**T.C.**

**BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ**

**LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**EĞİTİM BİLİMLERİ ANABİLİM DALI**

**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

# İKİ PARAMETRELİ LOJİSTİK MODEL ALTINDA FARKLI YETENEK DAĞILIMLARININ MADDE PARAMETRE KESTİRİMİNE ETKİSİNİN İNCELENMESİ

**İSMAİL BAŞARAN**

**BOLU-2020**

**T.C.**

**BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ**

**LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**EĞİTİM BİLİMLERİ ANABİLİM DALI**

**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

# İKİ PARAMETRELİ LOJİSTİK MODEL ALTINDA FARKLI YETENEK DAĞILIMLARININ MADDE PARAMETRE KESTİRİMİNE ETKİSİNİN İNCELENMESİ

**Yüksek Lisans Tezi**

**Hazırlayan**

**İsmail BAŞARAN**

**Danışman**

**Doç. Dr. İ. Alper KÖSE**

**Bolu, Temmuz-2020**

**YÜKSEK LİSANS TEZ ONAY FORMU**

İsmail Başaran tarafından hazırlanan "İki Parametreli Lojistik Model Altında Farklı Yetenek Dağılımlarının Madde Parametre Kestirimine Etkisinin İncelenmesi" adlı çalışma Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalında Yüksek Lisans Tezi olarak kabul edilmiştir.

| **Akademik Unvan ve Adı Soyadı** | | **İmza** |
|---|---|---|
| Üye (Tez Danışmanı) | : Doç. Dr. İ. Alper KÖSE | ……………….. |
| Üye | : Doç. Dr. C. Deha DOĞAN | ……………….. |
| Üye | : Dr. Öğr. Üyesi Alperen YANDI | ……………….. |

**Lisansüstü Eğitim Enstitüsünün Onayı**

**Prof. Dr. Osman Görür**

**Lisansüstü Eğitim Enstitü Müdürü**

Yüksek Lisans Tezi olarak sunduğum “İki Parametreli Lojistik Model Altında Farklı Yetenek Dağılımlarının Madde Parametre Kestirimine Etkisinin İncelenmesi” adlı çalışmanın yazılmasında bilimsel ve etik kurallara uyduğumu, başkalarının eserlerinden yararlanılması durumunda atıfta bulunduğumu, kullanılan verilerde herhangi bir tahrifat yapmadığımı, tezin tamamının ya da bir kısmının bu üniversite ya da başka bir üniversitede bir tez çalışması olarak sunulmadığını beyan ederim. …/…/…

İsmail Başaran

# TEŞEKKÜR

Yüksek lisans eğitimim ve tezimin her aşamasında beni destekleyen, Madde Tepki Kuramı ile tanışmamı sağlayan ve bu alanda bana ışık tutan değerli hocam ve danışmanım Doç. Dr. İbrahim Alper Köse'ye,

Tez çalışmamda sorularımı sabırla cevaplayan Prof. Dr. Levent Kirişçi, Prof. Dr. Li Cai, Prof. Dr. R. Philip Chalmers, Dr. Öğr. Üyesi Murat Doğan Şahin ve Arş. Gör. Dr. Elif Kübra Demir'e,

Değerli vakitlerini ayırarak tezimi inceyelen Doç. Dr. C. Deha Doğan ve Dr. Öğr. Üyesi Alperen Yandı'ya,

Bu süreçte vaktini ayırma nezaketini göstererek yardımcı olan İngilizce Öğretmeni Elif Pehlivan'a,

Yüksek lisans eğitimimde ve tez çalışmamda desteklerini bir an bile esirgemeyen çok kıymetli aileme ve bu zamana kadar üzerimde emeği olan tüm öğretmenlerime çok teşekkür ederim.

İsmail BAŞARAN

# İÇİNDEKİLER

# TABLOLAR DİZİNİ

# ŞEKİLLER DİZİNİ

# GRAFİKLER DİZİNİ

# KISALTMALAR DİZİNİ

1PL: 1 Parametreli lojistik

2PL: 2 Parametreli lojistik

3PL: 3 Parametreli lojistik

CML: Conditional maximum likelihood

ÇK: Çarpıklık katsayısı

JML: Joint maximum likelihood

KTK: Klasik Test Kuramı

MC: Monte Carlo

MKE: Madde karakteristik eğrisi

ML: Maximum likelihood

MLE: Maximum likelihood estimation

MML: Marginal maximum likelihood

MTK: Madde Tepki Kuramı

RMSE: Root mean square error

RMSD: Root mean square difference

TBF: Test bilgi fonksiyonu

# ÖZET


## İKİ PARAMETRELİ LOJİSTİK MODEL ALTINDA FARKLI YETENEK DAĞILIMLARININ MADDE PARAMETRE KESTİRİMİNE ETKİSİNİN İNCELENMESİ



