kestiriminin de arttığı görülmektedir.  $\text{ÇK} = 0,00$  olan ve  $\text{ÇK} = 0,50$  olan dağılımlarda RMSE değeri 0,1'den düşüktür;  $\text{ÇK} = 1,00$ ,  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda 0,1'den yüksektir (Tate, 2000; De Mars, 2003; Sen, 2014).  $\text{ÇK} = 0,00$  ve  $\text{ÇK} = 0,50$  olan dağılımda RMSE kabul edilebilir düzeydeyken, diğer dağılımlar için yüksek değerler almaktadır. RMSE değerinin artması ölçme kesinliğinin azaldığını göstermektedir. Bu durumda  $\text{ÇK} = 0,00$  olan dağılımda ölçme kesinliği diğer dağılımlardan daha yüksek olduğu için gerçek parametrelerle kestirilen parametrelerin arasındaki fark daha azdır, bu kestirimlerin daha doğru olduğunun bir göstergesidir. Büyüköztürk (2012)'ye göre çarpıklık katsayısı  $\pm 1$  arasında kaldığı durumlarda ölçümler normal dağılımdan önemli bir sapma göstermemektedir. Buna göre normalliğin ihlal edildiği  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda RMSE değeri sırasıyla 0,509082 ve 1,317861'dir ve bu değerler oldukça yüksektir. Normalliğin ihlal edilmesi  $a$  parametresi kestirimlerini gerçek  $a$  parametresi değerlerinden uzaklaştırmakta ve ölçme kesinliği azalmaktadır.

Modeldeki  $a$  parametrelerinin her dağılım türü için RMSE değerleri grafiksel olarak Şekil 7'te verilmiştir.



Şekil 7.  $a$  parametresi için dağılım türü ile RMSE arasındaki ilişki

Şekil 7. incelendiğinde,  $\text{ÇK} = 1,00$  olan dağılım ile  $\text{ÇK} = 1,50$  olan dağılım arasındaki RMSE artışı, daha küçük çarpıklık katsayılarının olan dağılımlar arasındaki RMSE artışıyla karşılaştırıldığında daha büyüktür.  $\text{ÇK} = 0,00$  olan,  $\text{ÇK} = 0,50$  olan ve  $\text{ÇK} = 1,00$  olan dağılımların  $a$  parametresi için RMSE değerleri birbirine oldukça yakın olmasına rağmen,



$\text{ÇK} = 1,5$  olan ve  $\text{ÇK} = 2,00$  olan dağılımların  $a$  parametresi için RMSE değerleri diğer dağılımlara oranla oldukça yüksektir. Bu bağlamda çarpıklık katsayısı artarak dağılım normal dağılım türünden farklılaştığında  $a$  parametresinin RMSE değerlerinin de arttığı görülmektedir. Yani grafikte de görüldüğü gibi çarpıklık katsayısı arttıkça (normallikten fark arttıkça) RMSE artmaktadır ve çarpıklık katsayısı büyüdükçe RMSE'deki artış daha hızlı olmaktadır. Bu bulgu normallik ihlal edildiğinde ölçme kesinliğinin azaldığını kanıtlamaktadır.

### 1.2.Alt Probleme İlişkin Bulgular

" Samejima'nın Derecelendirilmiş Tepki Modeli ile  $b_1, b_2, b_3$  parametrelerinin kestiriminde normalliğin ihlali (çarpıklık katsayısı ( $\text{ÇK}$ ) = 1,50 ve 2,00 olan dağılımlar) hata kareleri ortalamasının karekökünü (RMSE) nasıl etkiler?" sorusuna yanıt aramak amacıyla yapılan analizlerin sonuçları Tablo 7'de verilmiştir.

Tablo 7. Her Dağılım Türü İçin  $b_1, b_2, b_3$  Parametresi RMSE Değerleri

Dağılım Türü	$b_1$ parametresi RMSE değerleri	$b_2$ parametresi RMSE değerleri	$b_3$ parametresi RMSE değerleri
$\text{ÇK} = 0,00$	0,010984	0,022638	0,039713
$\text{ÇK} = 0,50$	0,167868	0,129385	0,111942
$\text{ÇK} = 1,00$	0,234716	0,159561	0,172418
$\text{ÇK} = 1,50$	0,407426	1,087408	1,233695
$\text{ÇK} = 2,00$	42,56798	120,2494	73,47976

Tablo 7'deki RMSE değerleri incelendiğinde en düşük RMSE'nin normal dağılımda, en yüksek RMSE'nin ise  $\text{ÇK} = 2,00$  olan dağılımda olduğu görülmüştür. Ayrıca  $b_1, b_2, b_3$  parametreleri için dağılımların çarpıklık katsayısı arttıkça RMSE değerinin de arttığı ve  $\text{ÇK} = 2,00$  olan dağılımda RMSE'nin diğer dağılımların RMSE'lerine göre çok yüksek olduğu görülmektedir. Hata  $\text{ÇK}=2,00$  olan dağılım için  $b$  parametrelerinin kestirimin uygun olmadığı da anlaşılmaktadır.

Tablo 7'de  $b_1$  parametresi için RMSE ölçme kesinliği değerleri incelendiğinde en düşük RMSE'nin  $\text{ÇK} = 0,00$  olan dağılımda (0,010984), en yüksek RMSE'nin ise  $\text{ÇK} = 2,00$  olan dağılımda (42,56798) olduğu görülmüştür. Ayrıca  $b_1$  parametresi için dağılımların çarpıklık katsayısı arttıkça RMSE değerinin de arttığı görülmektedir.  $\text{ÇK} = 0,00$  olan dağılımda RMSE değeri 0,1'den düşükken;  $\text{ÇK} = 0,50, \text{ÇK} = 1,00, \text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda 0,1'den yüksektir (Tate, 2000; De Mars, 2003 ve Sen, 2014).  $\text{ÇK} = 0,00$  olan dağılımda RMSE kabul edilebilir düzeydeyken, diğer dağılımlar için yüksek değerler almaktadır. RMSE değerinin artması ölçme kesinliğinin azaldığını göstermektedir.

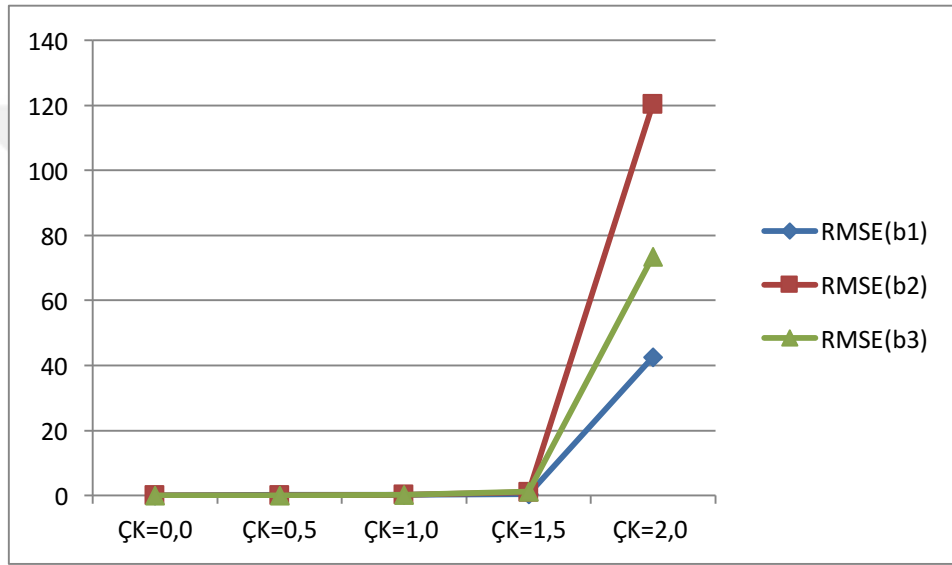
Bu durumda  $\text{ÇK} = 0,00$  olan dağılımda ölçme kesinliği diğer dağılımlardan daha yüksek olduğu için gerçek parametrelerle kestirilen parametrelerin arasındaki fark daha azdır, bu kestirimlerin daha doğru olduğunun bir göstergesidir. Normalliğin ihlal edildiği  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda RMSE değeri sırasıyla 0,407426 ve 42,56798'dir ve özellikle  $\text{ÇK} = 2,00$  olan dağılım için bu değerler oldukça yüksektir. Bu durum modelin uygun kestirim yapmadığını ve kullanılmaması gerektiğini kanıtlamaktadır. Normalliğin ihlal edilmesi  $b_1$  parametresi kestirimlerini gerçek  $b_1$  parametresi değerlerinden uzaklaştırmakta ve ölçme kesinliği azalmaktadır.

Modeldeki  $b_2$  parametresi için RMSE değerleri incelendiğinde en düşük RMSE'nin  $\text{ÇK} = 0,00$  olan dağılımda (0,022638), en yüksek RMSE'nin ise  $\text{ÇK} = 2,00$  olan dağılımda (120,2494) olduğu görülmüştür. Ayrıca  $b_2$  parametresi için dağılımların çarpıklık katsayısı arttıkça RMSE değerinin de arttığı görülmektedir.  $\text{ÇK} = 0,00$  olan dağılımda RMSE değeri 0,1'den düşüktür;  $\text{ÇK} = 0,50$ ,  $\text{ÇK} = 1,00$ ,  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda 0,1'den yüksektir (Tate, 2000; De Mars, 2003 ve Sen, 2014).  $\text{ÇK} = 0,00$  olan dağılımda RMSE kabul edilebilir düzeydeyken, diğer dağılımlar için yüksek değerler almaktadır. RMSE değerinin artması ölçme kesinliğinin azaldığını göstermektedir. Bu durumda  $\text{ÇK} = 0,00$  olan dağılımda ölçme kesinliği diğer dağılımlardan daha yüksek olduğu için gerçek parametrelerle kestirilen parametrelerin arasındaki fark daha azdır, bu kestirimlerin daha doğru olduğunun bir göstergesidir. Normalliğin ihlal edildiği  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda RMSE değeri sırasıyla 1,087408 ve 120,2494'tür ve özellikle  $\text{ÇK} = 2,00$  olan dağılım için bu değerler oldukça yüksektir. Normalliğin ihlal edilmesi  $b_2$  parametresi kestirimlerini gerçek  $b_2$  parametresi değerlerinden uzaklaştırmakta ve ölçme kesinliği azalmaktadır.

Modeldeki  $b_3$  parametresi için RMSE değerleri incelendiğinde ise en düşük RMSE'nin  $\text{ÇK} = 0,00$  olan dağılımda (0,039713), en yüksek RMSE'nin ise  $\text{ÇK} = 2,00$  olan dağılımda (73,47976) olduğu görülmüştür. Ayrıca  $b_3$  parametresi için dağılımların çarpıklık katsayısı arttıkça RMSE değerinin de arttığı görülmektedir.  $\text{ÇK} = 0,00$  olan dağılımda RMSE değeri 0,1'den düşüktür;  $\text{ÇK} = 0,50$ ,  $\text{ÇK} = 1,00$ ,  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda 0,1'den yüksektir (Tate, 2000; De Mars, 2003 ve Sen, 2014).  $\text{ÇK} = 0,00$  olan dağılımda RMSE kabul edilebilir düzeydeyken, diğer dağılımlar için yüksek değerler almaktadır. RMSE değerinin artması ölçme kesinliğinin azaldığını göstermektedir. Bu durumda  $\text{ÇK} = 0,00$  olan dağılımda ölçme kesinliği diğer dağılımlardan daha yüksek olduğu için gerçek

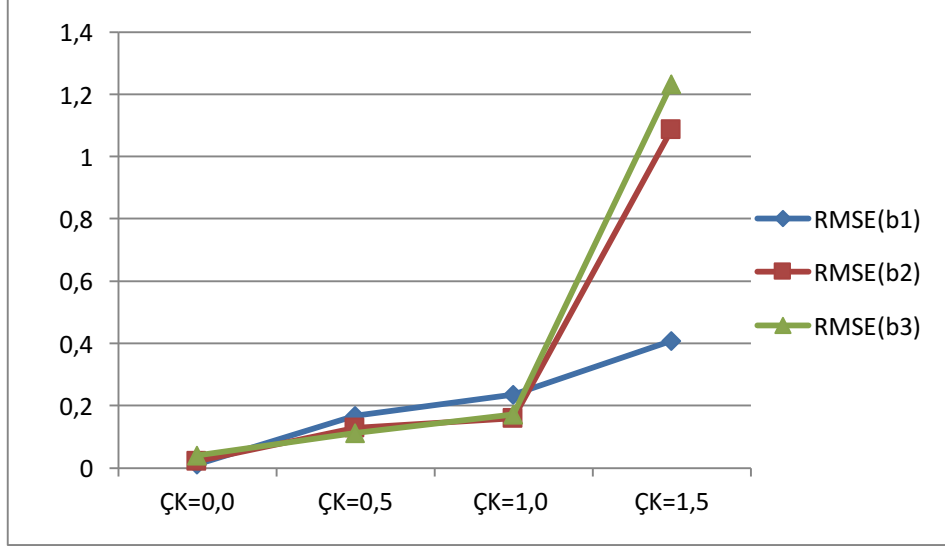
parametrelerle kestirilen parametrelerin arasındaki fark daha azdır, bu kestirimlerin daha doğru olduğunun bir göstergesidir. Normalliğin ihlal edildiği  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda RMSE değeri sırasıyla 1,233695 ve 73,47976'dır ve özellikle  $\text{ÇK} = 2,00$  olan dağılım için bu değerler oldukça yüksektir. Normalliğin ihlal edilmesi  $b_3$  parametresi kestirimlerini gerçek  $b_3$  parametresi değerlerinden uzaklaştırmakta ve ölçme kesinliği azalmaktadır.

DTM  $b_1$ ,  $b_2$ ,  $b_3$  parametrelerinin her dağılım türü için RMSE değerleri grafiksel olarak Şekil 8'de verilmiştir.



Şekil 8. DTM  $b_1$ ,  $b_2$ ,  $b_3$  parametreleri için dağılım türü ile RMSE arasındaki ilişki

Şekil 8 incelendiğinde,  $\text{ÇK} = 1,50$  olan dağılım ile  $\text{ÇK} = 2,00$  olan dağılım arasındaki RMSE artışı, diğer dağılımların arasındaki RMSE artışıyla karşılaştırıldığında oldukça büyüktür. Bu grafikte  $\text{ÇK} = 0,00$ ,  $\text{ÇK} = 0,50$ ,  $\text{ÇK} = 1,00$  ve  $\text{ÇK} = 1,50$  olan dağılımların  $b_1$ ,  $b_2$ ,  $b_3$  parametreleri için RMSE değerleri birbirine oldukça yakın görülmektedir ancak  $\text{ÇK} = 2,00$  olan dağılımın RMSE'sinin aşırı derecede yüksek olması nedeniyle bu dağılımlar arasındaki farklılık net olarak anlaşılmamaktadır. Bu nedenle Şekil 9'da  $\text{ÇK} = 2,00$  olan dağılıma ait RMSE değerinin bulunmadığı bir grafik verilmiştir.



Şekil 9. DTM  $b_1$ ,  $b_2$ ,  $b_3$  parametreleri için dağılım türü ile RMSE arasındaki ilişki (ÇK = 2,00 olan dağılım hariç)

Şekil 9 incelendiğinde, ÇK = 1,00 olan dağılım ile ÇK = 1,50 olan dağılım arasındaki RMSE artışı, diğer dağılımların arasındaki RMSE artışıyla karşılaştırıldığında özellikle  $b_2$  ve  $b_3$  parametreleri için daha büyüktür. ÇK = 0,00 olan, ÇK = 0,50 olan ve ÇK = 1,00 olan dağılımların  $b$  parametreleri için RMSE değerleri birbirine oldukça yakın olmasına rağmen, ÇK = 1,5 olan dağılımın  $b$  parametreleri için RMSE değerleri diğer dağılımlara oranla yüksektir. Bu bağlamda çarpıklık katsayısı artarak dağılım normal dağılım türünden farklılaştığında  $b_1$ ,  $b_2$ ,  $b_3$  parametrelerinin RMSE değerlerinin de arttığı söylenebilir. Bu durum normallik ihlal edildiğinde ölçme kesinliğinin azaldığını kanıtlamaktadır.

## 2. Alt Probleme İlişkin Bulgular

Bu araştırmanın 2. alt problemi “Samejima’nın Derecelendirilmiş Tepki Modeli ile madde parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı (ÇK) = 1,50 ve 2,00 olan dağılımlar) ortalama mutlak farklılığı (AAD) nasıl etkiler?” olarak ifade edilmiştir. Bu alt problem 2 ayrı kademede incelenecektir.

### 2.1. Alt Probleme İlişkin Bulgular

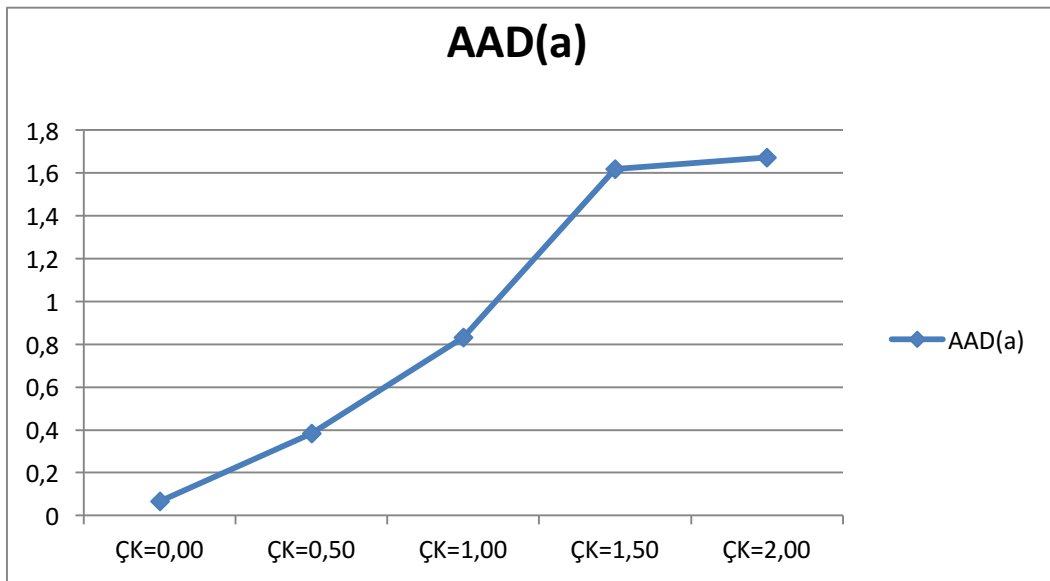
“Samejima’nın Derecelendirilmiş Tepki Modeli ile  $a$  parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı (ÇK) = 1,50 ve 2,00 olan dağılımlar) ortalama mutlak farklılığı (AAD) nasıl etkiler?” sorusuna yanıt aramak amacıyla yapılan analizlerin sonuçları Tablo 8’de verilmiştir.

Tablo 8. Her Dağılım Türü İçin a Parametresi AAD Değerleri

Dağılım Türü	a parametresi AAD değerleri
ÇK = 0,00	0,0672
ÇK = 0,50	0,3832
ÇK = 1,00	0,83139
ÇK = 1,50	1,617848
ÇK = 2,00	1,673562

Tablo 8'deki AAD değerleri incelendiğinde en düşük AAD'nin  $\text{ÇK} = 0,00$  olan dağılımda (0,0672), en yüksek AAD'nin ise  $\text{ÇK} = 2,00$  olan dağılımda (1,673562) olduğu görülmüştür. Ayrıca a parametresi için dağılımların çarpıklık katsayısı arttıkça AAD değerinin de arttığı görülmektedir. AAD değerinin artması ölçme kesinliğinin azaldığını göstermektedir. Bu durumda  $\text{ÇK} = 0,00$  olan dağılımda ölçme kesinliği diğer dağılımlardan daha yüksek olduğu için gerçek parametrelerle kestirilen parametrelerin arasındaki fark daha azdır, bu kestirimlerin daha doğru olduğunun bir göstergesidir. Büyüköztürk (2012)'ye göre çarpıklık katsayısı  $\pm 1$  arasında kaldığı durumlarda ölçümler normal dağılımdan önemli bir sapma göstermemektedir. Buna göre normalliğin ihlal edildiği  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda AAD değeri sırasıyla 1,617848 ve 1,673562'dir ve bu değerler oldukça yüksektir. Normalliğin ihlal edilmesi a parametresi kestirimlerini gerçek a parametresi değerlerinden uzaklaştırmakta ve ölçme kesinliği azalmaktadır.

Şekil 10'da a parametrelerinin her dağılım türü için AAD değerleri grafiksel olarak verilmiştir.



Şekil 10. a parametresi için dağılım türü ile AAD arasındaki ilişki

Şekil 10. incelendiğinde,  $\text{ÇK} = 1,00$  olan dağılım ile  $\text{ÇK} = 1,50$  olan dağılım arasındaki AAD artışı, diğer dağılımların arasındaki AAD artışıyla karşılaştırıldığında daha büyüktür.  $\text{ÇK} = 1,5$  olan ve  $\text{ÇK} = 2,00$  olan dağılımların a parametresi için AAD değerleri diğer dağılımlara oranla yüksek ve birbirlerine yakındır. Bu bağlamda çarpıklık katsayısı artarak dağılım normal dağılım türünden farklılaştığında a parametresinin AAD değerlerinin de arttığı söylenebilir. Bu durum normalliğin ihlal edildiğinde ölçme kesinliğinin azaldığı şeklinde yorumlanabilir.

## 2.2. Alt Probleme İlişkin Bulgular

“Samejima’nın Derecelendirilmiş Tepki Modeli ile  $b_1, b_2, b_3$  parametrelerinin kestiriminde normalliğin ihlali (çarpıklık katsayısı ( $\text{ÇK}$ ) = 1,50 ve 2,00 olan dağılımlar) ortalama mutlak farklılığı (AAD) nasıl etkiler?” sorusuna yanıt aramak amacıyla yapılan analizlerin sonuçları Tablo 9’da verilmiştir.

Tablo 9. Her Dağılım Türü İçin  $b_1, b_2, b_3$  Parametresi AAD Değerleri

Dağılım Türü	$b_1$ parametresi AAD değerleri	$b_2$ parametresi AAD değerleri	$b_3$ parametresi AAD değerleri
$\text{ÇK} = 0,00$	0,040476	0,079638	0,126495
$\text{ÇK} = 0,50$	0,83141	0,62699	0,507638
$\text{ÇK} = 1,00$	1,157848	0,77019	0,585676
$\text{ÇK} = 1,50$	1,638019	2,772248	4,155143
$\text{ÇK} = 2,00$	22,09926	56,61021	36,2361

Tablo 9’deki AAD ölçme kesinliği değerleri incelendiğinde en düşük AAD’nin normal dağılımda, en yüksek AAD’nin ise  $\text{ÇK} = 2,00$  olan dağılımda olduğu görülmüştür. Ayrıca  $b_1, b_2, b_3$  parametreleri için dağılımların çarpıklık katsayısı arttıkça AAD değerlerinin arttığı ve  $\text{ÇK} = 2,00$  olan dağılımda AAD’nin diğer dağılımların AAD’lerine göre çok yüksek olduğu görülmüştür.

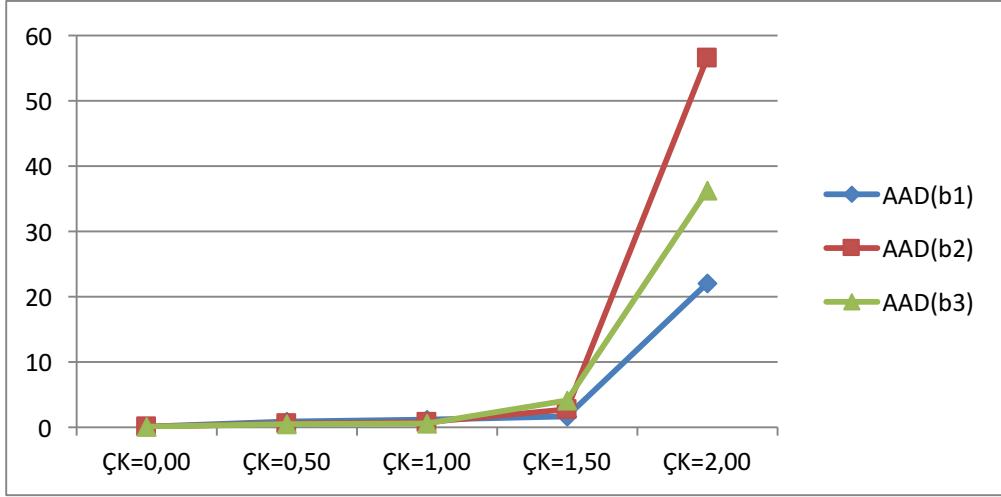
Tablo 9’da  $b_1$  parametresi için AAD ölçme kesinliği değerleri incelendiğinde en düşük AAD’nin  $\text{ÇK} = 0,00$  olan dağılımda (0,040476), en yüksek AAD’nin ise  $\text{ÇK} = 2,00$  olan dağılımda (22,09926) olduğu görülmüştür. Ayrıca  $b_1$  parametresi için dağılımların çarpıklık katsayısı arttıkça AAD değerinin de arttığı görülmektedir. AAD değerinin artması ölçme kesinliğinin azaldığını göstermektedir. Bu durumda  $\text{ÇK} = 0,00$  olan dağılımda ölçme kesinliği diğer dağılımlardan daha yüksek olduğu için gerçek

parametrelerle kestirilen parametrelerin arasındaki fark daha azdır, bu kestirimlerin daha doğru olduğunun bir göstergesidir. Normalliğin ihlal edildiği  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda AAD değeri sırasıyla 1,638019 ve 22,09926'dır ve özellikle  $\text{ÇK} = 2,00$  olan dağılım için bu değerler oldukça yüksektir. Normalliğin ihlal edilmesi  $b_1$  parametresi kestirimlerini gerçek  $b_1$  parametresi değerlerinden uzaklaştırmakta ve ölçme kesinliği azalmaktadır.

DTM  $b_2$  parametresi için AAD ölçme kesinliği değerleri incelendiğinde en düşük AAD'nin  $\text{ÇK} = 0,00$  olan dağılımda (0,079638), en yüksek AAD'nin ise  $\text{ÇK} = 2,00$  olan dağılımda (56,61021) olduğu görülmüştür. Ayrıca  $b_2$  parametresi için dağılımların çarpıklık katsayısı arttıkça AAD değerinin de arttığı görülmektedir. AAD değerinin artması ölçme kesinliğinin azaldığını göstermektedir. Bu durumda  $\text{ÇK} = 0,00$  olan dağılımda ölçme kesinliği diğer dağılımlardan daha yüksek olduğu için gerçek parametrelerle kestirilen parametrelerin arasındaki fark daha azdır, bu kestirimlerin daha doğru olduğunun bir göstergesidir. Normalliğin ihlal edildiği  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda AAD değeri sırasıyla 2,772248 ve 56,61021'dir ve özellikle  $\text{ÇK} = 2,00$  olan dağılım için bu değerler oldukça yüksektir. Normalliğin ihlal edilmesi  $b_2$  parametresi kestirimlerini gerçek  $b_2$  parametresi değerlerinden uzaklaştırmakta ve ölçme kesinliği azalmaktadır.

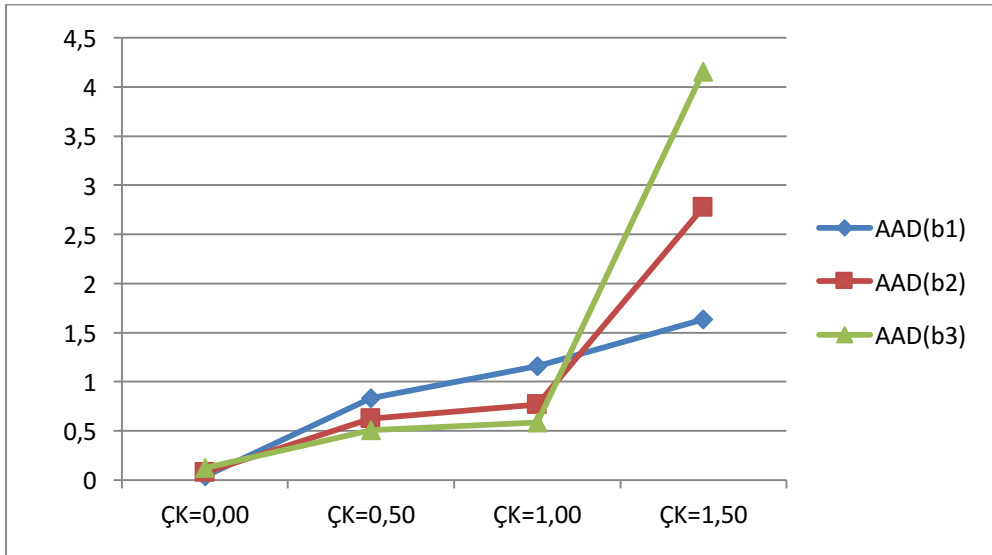
DTM  $b_3$  parametresi için AAD ölçme kesinliği değerleri incelendiğinde ise en düşük AAD'nin  $\text{ÇK} = 0,00$  olan dağılımda (0,126495), en yüksek AAD'nin ise  $\text{ÇK} = 2,00$  olan dağılımda (36,2361) olduğu görülmüştür. Ayrıca  $b_3$  parametresi için dağılımların çarpıklık katsayısı arttıkça AAD değerinin de arttığı görülmektedir. AAD değerinin artması ölçme kesinliğinin azaldığını göstermektedir. Bu durumda  $\text{ÇK} = 0,00$  olan dağılımda ölçme kesinliği diğer dağılımlardan daha yüksek olduğu için gerçek parametrelerle kestirilen parametrelerin arasındaki fark daha azdır, bu kestirimlerin daha doğru olduğunun bir göstergesidir. Normalliğin ihlal edildiği  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlarda AAD değeri sırasıyla 4,155143 ve 36,2361'dir ve özellikle  $\text{ÇK} = 2,00$  olan dağılım için bu değerler oldukça yüksektir. Normalliğin ihlal edilmesi  $b_3$  parametresi kestirimlerini gerçek  $b_3$  parametresi değerlerinden uzaklaştırmakta ve ölçme kesinliği azalmaktadır.