Başaran, İsmail
Yüksek Lisans Tezi
Eğitim Bilimleri Anabilim Dalı
Eğitimde Ölçme ve Değerlendirme Bilim Dalı
Tez Danışmanı: Doç. Dr. İ. Alper Köse
Temmuz, 2020, 116 sayfa



Eğitimde ve Psikolojide ikili puanlanan maddelerden oluşan testler sıklıkla kullanılmaktadır. Madde tepki kuramı altında lojistik modellerle kullanılabilen bu testlerin madde parametreleri kestirilirken kestirimlerin daha keskin olmasını sağlayan bazı özellikler vardır ancak testlerin uygulandığı gruplar bu özellikleri her zaman sağlayamayabilir.

Bu araştırmanın amacı, ikili puanlanan maddelerden oluşan bir veri setinin 2 parametreli lojistik (2PL) model ile analizinde veri setinin çeşitli özelliklerinin parametre kestirimlerinin keskinliğine olan etkilerini incelemektir. Bu araştırma, ikili puanlanan testlerden elde edilen yetenek parametrelerinin normal dağılmadığı durumların ve örneklem büyüklüğünün parametre kestirimlerinin keskinliğini nasıl etkileyeceğini açıklayacağından önemlidir.

Araştırmanın amacı doğrultusunda çarpıklık katsayıları 2,00, 1,00, 0,00, -1,00 ve -2,00 olan ve 250, 500, 1,000 ve 2,000 örneklem büyüklüklerinde veriler ve uzunluğu 30 maddeden oluşan bir test için madde parametreleri R programlama dilinde üretilmiştir. Üretilen her bir veri seti için 100 replikasyon gerçekleştirilmiş ve madde parametrelerinin kestirimleri R programlama dili kullanılarak *mirt* paketinde marginal maximum

likelihood (MML) kestirim yöntemi yardımıyla gerçekleştirilmiştir. Parametre kestirim keskinliğini değerlendirmek içinse hata kareleri ortalamasının karekökü (root mean squared error-RMSE) ve Bias istatistikleri kullanılmıştır.

Araştırmanın bulgularında çarpıklık katsayıları mutlak değerce büyüdüğünde a parametreleri için RMSE değerlerinin büyüdüğü ve Bias değerlerinin sıfırdan uzaklaştığı, b parametreleri için çarpıklık katsayıları mutlak değerce büyüdüğünde hemen hemen aynı RMSE ve Bias değerlerinin elde edildiği görülmüştür. Örneklem büyüklükleri arttığında a parametreleri için tüm dağılımlarda RMSE değerlerinin küçüldüğü ve Bias değerlerinin hemen hemen aynı değerlerde olduğu, b parametreleri içinse örneklem büyüklüğü arttıkça RMSE değerlerinin küçüldüğü ve Bias değerlerinin çarpıklık katsayısına göre bazen sıfıra yaklaştığı bazen de sıfırdan uzaklaştığı görülmüştür. Normal dağılımdan elde edilen sonuçlar diğer çarpıklık katsayılarına sahip dağılımlardan elde edilen sonuçlar ile karşılaştırıldığında en küçük RMSE ve sıfıra en yakın Bias değerlerini ürettiği görülmüştür.

**Anahtar Kelimeler:** Madde Tepki Kuramı, Parametre Kestirimi, Monte Carlo Simülasyonu, Çarpık Dağılım

# ABSTRACT

INVESTIGATION OF THE EFFECT OF DIFFERENT ABILITY DISTRIBUTIONS ON ITEM PARAMETER ESTIMATION UNDER TWO-PARAMETER LOGISTICS MODEL

Başaran, Ismail

Master Thesis

Department Of Educational Sciences

Measurement and Evaluation in Education

Thesis Advisor: Assoc. Dr. I. Alper Köse

July, 2020, 116 page

Tests consisting of dichotomously scored items are frequently used in education and psychology. These tests, which can be used with logistic models under item response theory, have some features that make the estimation more accurate when estimating item parameters but groups in which tests are applied may not always provide these features.

The aim of this study is to analyze the effects of various features of the data set on the accuracy of the parameter estimates in the analysis of a data set consisting of dichotomously scored items with a 2 parameter logistic (2 PL) model. This study is important because it will explain how the ability parameters obtained from dichotomously scored tests have not normal distribution and the sample size will affect the accuracy of parameter estimates.