MTK  $b_1$ ,  $b_2$ ,  $b_3$  parametrelerinin her dağılım türü için AAD değerleri grafiksel olarak Şekil 11'de verilmiştir.



Şekil 11.  $b_1$ ,  $b_2$ ,  $b_3$  parametreleri için dağılım türü ile AAD arasındaki ilişki

Şekil 11 incelendiğinde,  $\text{ÇK} = 1,50$  olan dağılım ile  $\text{ÇK} = 2,00$  olan dağılım arasındaki AAD artışı, diğer dağılımların arasındaki AAD artışıyla karşılaştırıldığında oldukça büyüktür. Bu grafikte  $\text{ÇK} = 0,00$ ,  $\text{ÇK} = 0,50$ ,  $\text{ÇK} = 1,00$  dağılımların  $b_1$ ,  $b_2$ ,  $b_3$  parametreleri için AAD değerleri birbirine oldukça yakın görülmektedir ancak  $\text{ÇK} = 2,00$  olan dağılımın AAD'sinin aşırı derecede yüksek olması nedeniyle bu dağılımlar arasındaki farklılık net olarak anlaşılmamaktadır. Bu nedenle Şekil 12'de  $\text{ÇK} = 2,00$  olan dağılıma ait AAD değerinin bulunmadığı bir grafik verilmiştir.



Şekil 12.  $b_1$ ,  $b_2$ ,  $b_3$  parametreleri için dağılım türü ile AAD arasındaki ilişki ( $\text{ÇK} = 2,00$  olan dağılım hariç)



Şekil 12 incelendiğinde,  $\text{ÇK} = 1,00$  olan dağılım ile  $\text{ÇK} = 1,50$  olan dağılım arasındaki AAD artışı, diğer dağılımların arasındaki AAD artışıyla karşılaştırıldığında özellikle  $b_2$  ve  $b_3$  parametreleri için daha büyüktür.  $\text{ÇK} = 0,00$  olan dağılımın AAD'sinin 0'a oldukça yakın olmasıyla birlikte  $\text{ÇK} = 0,00$ ,  $\text{ÇK} = 0,50$  ve  $\text{ÇK} = 1,00$  olan dağılımların  $b$  parametreleri için AAD değerleri birbirine yakındır ancak  $\text{ÇK} = 1,5$  olan dağılımın  $b$  parametreleri için AAD değerleri diğer dağılımlara oranla yüksektir. Bu bağlamda çarpıklık katsayısı artarak dağılım normal dağılım türünden farklılaştığında  $b_1$ ,  $b_2$ ,  $b_3$  parametrelerinin AAD değerlerinin de arttığı görülmektedir. Bu bulgu normallik ihlal edildiğinde ölçme kesinliğinin azaldığını kanıtlamaktadır.

### 3. Alt Probleme İlişkin Bulgular

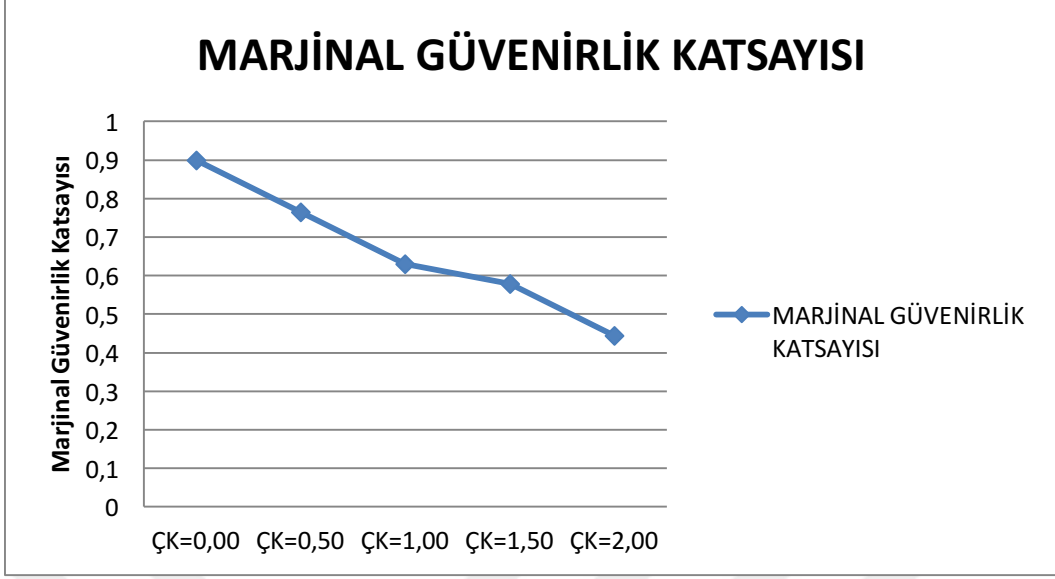
“Samejima'nın Derecelendirilmiş Tepki Modeli ile madde parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı ( $\text{ÇK}$ ) = 1,50 ve 2,00 olan dağılımlar) marjinal güvenilirliği nasıl etkiler?” sorusuna yanıt aramak amacıyla yapılan analizlerin sonuçları Tablo 10'da verilmiştir.

Tablo 10. Her Dağılım Türü İçin Marjinal Güvenirlik Katsayısı

Dağılım Türü	Marjinal Güvenirlik Katsayısı
$\text{ÇK} = 0,00$	0,898764
$\text{ÇK} = 0,50$	0,764116
$\text{ÇK} = 1,00$	0,629808
$\text{ÇK} = 1,50$	0,578908
$\text{ÇK} = 2,00$	0,443537

Tablo 10 incelendiğinde en düşük marjinal güvenilirlik katsayısı  $\text{ÇK} = 2,00$  olan dağılımda gözlenirken, en yüksek marjinal güvenilirlik katsayısı  $\text{ÇK} = 0,00$  olan dağılımda gözlenmiştir. Ayrıca çarpıklık katsayısı arttıkça marjinal güvenilirlik katsayısının azaldığı da görülmektedir.  $\text{ÇK} = 0,00$  olan yani  $N(0,1)$  dağılımında marjinal güvenilirlik oldukça yüksektir.  $\text{ÇK} = 0,50$  ve  $\text{ÇK} = 1,00$  olan dağılımların marjinal güvenilirlik katsayısı ise kabul edilir bir düzeydedir (de Ayala, 2009). Normalliğin ihlal edildiği durumlarda ( $\text{ÇK} = 1,50$  olan ve  $\text{ÇK} = 2,00$  olan dağılımlarda) özellikle  $\text{ÇK} = 2,00$  olan dağılımda olmak üzere güvenilirlik katsayısı diğer dağılımlara göre daha düşüktür.

Her dağılım türü ile marjinal güvenilirlik katsayısı arasındaki ilişki grafiksel olarak Şekil 13'te verilmiştir.



Şekil 13. Dağılım türleri ile marjinal güvenilirlik katsayısı arasındaki ilişki

Şekil 13 incelendiğinde çarpıklık katsayısı arttıkça marjinal güvenilirlik katsayısının azaldığı görülmüştür. Bu bulgu puan dağılımı normal dağılımdan  $(N(0,1))$  uzaklaştıkça marjinal güvenilirlik katsayısının da azaldığını kanıtlamaktadır.

#### 4. Alt Probleme İlişkin Bulgular

“Samejima’nın Derecelendirilmiş Tepki Modeli ile madde parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı (ÇK) = 1,50 ve 2,00 olan dağılımlar) EM cycle yakınsaklık kriterini sağlamama yüzdesini nasıl etkiler?” sorusuna yanıt aramak amacıyla her dağılıma ait tüm replikasyonlar için maximum intercycle parameter change değerleri Tablo 11’de verilmiştir.

Tablo 11. Her Dağılım Türü ve Replikasyon İçin Maximum Intercycle Parameter Change Değerleri

Replikasyon Numarası	ÇK=0,0	ÇK=0,5	ÇK=1,0	ÇK=1,5	ÇK=2,0
1	0,00094	0,00094	0,00097	0,0008	1,33071
2	0,00090	0,00086	0,00090	0,00513	1,07737
3	0,00090	0,00097	0,00084	0,00062	0,00055
4	0,00091	0,00083	0,00094	0,00056	0,01765
5	0,00093	0,00085	0,00086	0,00018	0,01064
6	0,00092	0,00088	0,00086	0,0062	1,28342
7	0,00094	0,00094	0,00087	0,09995	0,3296
8	0,00092	0,00095	0,00089	0,00105	3,61278
9	0,00093	0,00095	0,00093	0,00040	0,05770
10	0,00093	0,00086	0,00094	0,00152	0,04183
11	0,00089	0,00097	0,00089	0,00128	0,03336
12	0,00090	0,00089	0,00091	0,00039	0,03051
13	0,00092	0,00090	0,00089	0,00055	0,02403
14	0,00089	0,00091	0,00093	0,01311	0,00520
15	0,00090	0,00092	0,00095	0,00018	0,02550
16	0,00093	0,00091	0,00087	0,01900	0,00040
17	0,00093	0,00088	0,00090	0,00233	0,01019
18	0,00090	0,00092	0,00098	0,00079	0,00280
19	0,00092	0,00086	0,00090	0,21721	0,05160
20	0,00089	0,00089	0,00089	0,00039	0,04799
21	0,00090	0,00085	0,00096	0,00354	0,00257
22	0,00090	0,00090	0,00087	0,02478	0,04431
23	0,00091	0,00094	0,00085	0,03994	0,00289
24	0,00091	0,00089	0,00084	0,00013	0,06381
25	0,00092	0,00094	0,00099	1,16913	0,06127

EM cycle yakınsaklık kriteri MULTILOG 7.03 programında 0,001 olarak belirlenmiştir. Tablo 11'de verilen maximum intercycle parameter change değerleri incelendiğinde ÇK = 0,00 olan, ÇK = 0,50 olan ve ÇK = 1,00 olan dağılımlarda 0,001'den yüksek bir maximum intercycle parameter change değeri bulunmazken, ÇK = 1,50 ve ÇK = 2,00 olan dağılımlarda 0,001'den yüksek maximum intercycle parameter change değerleri bulunmaktadır. Kriteri sağlayan ve sağlamayan replikasyon frekansı ve yüzdesi Tablo 12'de verilmiştir.

Tablo 12. Her Dağılım İçin Kriteri Sağlamayan Replikasyonların Sayısı ve Yüzdeleri

Dağılım Türü	Kriteri sağlamayan replikasyon sayısı	Kriteri sağlamayan replikasyon yüzdesi
ÇK = 0,00	0	%0
ÇK = 0,50	0	%0
ÇK = 1,00	0	%0
ÇK = 1,50	14	%56
ÇK = 2,00	23	%92

Tablo 12 incelendiğinde EM cycle yakınsaklık kriterini sağlamayan yani 0,001 değerinden yüksek olan replikasyon sayısı normalliğin ihlal edilmediği durumlarda (ÇK = 0,00, ÇK = 0,50 ve ÇK = 1,00 olan dağılımlarda) 0 iken normalliğin ihlal edildiği durumlarda (ÇK = 1,50 ve ÇK = 2,00 olan dağılımlarda) sırasıyla 14 ve 23'tür. Normalliğin ihlal edilmediği dağılımlarda tüm replikasyonlarda kriter sağlanırken, normalliğin sağlanmadığı durumlarda çarpıklık katsayısı arttıkça kriter sağlanmayan replikasyon yüzdesi artmaktadır. Kriter sağlamayan replikasyon yüzdeleri ÇK = 1,50 olan ve ÇK = 2,00 olan dağılımlar için sırasıyla %56 ve %92'dir. ÇK = 2,00 olan dağılım için neredeyse tüm replikasyonlarda bu kriter sağlanamamıştır.

### 5. Alt Probleme İlişkin Bulgular

“Samejima'nın Derecelendirilmiş Tepki Modeli ile madde parametresi kestiriminde normalliğin ihlali (çarpıklık katsayısı (ÇK) = 1,50 ve 2,00 olan dağılımlar) MULTILOG programının verdiği hata sayısını nasıl etkiler?” sorusuna yanıt aramak amacıyla her dağılıma ait tüm replikasyonlar için hata frekansları Tablo 13'te verilmiştir.

Tablo 13. Her Dağılım Türü İçin Tüm Replikasyonlar MULTILOG Programının Verdiği Hata Sayıları

Replikasyon Numarası	ÇK=0,00 Olan	ÇK=0,5 Olan	ÇK=1,00 Olan	ÇK=1,5 Olan	ÇK=2,0 Olan
1	0	0	0	1	6
2	0	0	0	0	6
3	0	0	0	1	7
4	0	0	0	0	8
5	0	0	0	1	9
6	0	0	0	2	7
7	0	0	0	1	6
8	0	0	0	1	7
9	0	0	0	1	8
10	0	0	0	0	11
11	0	0	0	2	8
12	0	0	0	1	8
13	0	0	0	2	10
14	0	0	0	2	11
15	0	0	0	1	8
16	0	0	0	1	5
17	0	0	0	1	6
18	0	0	0	1	7
19	0	0	0	0	7
20	0	0	0	2	9
21	0	0	0	1	10
22	0	0	0	1	11
23	0	0	0	1	10
24	0	0	0	4	7
25	0	0	0	2	6

Tablo 13 incelendiğinde normalliğin ihlal edilmediği  $\text{ÇK} = 0,00$ ,  $\text{ÇK} = 0,5$  ve  $\text{ÇK} = 1,00$  olan dağılımlarda hiç bir replikasyon için MULTILOG 7.03 programının "ERROR MESSAGE FROM SUBROUTINE INVSD: MATRIX IS SINGULAR" hatası vermediği görülmektedir. Bu, normalliğin ihlal edilmediği dağılımlarda tüm kestirimlerin yapıldığı ve b parametrelerinin  $b_1 < b_2 < b_3$  şeklinde sıralandığı anlamına gelmektedir ve bu dağılımlarda  $b_3$  parametresi hata kaynaklı olarak 1,1 değerinde kestirilmemiştir. Bu durumun aksine normalliğin ihlal edildiği  $\text{ÇK} = 1,5$  ve  $\text{ÇK} = 2,0$  olan dağılımlarda hata veren replikasyonlar vardır. Çarpıklık katsayısının replikasyonlardaki hata sayılarını nasıl etkilediğini incelemek için her dağılım türüne ait ortalama hata sayıları Tablo 14'te verilmiştir.

Tablo 14. Her Dağılım Türü İçin Replikasyonlardaki Ortalama Hata Sayısı

Dağılım Türü	Ortalama Hata Sayısı
ÇK = 0,00	0
ÇK = 0,50	0
ÇK = 1,00	0
ÇK = 1,50	1,2
ÇK = 2,00	7,92

Tablo 14 incelendiğinde normalliğin ihlal edildiği durumlarda çarpıklık katsayısının arttığı ve ÇK = 1,50 olan dağılımın ortalama hata sayısının yaklaşık 1 olduğu ÇK = 2,00 olan dağılımının ise yaklaşık 8 olduğu görülmektedir. Normalliğin ihlal edilmediği durumlarda ise daha önce de belirtildiği gibi hata veren replikasyon yoktur ve bu nedenle ortalama hata sayısı 0'dır. ÇK = 2,00 olan dağılımdaki hata sayısının ÇK = 1,5 olan dağılımın hata sayısının yaklaşık 7 katı olması normallikten saptıkça programın kestirim yapamayarak "ERROR MESSAGE FROM SUBROUTINE INVSD:MATRIX IS SINGULAR" hatası verdiğini ve bu hata sayısı kadar maddede b parametresi kestirimlerinin  $b_1 < b_2 < b_3$  şeklinde sıralanmadığını göstermektedir.

Her replikasyon için hata sayısının ve ortalama hata sayısının yanı sıra hata veren ve vermeyen replikasyonların frekansları ve yüzdelerini incelemek amacıyla Tablo 15 aşağıda verilmiştir.

Tablo 15. Her Dağılım Türü İçin Hata Veren ve Vermeyen Replikasyon Sayısı ve Yüzdesi

Dağılım Türü	Hata veren replikasyon sayısı	Hata veren replikasyon yüzdesi
ÇK = 0,00	0	%0
ÇK = 0,50	0	%0
ÇK = 1,00	0	%0
ÇK = 1,50	21	%84
ÇK = 2,00	25	%100

Tablo 15 incelendiğinde "ERROR MESSAGE FROM SUBROUTINE INVSD:MATRIX IS SINGULAR" hatası veren replikasyon sayısı normalliğin ihlal edilmediği durumlarda (ÇK = 0,00, ÇK = 0,50 ve ÇK = 1,00 olan dağılımlarda) 0 iken normalliğin ihlal edildiği durumlarda (ÇK = 1,50 ve ÇK = 2,00 olan dağılımlarda) sırasıyla 21 ve 25'tir. Normalliğin ihlal edilmediği dağılımlarda hiç bir replikasyonda hata söz konusu değilken, normallik sayılıtısının sağlanmadığı durumlarda çarpıklık katsayısı arttıkça hata veren replikasyon yüzdesi artmaktadır. Hata veren replikasyon yüzdeleri ÇK = 1,50 olan ve ÇK = 2,00 olan dağılımlar için sırasıyla %84 ve %100'dür. ÇK = 2,00 olan dağılım için tüm replikasyonlarda hata bulunmaktadır.

## BÖLÜM V

### SONUÇ ve TARTIŞMA

Bu çalışmada Samejima'nın (1969) DTM'si ile yapılan madde parametreleri kestiriminde normallik ihlal edildiğinde RMSE ve AAD'nin, marjinal güvenirliliğin, EM cycle yakınsaklık kriterini sağlamama yüzdesinin ve madde parametresi kestiriminde programın verdiği hata sayısının nasıl etkilendiği incelenmiştir ve buna ilişkin sonuçlar aşağıda verilerek diğer çalışmaların sonuçlarıyla tartışılmıştır.

#### 1. Alt Probleme İlişkin Sonuçlar

Bu alt probleme ilişkin sonuçlar 1.1. alt probleme ve 1.2. alt probleme ilişkin sonuçlar olmak üzere 2 ayrı başlık altında ele alınmıştır.

##### 1.1. Alt Probleme İlişkin Sonuçlar

Analizler göstermiştir ki çarpıklık katsayısının artması  $a$  parametresi için RMSE'nin de artmasına neden olmuştur ve normalliğin ihlal edilmediği durumlarda birbirine yakın RMSE değerleri elde edilirken, normalliğin ihlal edildiği durumlarda RMSE değerleri oldukça yüksektir. Normallik ihlali arttıkça RMSE değerlerinin artması ölçme kesinliğinin azaldığını ve kestirilen parametrelerin gerçek parametrelerden uzaklaştığını göstermektedir. Sonuç olarak dağılımların çarpıklık katsayısının artması Samejima'nın DTM'sinde yapılan kestirimlerin uygunluk düzeyinin azalmasına yol açmaktadır. Bu araştırmanın sonuçlarına benzer olarak  $a$  parametresi için De Mars (2002) 500 kişilik örneklem büyüklüğünde normal dağılıma ait RMSE'nin çarpık dağılıma ( $\alpha = 2, \beta = 5,5$ ) ait RMSE'den daha yüksek olduğunu raporlamışlardır. Ayrıca madde parametresi kestirimlerinde Lautenschlager vd. (2006) hata karelerinin ortalamasının (MSE) çarpık dağılımda ( $\text{ÇK} = 1,25$ ) çok yüksek olduğunu, Kieftenbeld ve Natesan (2012) de çarpıklığın ( $\text{ÇK} = 1,25$ ) negatif bir etkisi olduğunu, Bahry (2012) çalıştığı örneklem büyüklüklerinde

çarpıklık katsayısının artmasıyla RMSE'nin yükseldiğinin ve en zayıf kestirimlerin en çarpık dağılımda ( $\text{ÇK} = 1,00$ ) olduğunu belirtmişlerdir. Doğan ve Tezbaşaran (2003)'ün çalışmasında ise 2PLM ve 3PLM'de a parametrelerinin evren a parametreleriyle korelasyonunun en düşük sağa çarpık dağılımda olduğu ve a parametrelerinin dağılımın çarpıklığından daha fazla etkilendiği görülmüştür.

## 1.2. Alt Probleme İlişkin Sonuçlar

DTM b parametreleri için çarpıklık katsayısının artması RMSE'nin artmasına neden olmuştur ve normalliğin ihlal edilmediği durumlarda birbirine yakın RMSE değerleri elde edilirken, normalliğin ihlal edildiği durumlarda (özellikle  $\text{ÇK} = 2,00$  olan dağılımda) RMSE değerleri oldukça yüksektir. Normallik sağlanmadığında RMSE değerlerinin artması ölçme kesinliğinin azaldığını ve kestirilen parametrelerin gerçek parametrelerden uzaklaştığını göstermektedir. Sonuç olarak dağılımların çarpıklık katsayısının artması Samejima'nın DTM'sinde yapılan kestirimlerin uygun olmamasına yol açmaktadır. Bu araştırmanın sonuçlarına benzer olarak b parametreleri için De Mars (2002) ele aldığı iki örneklem büyüklüğünde de normal dağılıma ait RMSE'nin çarpık dağılıma ( $\alpha = 2, \beta = 5,5$ ) ait RMSE'den daha yüksek olduğunu raporlamışlardır. Madde parametresi kestiriminde; Lautenschlager vd. (2006) hata karelerinin ortalamasının (MSE) çarpık dağılımda ( $\text{ÇK} = 1,25$ ) çok yüksek olduğunu, Kieftenbeld ve Nattesan (2012) çarpıklığın ( $\text{ÇK} = 1,25$ ) negatif bir etkisi olduğunu, Bahry (2012) çalıştığı örneklem büyüklüklerinde çarpıklık katsayısının artmasıyla RMSE'nin yükseldiğinin ve en zayıf kestirimlerin en çarpık dağılımda ( $\text{ÇK} = 1,00$ ) olduğunu belirtmişlerdir. Doğan ve Tezbaşaran (2003)'ün çalışmasında ise 1PLM, 2PLM ve 3PLM'de b parametrelerinin evren b parametreleriyle korelasyonunun en düşük olduğu dağılımların sırasıyla sağa çarpık, sola çarpık ve basık dağılımda olduğu ve en yüksek korelasyonların da evrenden seçilen randum dağılımda olduğu görülmüştür. Gelbal (1994) bu çalışmanın aksine Türkçe ve Matematik testi için hem normal hem de normal olmayan dağılımda  $b_g$  ve  $z_g$  arasındaki korelasyonu 1,00 bulmuş, standart hatayı ise normal dağılımda normal olmayan dağılımdan daha yüksek bulmuştur. Ancak normal dağılımda standart hatanın normal olmayan dağılımda daha fazla olmasının nedenini örneklem büyüklüğünün normal olmayan dağılımda daha fazla olmasıyla açıklamıştır.



## 2. Alt Probleme İlişkin Sonuçlar

Bu alt probleme ilişkin sonuçlar 2.1. alt probleme ve 2.2. alt probleme ilişkin sonuçlar olmak üzere 2 ayrı başlık altında ele alınmıştır.

### 2.1. Alt Probleme İlişkin Sonuçlar

DTM a parametresi için çarpıklık katsayısının artması AAD'nin artmasına neden olmuştur ve özellikle  $N(0,1)$  simetrik dağılımda olmak üzere normalliğin ihlal edilmediği durumlarda AAD değerleri düşükken, normalliğin ihlal edildiği durumlarda AAD değerleri sayıltının ihlal edilmediği dağılımlara göre daha yüksektir. Ayrıca normalliğin ihlal edildiği durumlarda a parametresi için AAD değerlerinin de birbirine yakın olduğu sonucuna varılmıştır. AAD değerlerinin artması ölçme kesinliğinin azaldığını ve kestirilen parametrelerin gerçek parametrelerden uzaklaştığını göstermektedir. Sonuç olarak dağılımların çarpıklık katsayısının artması Samejima'nın DTM'sinde yapılan kestirimlerin doğruluk düzeyinin azalmasına yol açmaktadır.

### 2.2. Alt Probleme İlişkin Sonuçlar

DTM b parametreleri için çarpıklık katsayısının artması AAD'nin artmasına neden olmuştur ve özellikle  $N(0,1)$  simetrik dağılımda olmak üzere normalliğin ihlal edilmediği durumlarda AAD değerleri düşükken, normalliğin ihlal edildiği durumlarda AAD değerleri çok (özellikle  $\text{ÇK} = 2,00$  olan dağılımda) yüksektir. AAD değerlerinin artması ölçme kesinliğinin azaldığını ve kestirilen parametrelerin gerçek parametrelerden uzaklaştığını göstermektedir. Sonuç olarak dağılımların çarpıklık katsayısının artması Samejima'nın DTM'sinde yapılan kestirimlerin doğruluk düzeyinin azalmasına yol açmaktadır.

## 3. Alt Probleme İlişkin Sonuçlar

Çarpıklık katsayısının artması marjinal güvenilirlik katsayısının azalmasına yol açmaktadır.  $N(0,1)$  simetrik dağılımında marjinal güvenilirlik oldukça yüksektir ve normallik sayıltısının ihlal edilmediği diğer dağılımlarda marjinal güvenilirlik katsayısı kabul edilebilir düzeydedir. Buna karşın normalliğin ihlal edildiği durumlarda marjinal güvenilirlik katsayıları daha düşüktür. Marjinal güvenilirlik katsayısı normallik ihlal edildiğinde azalarak ölçümlerin güvenilirliğini önemli ölçüde düşürmektedir.

#### 4. Alt Probleme İlişkin Sonuçlar

Dağılımların normallikten sapması kestirim yapılan programda EM cycle yakınsaklık kriterini sağlamama yüzdesini arttırmaktadır. Normalliğin ihlal edilmediği dağılımlarda kriterin sağlanmadığı durum bulunmazken, ihlal edildiği dağılımlarda kriter çoğunlukla sağlanamamaktadır ve cycle tamamlanamayarak kestirimlerin tam olarak yapılamamasına neden olmaktadır. Ayrıca normalliğin sağlanmadığı durumlarda çarpıklık katsayısı arttıkça kriterin sağlanmama yüzdesinin arttığı sonucuna varılmıştır. Bu sonuca benzer olarak Bahry (2012) de yaptığı çalışmada 3000 kişilik örneklem için çarpıklık katsayısı 0,00 ve 0,50 olduğunda kriterin her replikasyonda sağlandığını, ancak çarpıklık katsayısı 1,00 olduğunda replikasyonların bir kısmının kriteri sağlamadığı sonucuna ulaşmıştır.

#### 5. Alt Probleme İlişkin Sonuçlar

Dağılımların normallikten sapması kestirim yapılan programda hata verme yüzdesini arttırmaktadır. Normalliğin ihlal edilmediği dağılımlarda hata veren bir replikasyon söz konusu değilken, ihlal edildiği dağılımlarda çoğu replikasyonda hata bulunmaktadır ve bu hata madde parametresi kestirimlerinin yapılamadığının bir göstergesidir ve hata sayısı kadar maddede b parametreleri olması gerektiği gibi  $b_1 < b_2 < b_3$  şeklinde kestirilememektedir. Ayrıca normalliğin sağlanmadığı durumlarda çarpıklık katsayısı arttıkça hata veren replikasyon sayısının, replikasyonlardaki hata sayısının ve tüm replikasyonlara ait ortalama hata sayısının arttığı sonucuna varılmıştır.

Bulgularında kanıtladığı gibi normallik önemli ölçüde sağlanmadığında ( $\text{ÇK} > 1,00$ ) Samejima'nın Derecelendirilmiş Tepki Modelinde

- 1- Modeldeki a parametrelerinin kestirimindeki hata önemli ölçüde artmaktadır.
- 2- Modeldeki b parametrelerinin kestirimindeki hata önemli ölçüde artmaktadır.
- 3- Marjinal güvenilirlik katsayısı azalmakta ve ölçümlerin güvenilirliği düşmektedir.
- 4- Madde parametrelerinin kestirimde kullanılan MULTILOG 7.03 programında EM cycle yakınsaklık kriterini sağlamama yüzdesi artmaktadır.
- 5- MULTILOG 7.03 ile yapılan madde parametresi kestirimlerinde hata veren replikasyon sayısı ve her bir replikasyonun verdiği ortalama hata sayısı artmaktadır.