For the purpose of the study, item parameters for a test with skewness coefficients 2,00, 1,00, 0,00, -1,00 and -2,00 and with sample sizes of 250, 500, 1,000 and 2,000 and a length of 30 items were produced in the R programming language. 100 replications were performed for each data set produced and the estimations of the item parameters were performed with the help of the marginal maximum likelihood (MML) estimation method in the mirt pack using the R programming language. To evaluate parameter estimation accuracy, root mean squared error (RMSE) and Bias statistics were used.

The results of the study showed that RMSE values for parameters a increased when the skewness coefficients increased by absolute value and Bias values moved away from zero, and that almost identical RMSE and Bias values were obtained when the skewness coefficients for parameters b increased by absolute value. When the sample sizes increased, it was observed that RMSE values decreases in all distributions for the a parameters and the Bias values were almost the same, for the b parameters, the RMSE values decreases as the sample size increased, and the Bias values sometimes approached to zero compared to the skewness coefficient. When the results obtained from the normal distribution are compared with the results obtained from the distributions with other skewness coefficients, it is seen that it produces the smallest RMSE and the Bias values closest to zero.

**Keywords:** Item Response Theory, Parameter Estimation, Monte Carlo Simulation, Skewed Distribution

# BÖLÜM I

## 1. Giriş

Bu bölümde problem durumuna, problem cümlesi, alt problemler, tanım, araştırmanın amacı, önemi ve sınırlılıkları bölümlerine yer verilmiştir.

### 1.1. Problem Durum

İnsanoğlu hiç şüphesiz merakları ve ihtiyaçları olan bir canlıdır ve insanı diğer canlılardan ayıran özelliklerinden biri de ihtiyaçlarından doğan bu meraklarını bilimsel yöntemlerle sınamasıdır. İhtiyaçlarından doğan en temel meraklardan biri de iki varlık arasındaki farklılıkların, benzerliklerin ya da ilişkilerin kuvvetinin ne kadar olduğunu, bunların kaynaklarının ne olduğunu bulmak istemesidir. İşte bu istek binlerce yıl önce ölçme faaliyetlerini doğurmuştur. Örneğin binlerce yıl önce Mısır'da yaşayan insanlar hasat ve taşkın zamanlarını doğru belirleyebilmek için Sirius yıldızını gözlemleyerek 365 günlük takvimi icat etmişlerdir (Placko, 2007). İnsanın bu merakı yine onu tanımaya ve eğitmeye itmiş, bu faaliyetler de ölçme ve değerlendirme çalışmalarını doğurmuştur (Turgut ve Baykul, 2015). Ancak biliyoruz ki psikolojik bir yapıyı ölçmek bir yıldızı gözlemlemekten daha zordur çünkü bu yapılar doğrudan gözlemlenemezler.

Psikolojik yapılar örneğin matematik yeteneği doğrudan gözlemlenip ölçülemediğinden bu yapıları dolaylı olarak ölçebilecek ölçme araçları tasarlanmıştır (Crocker ve Algina, 2008; Lord ve Novick, 2008). Bu araçlardan çoktan seçmeli testler eğitimde birçok özelliği ya da psikolojik yapıyı ölçmek için geliştirilen ve geniş bir kullanım alanı olan ölçme araçlarıdır (Ebel ve Frisbie, 1991). Bu araçlardan içinde doğru cevabın da yer aldığı birden fazla seçenekten seçerek cevaplanan ölçme aracı en sık kullanılandır (Marso ve Piggie, 1988). Çoktan seçmeli ölçme araçları soru kökü adı verilen ve cevaplanması istenen soruyu içeren kısımla cevabın birden fazla seçenek arasına gizlendiği, eşleştirildiği gibi çeşitlilik gösterebilen cevap kısmından oluşur. (Miller, Linn ve Gronlund, 2009; Mehrens ve Lehmann, 1991). Kısa cevaplı, doğru-yanlış ve eşleştirme gibi türlerine göre isimler alan çoktan seçmeli ölçme araçlarının yaygın kullanımı birçok öğrenme çıktısını ölçebilme, farklı içeriklerde hazırlanabilme, büyük gruplara uygulamada kolaylık sağlaması, test ve madde istatistiklerinin kolay hesaplanabilmesi gibi çeşitli özelliklerden kaynaklanmaktadır (Miller, Linn ve Gronlund, 2009; Mehrens ve Lehmann, 1991; Atılgan, Kan ve Doğan, 2009).