Bu sonuçlardan anlaşılacağı üzere Samejima'nın Derecelendirilmiş Tepki Modelinde her ne kadar lojistik model kullanılsa da puan dağılımının normalliği mutlaka sayıltı olarak

kontrol edilmelidir. Özellikle normalliğin önemli ölçüde ihlal edildiği dağılımlarda lojistik fonksiyona dayalı da olsa Samejima'nın DTM'sinin kullanımı uygun olmamaktadır.

Bir sonraki bölümde bu başlık altında tartışılan sonuçlara dayanarak araştırmacılara yönelik öneriler ve yeni araştırma önerileri verilmiştir.





## BÖLÜM VI

### ÖNERİLER

#### Uygulayıcılar İçin Öneriler

Araştırmanın tüm alt problemlerinin sonuçları doğrultusunda MULTİLOG 7.03 programında derecelendirilmiş tepki modeline göre marjinal en çok olabilirlik kestirim yöntemi ile madde parametresi kestirimi yapacak araştırmacıların normallik sayılığını kontrol etmeleri gerekmektedir. Normallik sayılığının ihlal edildiği durumlarda bu programda DTM'ye göre MML kestirim yöntemi ile yapılacak  $a_i$  ve  $b_{ik}$  parametreleri kestirimlerinde (bu araştırma için  $\text{ÇK} = 1,5$  olan ve  $\text{ÇK} = 2,00$  olan dağılımlar) kestirilen parametre değerleri gerçek parametre değerlerinden oldukça uzaklaşmaktadır ve araştırmacıların maddeleri yanlış yorumlamasına neden olmaktadır. Bu nedenle araştırmacıların normalliğın sağlanmadığı durumlarda bu modeli kullanmamaları önerilmektedir. Normallik sağlandığında ise (bu araştırma için  $\text{ÇK} = 0,00$  olan,  $\text{ÇK} = 0,50$  ve  $\text{ÇK} = 1,00$  olan dağılımlar) kestirilen parametre değerleri ile gerçek parametre değerleri arasındaki fark kabul edilebilir derecededir ve sayılı sağlandığında araştırmacılar çok kategorili maddeler içeren bir ölçme aracı kullandıklarında bu modele dayanarak madde parametresi kestirimi yapabilirler.

#### Araştırmacılar İçin Öneriler

Bu araştırmada normallik sayılığının ihlal edilmesinin MULTİLOG 7.03 programında Samejima'nın derecelendirilmiş tepki modeline göre marjinal en çok olabilirlik kestirim yöntemi ile madde parametresi kestiriminde ölçme kesinliğini nasıl etkilediği RMSE, AAD istatistikleri ile incelenmiştir. Bu doğrultuda  $\text{ÇK} = 0,00$ ,  $\text{ÇK} = 0,50$ ,  $\text{ÇK} = 1,00$ ,  $\text{ÇK} = 1,50$  ve  $\text{ÇK} = 2,00$  olan dağılımlar üretilmiştir.

1. Bu araştırmada normallik sayılığının ihlali incelenirken benzer bir araştırma tek boyutluluk gibi sayılı için de yapılabilir.

2. Bu arařtırmada ölçme kesinliđini dađılım türü faktörünün nasıl etkilediđi incelenirken benzer bir çalıřmada test uzunluđu, örneklem büyüklüđu, kategori sayısı ve kestirim yöntemi vb. bir faktörün nasıl etkilediđini arařtırılabilir.
3. Bu arařtırmada dađılımların çarpıklık katsayıları 0,00; 0,50; 1,00; 1,50 ve 2,00 olarak belirlenmiřken bir bařka arařtırmada farklı çarpıklık veya basıklık katsayıları kullanılarak arařtırma tekrarlanabilir.
4. Arařtırmada yapılan kestirimler DTM'de MML kestirim yöntemine göre yapılmıřtır. Bařka arařtırmacılar bu model ve kestirim yöntemi yerine bařka modeller ve kestirim yöntemlerini (STM, KPM, GKPM, Rasch, 1PLM, 2PLM, 3PLM vb.; EAP, ortak en çok olabilirlik (joint maximum likelihood vb.) çalıřabilirler.
5. Bu arařtırmada kullanılan MULTILOG 7.03 programı yerine temel aldıkları modele ve kestirim yaklařımına göre parametre kestirimi yapan diđer programları seçebilirler ya da birden fazla kestirim programı kullanarak programları ölçme kesinliđi açısından karşılařtırabilirler. Örneđin, PARSCALE, R Studio, BILOG-MG, LOGIST, WinStep, XCALIBRE, IRTPRO, Mplus gibi.
6. Bu arařtırmada normalliđin ihlalinin madde parametreleri kestirimlerinin dođruluđunu nasıl etkilediđi incelenirken diđer arařtırmacılar normalliđin ihlalinin birey parametrelerinin dođruluđunu nasıl etkilediđini inceleyebilirler.
7. Normallik sayılıtsı ihlal edildiđinde Samejima'nın DTM'sinin kullanılabilmesi için farklı dađılım normalleřtirme tekniklerinin parametre kestirim dođruluđuna etkisini arařtırabilirler.

## KAYNAKLAR

- Ankenmann, R. D. & Stone, C. A. (1992, Nisan). *A Monte Carlo study of marginal maximum likelihood parameter estimates for the graded model*. Ulusal Eğitimde Ölçme Kurulu'nda (NCME) sunulmuş bildiri, San Francisco, CA.
- Arkar, H.&Şafak C. (2004). Klinik bir örnekleme Beck Depresyon Envanteri'nin boyutlarının araştırılması. *Türk Psikoloji Dergisi*, 19(53), 117 – 123.
- Avşar, F. (2007). *Doğrulayıcı faktör analizi ve beck depresyon envanteri üzerine bir uygulama*. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Bahry, L. M. (2012). *Polytomous item response theory parameter recovery: An investigation of non-normal distributions and small sample size*. Yüksek Lisans Tezi, University of Alberta Department of Educational Psychology, Edmonton.
- Baker, F. B. (2001). *The basics of item response theory* (2. baskı). USA: Eric Clearinghouse on Assessment Evaluation.
- Baker, F. B. & Kim, S. (2004). *Item response theory parameter estimation techniques* (2. baskı). New York: Marcel Dekker.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması*(2. baskı). Ankara: Pegem Akademi.
- Beck, A. T., Rush, A. J., Shaw, B. F. & Emery, G.(1979). *Cognitive therapy of depression*. New York: Guilford.

- Beck, A. T., Steer, R. A., Ball, R. & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories –IA and –II in psychiatric outpatients. *Journal of Personality Assessment*, 67(3), 588-597.
- Beck, A., Steer, R., & Brown, G. (1996). *Beck depression inventory manual* (2. baskı). San Antonio, TX: The Psychological Association.
- Bilican Demir, S. (2014). *Çok kategorili puanlanan maddelerde madde işlev farklılığının mantel test ve olabilirlik oran testi ile karşılaştırılması*. Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Boughton, K. A., Klinger, D. A. & Gierl, M. J. (2001, Nisan). *Effects of random rater error on parameter recovery of the generalized partial credit model and graded response model*. Ulusal Eğitimde Ölçme Kurulu'nda (NCME) sunulmuş bildiri, Seattle, WA.
- Büyüköztürk, Ş., Çokluk, Ö. & Köklü, N. (2012). *Sosyal bilimler için istatistik* (11. Baskı). Ankara: Pegem Akademi.
- Büyüköztürk, Ş. (2012). *Sosyal bilimler için veri analizi el kitabı* (16. Baskı). Ankara: Pegem Akademi.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont CA: Wadsworth group/Thomson learning.
- De Ayala, R. J. (2009). *Theory and practice of item response theory*. USA: Guilford.
- De Ayala, R. J., Dodd, B. G. & Koch, W. R. (1991, Nisan). *An investigation of the robustness of a partial credit model-based computerized adaptive test to misfitting items*. Ulusal Eğitimde Ölçme Kurulu'nda (NCME) sunulmuş bildiri, Chicago.
- De Ayala, R. J. & Sava-Bolesta, M. (1999). Item parameter recovery for the nominal response model. *Applied Psychological Measurement*, 23(1), 3-19.
- De Mars, C. (2002, Nisan). *Recovery of graded response and partial credit parameters in MULTILOG and PARSCALE*. Amerikan Eğitim Araştırmaları Derneği'nde (AERA) sunulmuş bildiri, Chicago.
- De Mars, C. (2003). Sample size and the recovery of nominal response model item parameters. *Applied Psychological Measurement*, 27(4), 275-288.
- DeMars, C. (2010). *Item response theory*. New York: Oxford University.



- Doğan, N. & Tezbaşaran, A. A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklemeler bağlamında karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 25, 58-67.
- Dolma, S. (2009). *Çok ihtimalli rasch modeli ile derecelendirilmiş yanıt modelinin örtük özellikleri tahminleme performansı açısından simülasyon yöntemiyle karşılaştırılması*. Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Du Toit, M. (Ed.). (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. USA: Scientific Software International.
- Embretson, S. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme-I temel kavramlar ve işlemler*(1. Baskı). Ankara: Pegem Akademi.
- Erkuş, A. (2003). *Psikometri üzerine yazılar*(1.baskı). Ankara: Türk Psikologlar Derneği.
- Field, A. (2009). *Discovering statistics using SPSS* (3. baskı). London: Sage.
- Gelbal, S. (1994). *P madde güçlük indeksi ile Rasch modelinin b parametresi ve bunlara dayalı yetenek ölçüleri üzerine bir karşılaştırma*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Hambleton, R. K.&Swaminathan, H. (1985). *Item response theory principles and applications* (2. baskı). USA: Kluwer-Nijhoff.
- Hambleton, R.K.,Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Han, K. T. (2007). WinGen: Windows software that generates IRT parameters and item responses. *Applied Psychological Measurement*, 31(5), 457-459.
- Han, K. T., & Hambleton, R. K. (2007). *User's Manual: WinGen (Center for Educational Assessment Report No. 642)*. Amherst, MA: University of Massachusetts, School of Education.
- Hisli N. (1988). Beck depresyon envanterinin geçerliği üzerine bir çalışma. *Psikoloji Dergisi*,6(22), 118-122.

- Hisli, N. (1989). Beck depresyon ölçeğinin üniversite öğrencileri için geçerliği, güvenilirliği. *Psikoloji Dergisi*, 7(23), 3-13.
- Karasar, N. (2013). *Bilimsel araştırma yöntemleri* (25. baskı). Ankara: Nobel.
- Kapçı, E. G., Uslu, R., Türkçapar, H. & Karaoğlan, A. (2008). Beck Depression Inventory II: Evaluation of the psychometric properties and cut-off points in a Turkish adult population. *Depression and Anxiety*, 25, E104-E110.
- Kieftenbeld, V. & Natesan, P. (2012). Recovery of graded response model parameters: a comparison of marginal maximum likelihood and markov chain monte carlo estimation. *Applied Psychological Measurement*, 36(5), 399–419.
- Kline, T. J. B. (2005). *Psychological testing: a practical approach to design and evaluation*. USA: Sage.
- Kömürcü, A. (1992). *Bilim ve İnsan* (2. baskı). Ankara: Biltav.
- Lautenschlager, G. J., Meade, A. W., Kim, S. (2006, Nisan). *Cautions regarding sample characteristics when using the graded response model*. 21. Endüstri ve Örgüt Psikolojisi Derneği Yıllık Konferansı'nda sunulmuş bildiri, Dallas.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. NJ: Lawrence Erlbaum Associates.
- Mergen, H., Bernstein, I. H., Tavlı, V., Öngel, K., Tavlı, T. & Tan, Ş. (2011). Comparative validity and reliability study of the QIDS-SR<sub>16</sub> in Turkish and American college student samples. *Klinik Psikofarmoloji Bülteni*, 21(4), 289 - 301.
- Nartgün, Z. (2002). *Aynı tutumu ölçmeye yönelik likert tipi ölçek ile metrik ölçeğin madde ve ölçek özelliklerinin klasik test kuramı ve örtük özellikler kuramına göre incelenmesi* (Doktora Tezi). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Ostini, R. & Nering, M. L. (2006). *Polytomous item response theory models*. USA: Sage.
- Ostini, R. & Nering, M. L. (2010). *Handbook of polytomous item response theory models*. New York: Taylor & Francis.
- Özdemir, D. (2004). Çoktan seçmeli testlerin klasik test teorisi ve örtük özellikler teorisine göre hesaplanan psikometrik özelliklerinin iki kategorili ve

- ağırlıklandırılmış puanlanması yönünden karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 26, 117-123.
- Ree, M. J. (1979). Estimating item characteristic curves. *Applied psychological Measurement*, 3(3), 371-385.
- Reise, S. P. & Yu, J. (1990). Parameter recover in the graded response model using MULTILOG. *Journal of Educational Measurement*, 27(2), 133-144.
- Roberts, J. S., Donue, J. R. & Laughlin, J. E. (1999, Nisan). *Estimating parameters in the generalized graded unfolding model: sensitivity to the prior distribution assumption and the number of quadrature points used*. Ulusal Eğitimde Ölçme Kurulu'nda (NCME) sunulmuş bildiri, Montreal, PQ, Canada.
- Roberts, J. S., Donue, J. R. & Laughlin, J. E. (2002). Characteristics of MML/EAP parameter estimates in the generalized graded unfolding model. *Applied Psychological Measurement*, 26(2), 192-207.
- Russell, B. (2008). *Din ile Bilim* (3. baskı) (Akşit Göktürk, Çev.). İstanbul: YKY.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, (Suppl. 17), 1-100.
- Samejima, F.. (1996). Graded response model. van der Linden, W. J. & Hambleton, R. K. (Ed.). *Handbook of modern item response theory* (s. 85- 100). New York: Springer.
- Seong, T. J. (1990). Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions. *Applied Psychological Measurement*, 14(3), 299-311.
- Seong, T. J., Kim, S. H. & Cohen, A. S. (1997, Mart). *A Comparison of Procedures for Ability Estimation under the Graded Response Model*. Ulusal Eğitimde Ölçme Kurulu'nda (NCME) sunulmuş bildiri, Chicago.
- Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: an evaluation of MULTILOG. *Applied Psychological Measurement*, 16(1), 1-16.
- Swaminathan, H. & Gifford, J. A. (1979, Nisan). *Estimation of parameters in the three-parameter latent trait model*. AERA-NCME sempozyumunda sunulmuş bildiri, San Francisco.

- Tavşancıl, E. (2014). *Tutumların ölçülmesi ve SPSS ile veri analizi*(5. baskı). Ankara: Nobel.
- Teğin, B. (1980). *Depresyonda bilişsel bozukluklar: Beck modeline göre bir inceleme*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Tezbaşaran, A. A. (1997). *Likert tipi ölçek geliştirme kılavuzu*. Ankara: Türk Psikologlar Derneği.
- Thissen, D., Chen, W. H.& Bock, D. (2003). *MULTILOG 7.03*. Lincolnwood, IL: Scientific Software International.
- Turgut, M. F. & Baykul, Y. (1992).*Ölçekleme teknikleri*. Ankara: ÖSYM.
- Turgut, M. F. & Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme* (1. baskı). Ankara: Pegem Akademi.
- Ünver, Ö., Gamgam, H. & Altunkaynak, B. (2013). *Temel istatistik yöntemler: hipotez testleri-ilişki katsayıları-regresyon analizi* (7.baskı). Ankara: Seçkin.
- van der Linden, W. J. & Hambleton, R. K. (1996). Item response theory: brief history, common models, and extensions. van der Linden, W. J. & Hambleton, R. K. (Ed.). *Handbook of modern item response theory* (s. 1-28). New York: Springer.
- Willmott, C. J. & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79-82.
- Zengin, F. (1999). *Psychometric investigation of two turkish adaptations and short form of the beck depression inventory*. Yüksek Lisans Tezi, Boğaziçi Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.



**EKLER**

## EK 1. Beck Depresyon Envanteri

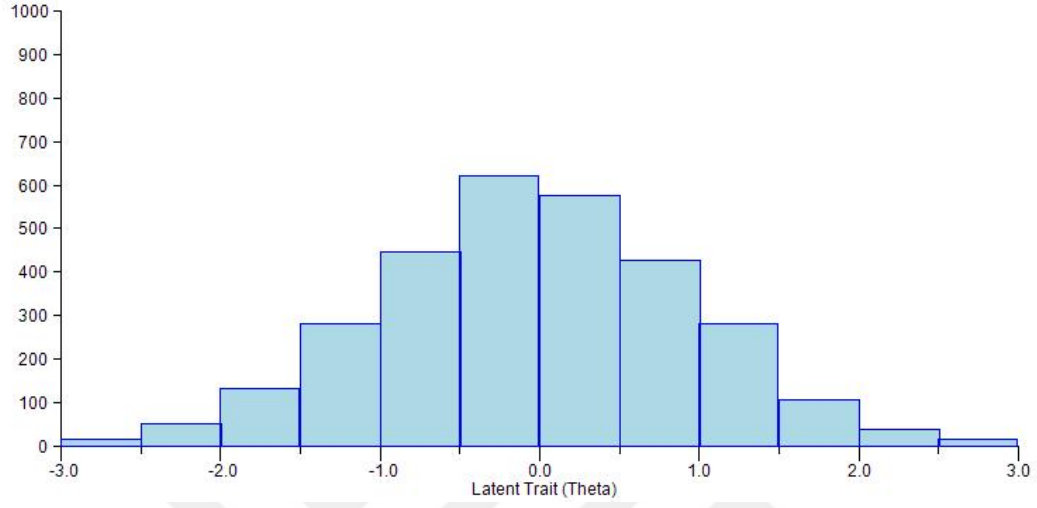
Beck Depresyon Envanteri maddelerini kendinizi son bir (1) haftada nasıl hissettiğinizi düşünerek yanıtlayınız.

1. (0) Kendimi üzüntülü ve sıkıntılı hissetmiyorum.  
(1) Kendimi üzüntülü ve sıkıntılı hissediyorum.  
(2) Hep üzüntülü ve sıkıntılıyım. Bundan kurtulamıyorum.  
(3) O kadar üzüntülü ve sıkıntılıyım ki artık dayanamıyorum.
2. (0) Gelecek hakkında umutsuz ve karamsar değilim.  
(1) Gelecek hakkında karamsarım.  
(2) Gelecekte beklediğim hiçbir şey yok.  
(3) Geleceğim hakkında umutsuzum ve sanki hiçbir şey düzelmeyecekmiş gibi geliyor.
3. (0) Kendimi başarısız bir insan olarak görmüyorum.  
(1) Başkalarından daha başarısız olduğumu hissediyorum.  
(2) Geçmişe baktığımda başarısızlıklarla dolu olduğumu görüyorum.  
(3) Kendimi tümüyle başarısız biri olarak görüyorum.
4. (0) Her şeyden eskisi kadar zevk alıyorum.  
(1) Birçok şeyden eskiden olduğu gibi zevk almıyorum.  
(2) Artık hiçbir şey bana tam anlamıyla zevk vermiyor.  
(3) Her şeyden sıkılıyorum.
5. (0) Kendimi herhangi bir şekilde suçlu hissetmiyorum.  
(1) Kendimi zaman zaman suçlu hissediyorum.  
(2) Çoğu zaman kendimi suçlu hissediyorum.  
(3) Kendimi her zaman suçlu hissediyorum.
6. (0) Kendimden memnunum.  
(1) Kendimden pek memnun değilim.  
(2) Kendime çok kızıyorum.  
(3) Kendimden nefret ediyorum.
7. (0) Başkalarından daha kötü olduğumu sanmıyorum.  
(1) Zayıf yanlarım ve hatalarım olduğumu düşünüyorum.  
(2) Hatalarımdan dolayı kendimden utanıyorum.  
(3) Her şeyi yanlış yapıyormuşum gibi geliyor ve hep kendimi kabahatli buluyorum.
8. (0) Kendimi öldürmek gibi düşüncelerim yok.  
(1) Zaman zaman kendimi öldürmeyi düşündüğüm oluyor ama yapmıyorum.  
(2) Kendimi öldürmek isterdim.  
(3) Fırsatını bulsam kendimi öldürürdüm.
9. (0) İçimden ağlamak geldiği pek olmuyor.  
(1) Zaman zaman içimden ağlamak geliyor.  
(2) Çoğu zaman ağlıyorum.  
(3) Eskiden ağlayabilirdim şimdi istesem de ağlayamıyorum.
10. (0) Şimdi her zaman olduğumdan daha canı sıkın ve sinirli değilim.  
(1) Eskisine kıyasla daha kolay canım sıkılıyor ve sinirleniyorum.  
(2) Her şey canımı sıkıyor ve kendimi hep sinirli hissediyorum.  
(3) Bir zamanlar beni sinirlendiren şeyler şimdi hiç sinirlendirmiyor.
11. (0) Başkaları ile görüşme, konuşma isteğimi kaybetmedim.  
(1) Başkaları ile eskiden daha az konuşmak, görüşmek istiyorum.  
(2) Birileriyle görüşüp konuşmak hiç içimden gelmiyor.  
(3) Hiç kimseyle konuşmak görüşmek istemiyorum.

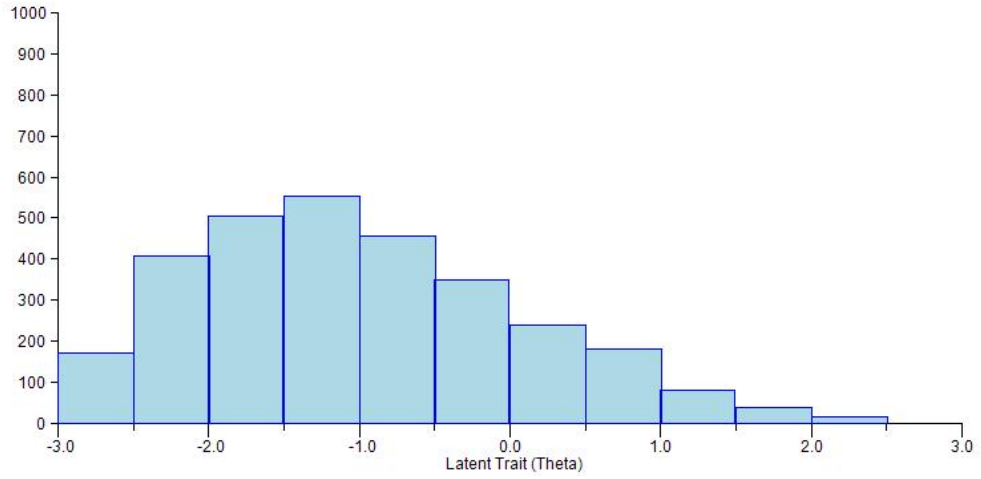
12. (0) Eskiden olduđu gibi kolay karar verebiliyorum.  
(1) Eskiden olduđu kadar kolay karar veremiyorum.  
(2) Karar verirken eskisine kıyasla çok güçlük çekiyorum.  
(3) Artık hiç karar veremiyorum.
13. (0) Aynada kendime baktığımda deđişiklik görmüyorum.  
(1) Aynada kendime her zamankinden kötü görünüyorum.  
(2) Aynaya baktığımda kendimi yaşlanmış ve çirkinleşmiş buluyorum.  
(3) Kendimi çok çirkin buluyorum.
14. (0) Eskisi kadar iyi çalışabiliyorum.  
(1) Bir şeyler yapabilmek için gayret göstermem gerekiyor.  
(2) Herhangi bir şeyi yapabilmek için kendimi çok zorlamam gerekiyor.  
(3) Hiçbir iş yapamıyorum.
15. (0) Her zamanki gibi iyi uyuyabiliyorum.  
(1) Eskiden olduđu gibi iyi uyuyamıyorum.  
(2) Her zamankinden 1-2 saat daha erken uyanıyorum ve tekrar uyuyamıyorum.  
(3) Her zamankinden çok daha erken uyanıyor ve tekrar uyuyamıyorum.
16. (0) Her zamankinden daha çabuk yorulmuyorum.  
(1) Her zamankinden daha çabuk yoruluyorum.  
(2) Yaptığım her şey beni yoruyor.  
(3) Kendimi hemen hiçbir şey yapamayacak kadar yorgun hissediyorum.
17. (0) İştahım her zamanki gibi.  
(1) Eskisinden daha iştahsızım.  
(2) İştahım çok azaldı.  
(3) Hiçbir şey yiyemiyorum.
18. (0) Son zamanlarda zayıflamadım.  
(1) Zayıflamaya çalışmadığım halde en az 2 Kg verdim.  
(2) Zayıflamaya çalışmadığım halde en az 4 Kg verdim.  
(3) Zayıflamaya çalışmadığım halde en az 6 Kg verdim.
19. (0) Sağlığım beni fazla endişelendirmiyor.  
(1) Ağrı, sancı, mide bozukluğu veya kabızlık gibi şikayetlerim oluyor ve bunlar beni endişelendiriyor.  
(2) Sağlığım beni endişelendirdiği için başka şeyleri düşünmek zorlaşıyor.  
(3) Sağlığım hakkında o kadar endişeliyim ki başka hiçbir şey düşünemiyorum.
20. (0) Son zamanlarda cinsel konulara olan ilgimde bir deđişme fark etmedim.  
(1) Cinsel konularla eskisinden daha az ilgiliyim.  
(2) Cinsel isteğim çok azaldı.  
(3) Cinsel konular olan ilgimi tamamen kaybettim.
21. (0) Cezalandırılması gereken şeyler yaptığımı sanmıyorum  
(1) Yaptıklarımın dolaylı cezalandırılabilceğimi düşünüyorum.  
(2) Cezamı çekmeyi bekliyorum.  
(3) Sanki cezamı bulmuşum gibi hissediyorum.