Alanyazında çoktan seçmeli maddelerden elde edilen verilerin hem Klasik Test Kuramı (KTK) hem de Madde Tepki Kuramı (MTK)' na göre madde ve test istatistikleri elde edilmektedir. KTK'ye göre kullanıldığında bu ölçme araçlarında madde güçlük ve ayırt edicilik katsayıları hesaplanarak ölçme aracının güvenirliği ve geçerliği belirlenir. Çoktan seçmeli ölçme araçlarının güvenirliğini belirlemede birçok yöntem bulunmakla birlikte genellikle Cronbach-α katsayısı, yapı geçerliliği için açımlayıcı ya da doğrulayıcı faktör analizi kullanılmaktadır (Finch ve French 2019). MTK'da ise güvenirliği kestirmek için bilgi fonksiyonları ve marjinal güvenirlik katsayısı kullanılmaktadır (DeMars, 2010; Bond ve Fox, 2015).

## 1.2. Problem Cümlesi

Tek boyutlu 2PL model altında yetenek (θ) parametrelerinin çeşitli dağılım özelliklerine ve farklı örneklem büyüklüklerine sahip olması parametre kestirim keskinliğini nasıl etkiler?

### 1.2.1. Alt problemler

1. Alt Problem: Tek boyutlu 2PL modelde madde parametrelerinin kestiriminde θ parametrelerinin çeşitli dağılım özellikleri gösterdiği durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 0,00, 1,00, 2,00) RMSE (root mean square error) değerlerini nasıl etkiler?

    1.1. Tek boyutlu 2PL modelde a parametresinin kestirimde θ parametrelerinin çeşitli dağılım özellikleri gösterdiği durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 0,00, 1,00, 2,00) RMSE değerlerini nasıl etkiler?

    1.2. Tek boyutlu 2PL modelde b parametresinin kestirimde θ parametrelerinin çeşitli dağılım özellikleri gösterdiği durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 0,00, 1,00, 2,00) RMSE değerlerini nasıl etkiler?

2. Alt Problem: Tek boyutlu 2PL modelde madde parametrelerinin kestiriminde θ parametrelerinin çeşitli dağılım özellikleri gösterdiği durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 0,00, 1,00, 2,00) Bias değerlerini nasıl etkiler?

    2.1. Tek boyutlu 2PL modelde a parametresinin kestirimde θ parametrelerinin çeşitli dağılım özellikleri gösterdiği durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 0,00 1,00, 2,00) Bias değerlerini nasıl etkiler?

    2.2. Tek boyutlu 2PL model ile b parametresinin kestirimde θ parametrelerinin çeşitli dağılım özellikleri gösterdiği durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 0,00 1,00, 2,00) Bias değerlerini nasıl etkiler?

3. Alt Problem: Tek boyutlu 2PL modelde madde parametrelerinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olması (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 0,00, 1,00, 2,00) RMSE değerlerini nasıl etkiler?

    3.1. Tek boyutlu 2PL modelde a parametresinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olması (çarpıklık katsayısı (ÇK)= -2,00, -1,00, 0,00, 1,00, 2,00) RMSE değerlerini nasıl etkiler?

3.2. Tek boyutlu 2PL modelde b parametresinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olması (çarpıklık katsayısı (ÇK)= -2,00, -1,00, 0,00, 1,00, 2,00) RMSE değerlerini nasıl etkiler?

4. Alt Problem: Tek boyutlu 2PL modelde madde parametrelerinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) θ parametrelerinin çeşitli dağılım özelliklerine sahip olduğu durumlarda (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 1,00, 2,00) Bias değerlerini nasıl etkiler?

4.1. Tek boyutlu 2PL modelde a parametresinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olması (çarpıklık katsayısı (ÇK)= -2,00, -1,00, 0,00, 1,00, 2,00) Bias değerlerini nasıl etkiler?

4.2. Tek boyutlu 2PL modelde b parametresinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olması (çarpıklık katsayısı (ÇK)= -2,00, -1,00, 0,00, 1,00, 2,00) Bias değerlerini nasıl etkiler?

## 1.3. Araştırmanın Amacı

Bu araştırmanın amacı, ikili puanlanan maddeler içeren bir test oluşturup 2PL model ile analizinde normallik varsayımının ihlalinin ve değişen örneklem büyüklüğünün parametre kestirim keskinliğine ne gibi etkileri olduğunu belirlemektir.