## EK 2. Dağılım Türlerinin Histogram Dağılım Diyagramları

ÇK = 0,00 OLAN DAĞILIM

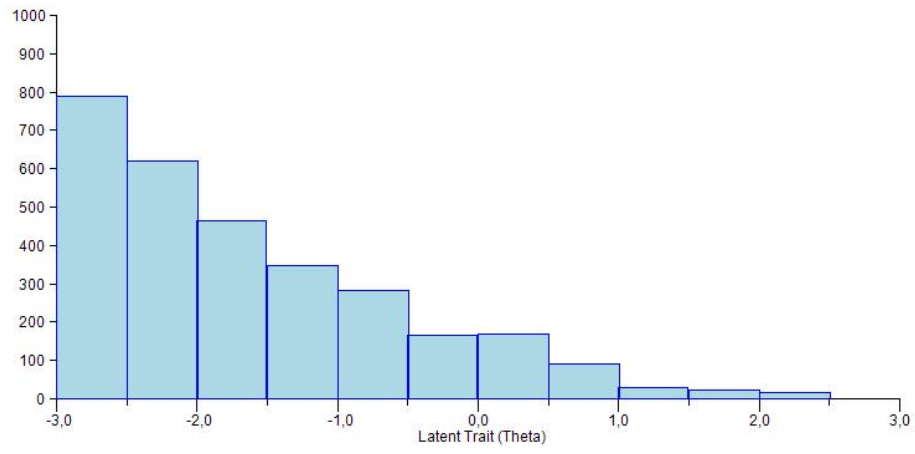


ÇK = 0,50 OLAN DAĞILIM

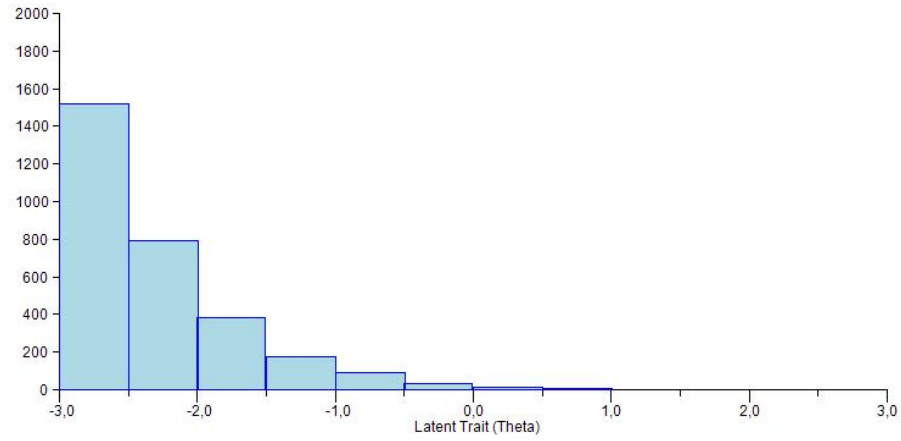




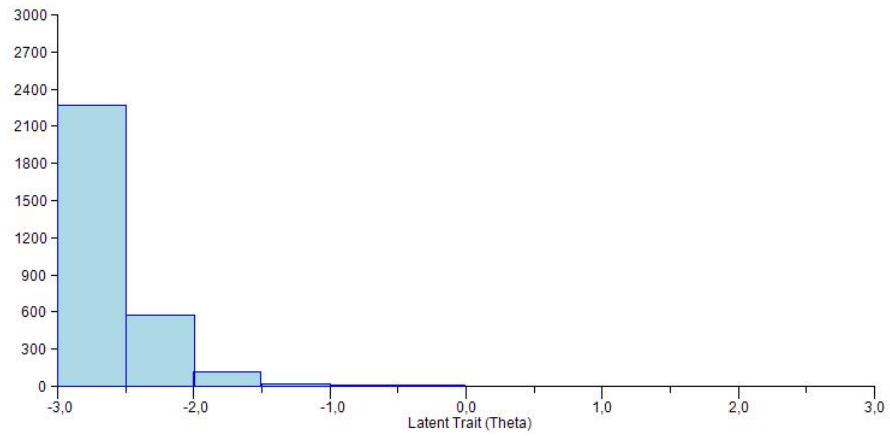
### ÇK = 1,00 OLAN DAĞILIM



### ÇK = 1,50 OLAN DAĞILIM

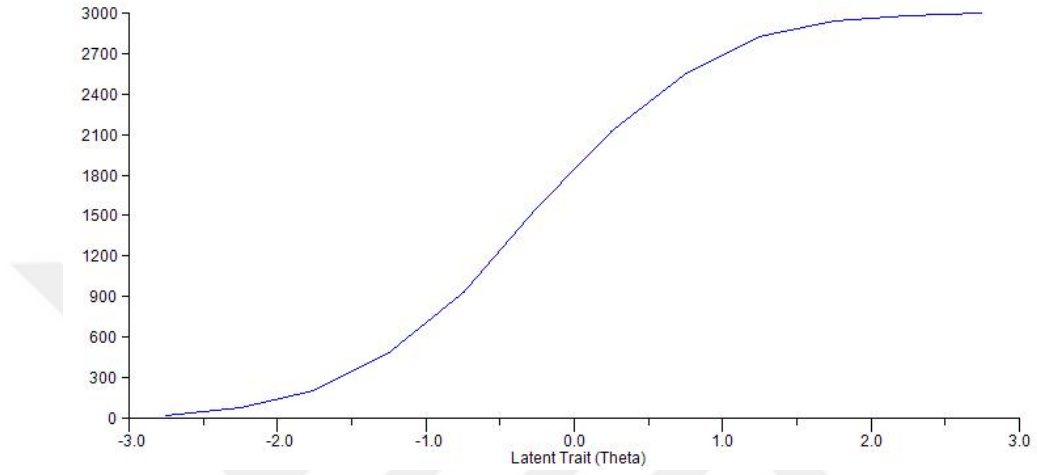


### ÇK = 2,00 OLAN DAĞILIM

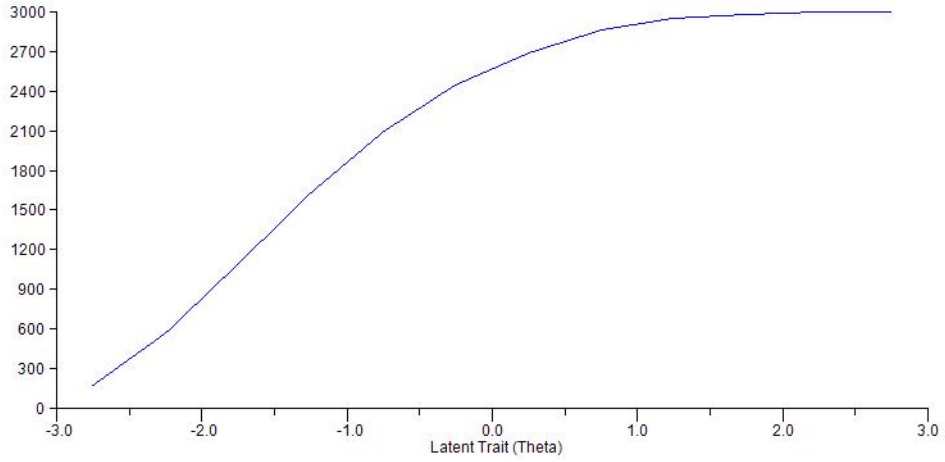


### EK 3. Dağılım Türlerinin Kümülatif Dağılım Diyagramları

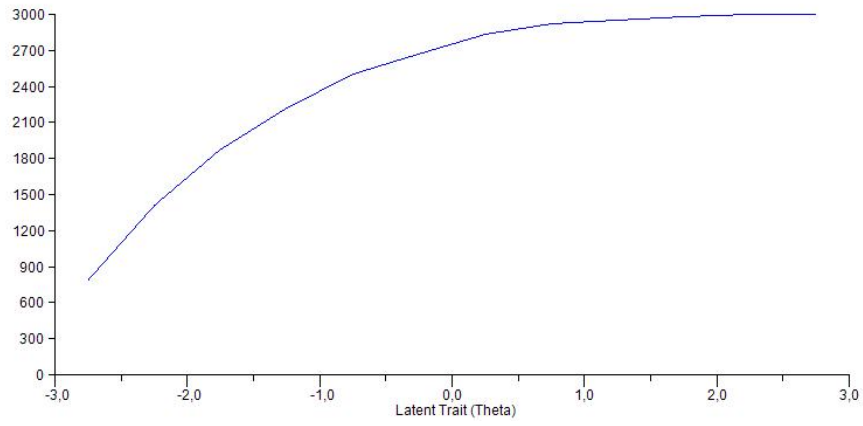
ÇK = 0,00 OLAN DAĞILIM



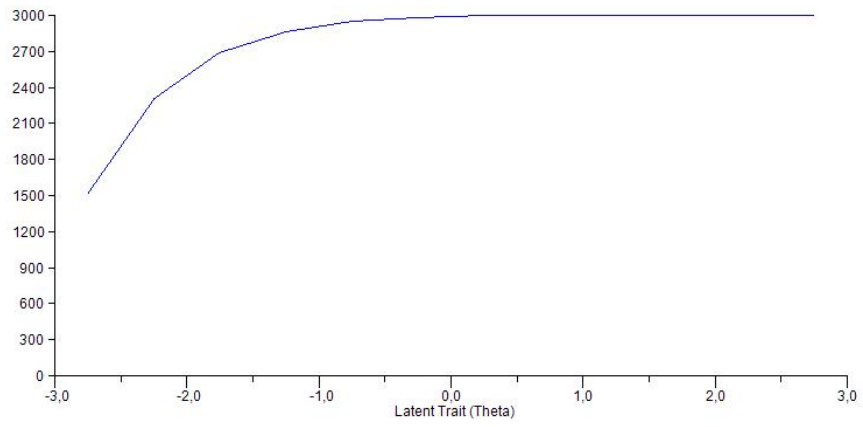
ÇK = 0,50 OLAN DAĞILIM



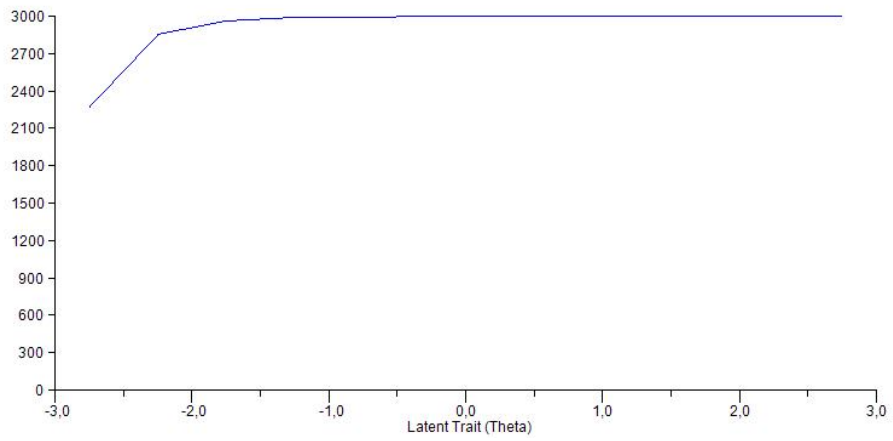
### ÇK = 1,00 OLAN DAĞILIM



### ÇK = 1,50 OLAN DAĞILIM

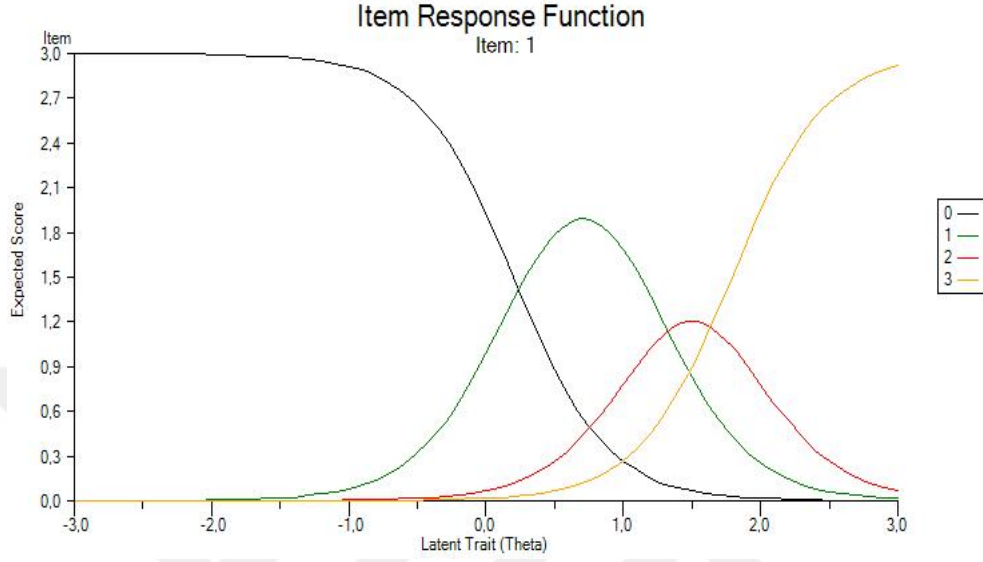


### ÇK = 2,00 OLAN DAĞILIM

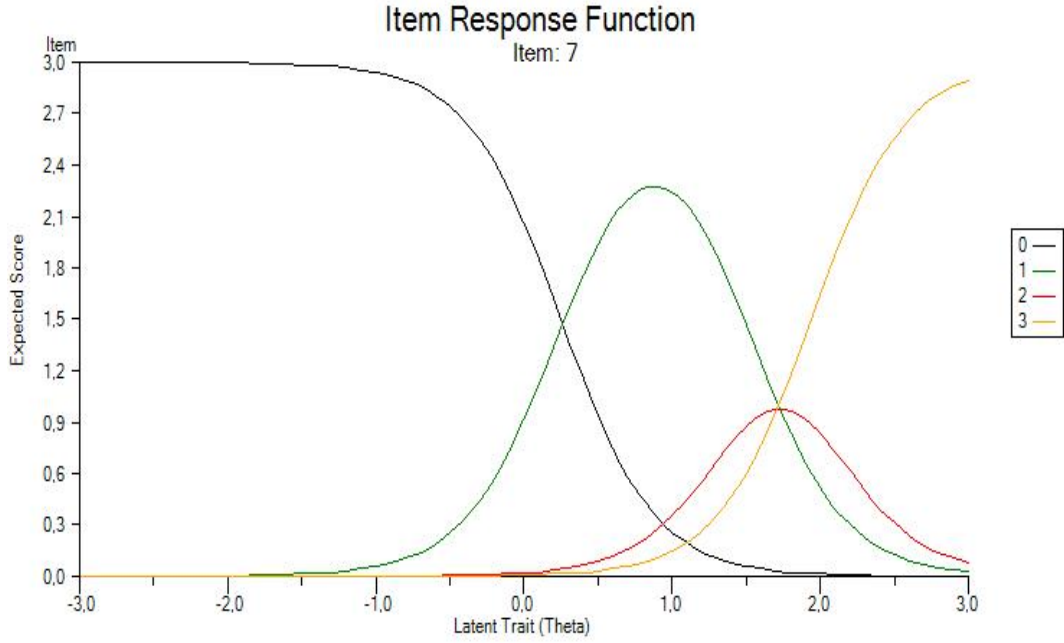


## EK 4. Bazı Maddelere Ait İşlevsel Karakteristik Eğrileri

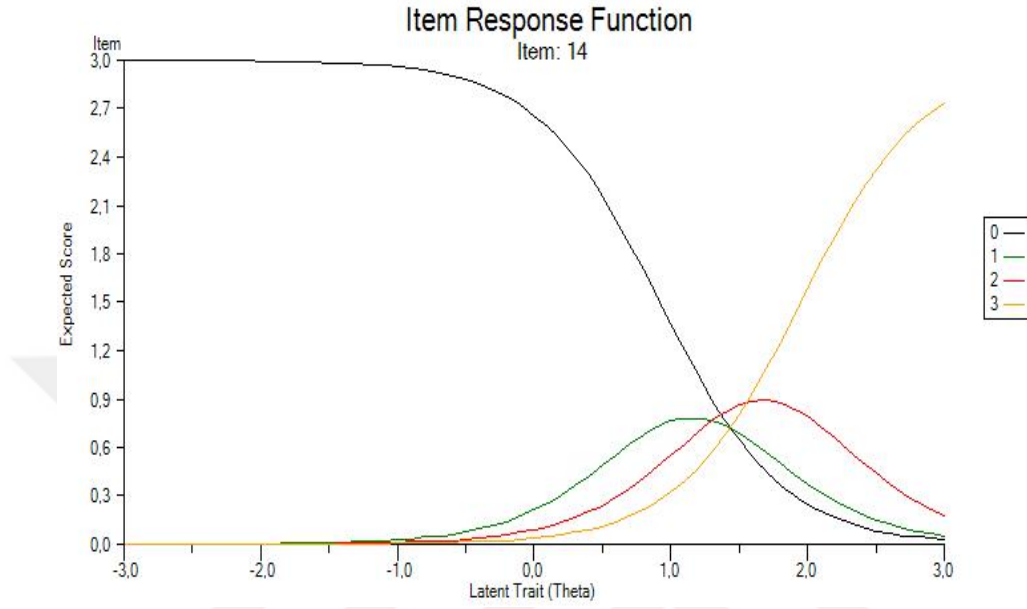
### 1. MADDEYE AİT İŞLEVSEL KARAKTERİSTİK EĞRİSİ



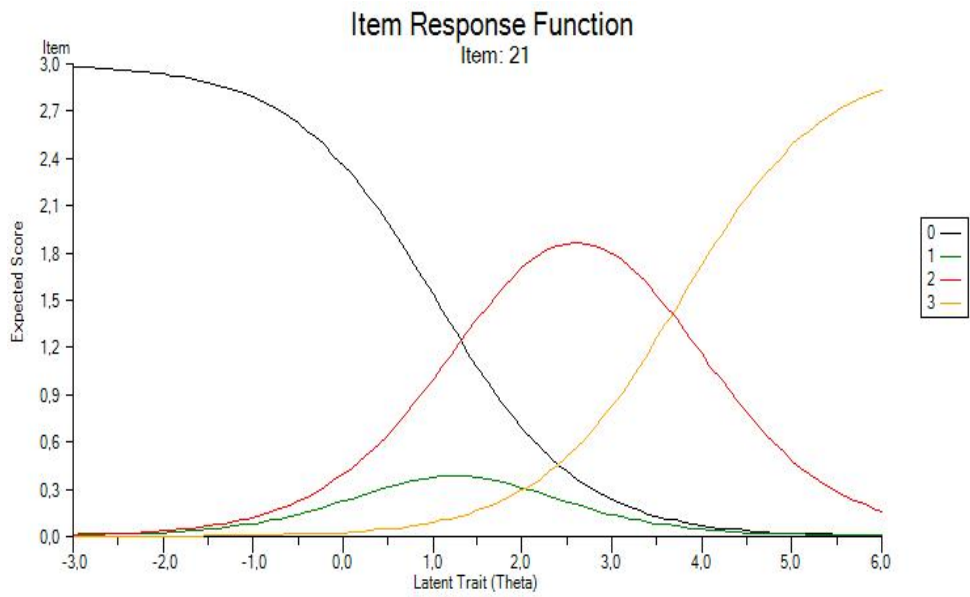
### 7. MADDEYE AİT İŞLEVSEL KARAKTERİSTİK EĞRİSİ



## 14. MADDEYE AİT İŞLEVSEL KARAKTERİSTİK EĞRİSİ

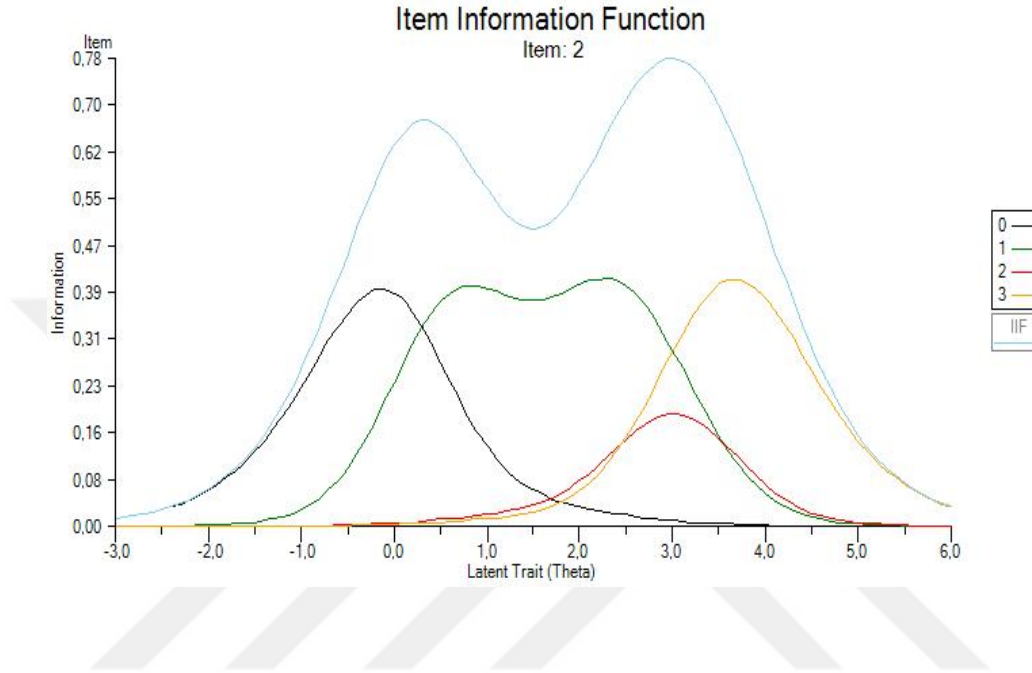


## 21. MADDEYE AİT İŞLEVSEL KARAKTERİSTİK EĞRİSİ

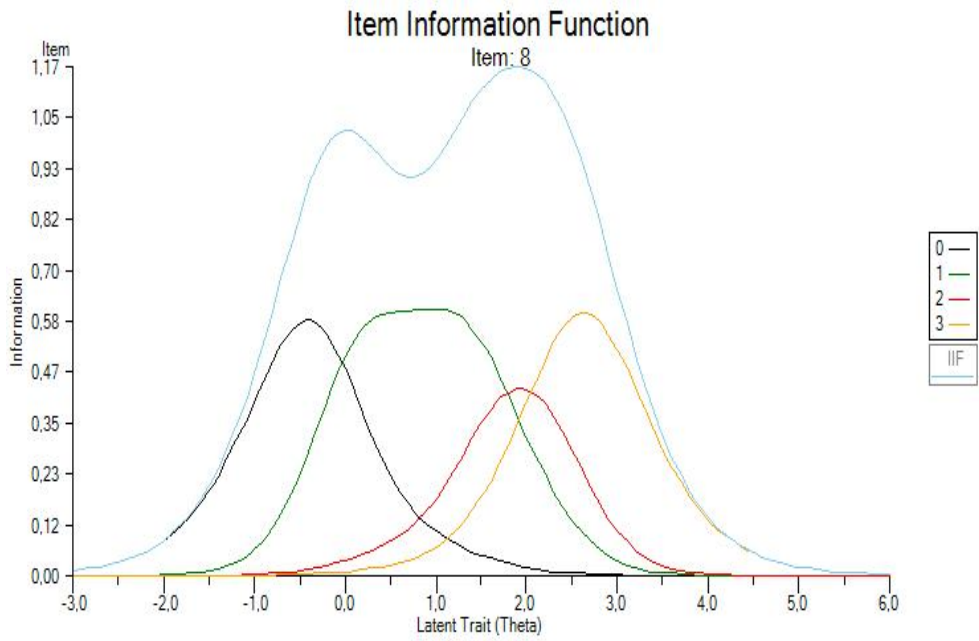


## EK 5. Bazı Maddelere Ait Madde Bilgi Fonksiyonları

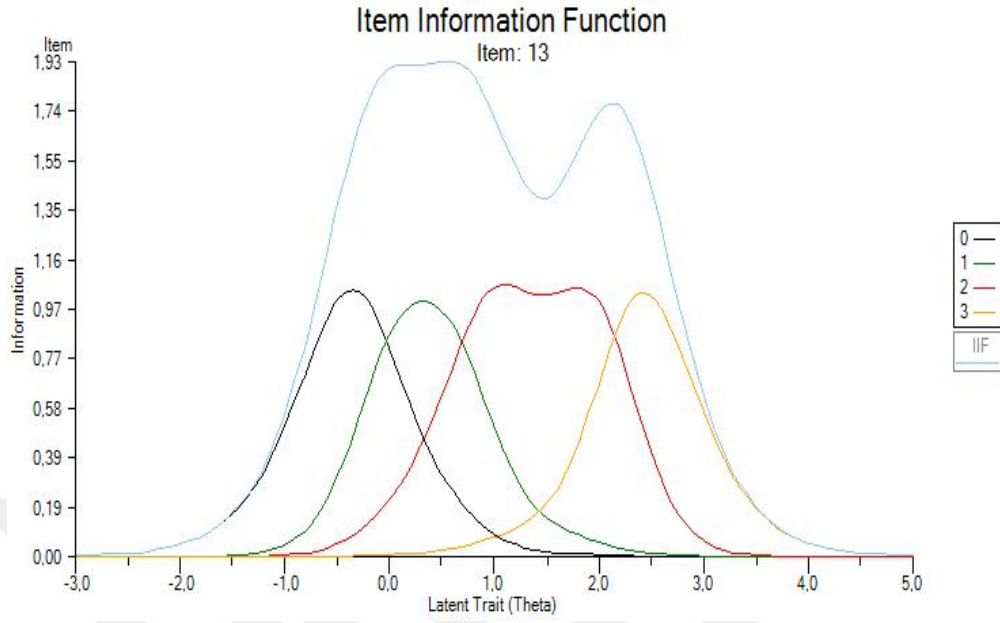
### 2. MADDEYE AİT MADDE BİLGİ FONKSİYONU



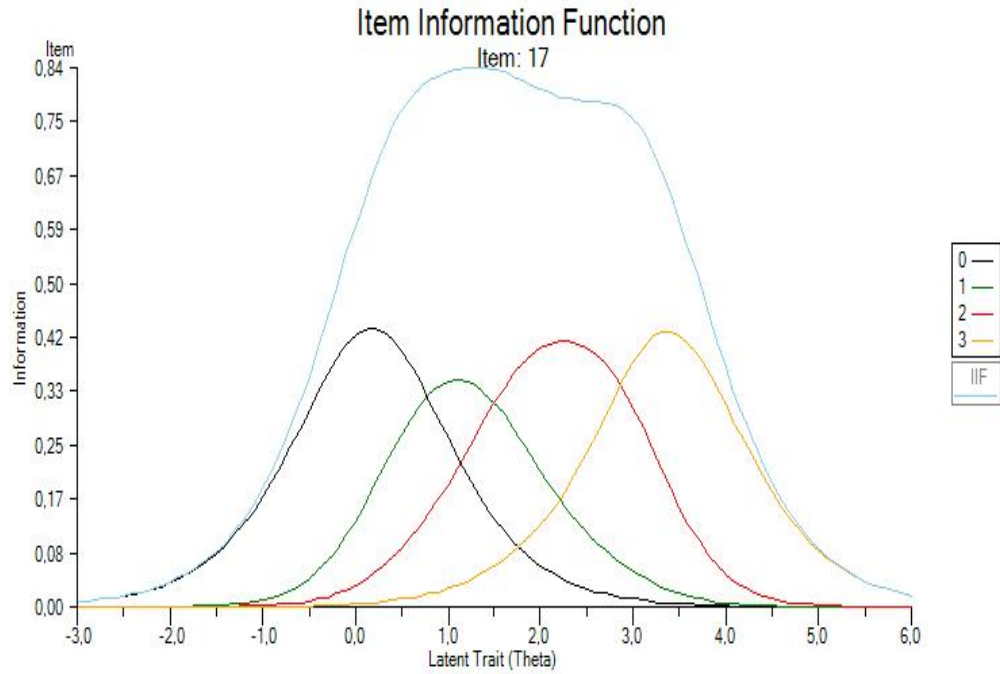
### 8. MADDEYE AİT MADDE BİLGİ FONKSİYONU



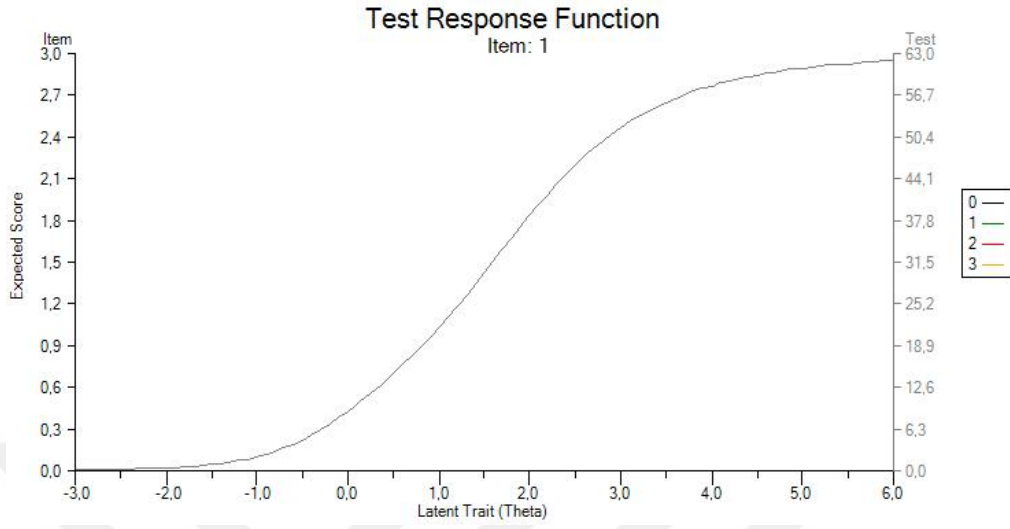
### 13. MADDEYE AİT MADDE BİLGİ FONKSİYONU



### 17. MADDEYE AİT MADDE BİLGİ FONKSİYONU

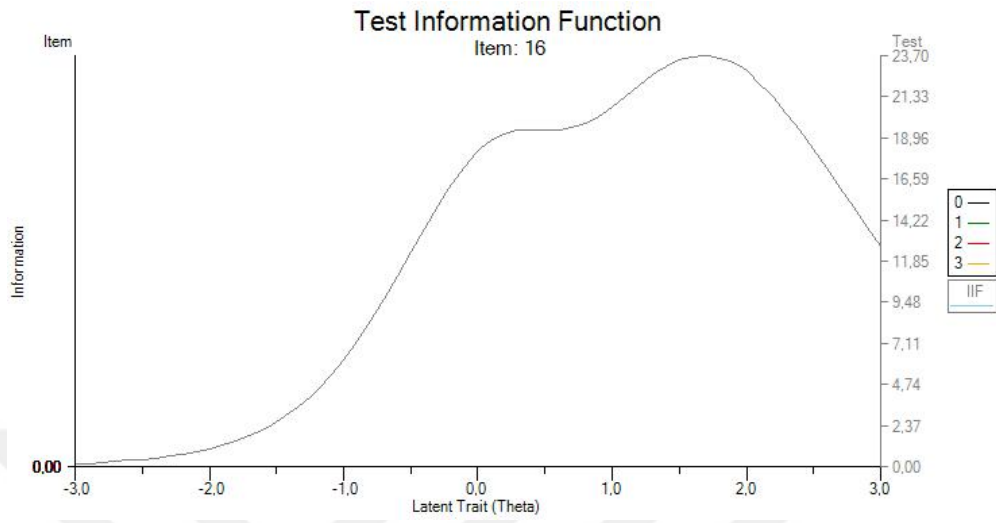


## EK 6. Test Cevap Fonksiyonu





## EK 7. Test Bilgi Fonksiyonu



## EK 8. Örnek MULTİLOG Komut Dosyası

MULTİLOG for Windows 7.00.2327.2

Created on: 24 July 2014, 09:13:23

```
>PROBLEM RANDOM,  
    INDIVIDUAL,  
    DATA = 'C:\Users\Kurtuluş\Desktop\1,0-131\1,0-131_1.DAT',  
    NITEMS = 21,  
    NGROUPS = 1,  
    NEXAMINEES = 3000,  
    NCHARS = 4;  
>TEST ALL,  
    GRADED,  
    NC = (4(0)21);  
>ESTIMATE,  
    NCYCLES = 99999,  
    ITERATIONS = 999;  
>END ;  
4  
1234  
11111111111111111111111111  
22222222222222222222222222  
33333333333333333333333333  
44444444444444444444444444  
(4A1,T5,21A1)
```