## 1.4. Araştırmanın Önemi

İkili puanlanan ölçme araçları eğitimde öğrenciyi herhangi bir kuruma yerleştirme, sistemin zayıflıklarını görme; özel sektörde işe alımlarda gerekli yeterliliklere sahip bireyleri seçme, çalışanların performansını gözlemleme gibi birçok alanda ve amaçta kullanılmaktadır. Son yıllarda ise bu ölçme araçlarından elde edilen veri setlerinin MTK modelleri ile analiz edilmesi gittikçe yaygınlaşmaktadır. Ancak MTK modellerinin büyük örneklem gerektirmesi, tek boyutluluk ve yerel bağımsızlık gibi varsayımları araştırmacıların gerçekleşebilecek muhtemel durumlar için veri toplamasını güçleştirmektedir. Gerçekleşmesi muhtemel bu durumlardan biri ise çalışma grubunun normal dağılmıyor olmasıdır. Eğitim, psikoloji gibi alanlarda çoğu zaman çalışma grubunun normal dağıldığı varsayılmaktadır ve çoğu zaman bu varsayım araştırmacı için işleri kolaylaştırmaktadır. Ancak gerçek hayatta ve eğitimde kendine has karakteristik özelliklere sahip bir grubun normal dağılması oldukça zordur hatta bazı durumlarda gruplar aşırı çarpık olabilmektedir (Seong, 1990). Bunun yanı sıra çarpık dağılan yetenek parametrelerinin madde parametre kestirimlerine etkisi ise hala net değildir (Kirişçi, 2001). Bu araştırma özellikle normal dağılımı bir varsayım olarak görmeyen lojistik modellerde normallikten hem negatif hem de pozitif yönde saptıkça parametre kestirim keskinliğinin nasıl etkilendiğine, bu etkinin örneklem büyüklüğünden etkilenip etkilenmediğine cevap olacağı için önemlidir. Bu çalışmayla araştırmacıların yapacakları yeni çalışmalarda kullanacakları örneklem büyüklüklerinin, dağılımın normal olmaması durumunda parametre kestirimlerinin keskinliğini nasıl etkileyeceğini veya kullanacakları çalışma gruplarının dağılımlarının normal dağılmadığı durumda parametre kestirim keskinliğine nasıl bir etkisi olacağını tahmin edebileceklerdir. Ayrıca araştırmacılar kullanacakları ya da elde ettikleri; a) yetenek parametrelerinin çarpıklık durumlarının, b) örneklem büyüklüklerinin çalışmalarının kestirim keskinliğini nasıl etkileyeceğini de tahmin edebileceklerdir. Yine bu çalışma araştırmacılar için R programlama dilinde bir simülasyon çalışmasının nasıl yapıldığı, sonuçların nasıl analiz edildiği konusunda rehber olacaktır. Son olarak bu çalışma simülasyondan ya da gerçek durumlardan elde edilmiş veriler ile gerçekleştirilecek olan çalışmalar için kabul edilebilir sınırlar için bir referans olacaktır.

## 1.5. Sınırlılıklar

1. Araştırma yalnızca yetenek dağılımlarının ve örneklem büyüklüğünün değişiminin incelenmesi ile sınırlıdır.
2. Araştırmada yalnızca -2,00, -1,00, 0,00, 1,00, 2,00 çarpıklık katsayısına sahip dağılımlar kullanılmıştır.
3. Araştırmada dağılımların basıklık değerleri incelenmemiştir.
4. Araştırma Marginal Maximum Likelihood (MML) kestirim yöntemi ile sınırlıdır.

## 1.6. Varsayımlar

Araştırmada üretilen verilerin gerçek durumları yansıttığı varsayılmaktadır.

# BÖLÜM II

## 2. Kuramsal Çerçeve ve İlgili Alanyazın

### 2.1. Kuramsal Çerçeve

#### 2.1.1. Klasik test kuramı

Gerçek Puan Kuramı olarak da bilinen KTK temel olarak gözlenen puanın gerçek puana rastgele hata puanının eklenmesiyle elde edildiği varsayımına dayanır (Crocker ve Algina, 2008; Lord ve Novick, 2008; Finch ve French, 2019; Cronbach, 1990; Allen ve Yen, 1979; Gulliksen, 1950). Formül ile gösterecek olursak;

$$X = T + E$$

Burada X gözlenen puana, T gerçek puana ve E rastgele hata puanına karşılık gelmektedir. Teknik olarak gerçek puan, sonsuza kadar tekrarlanan bir testte hata puanı dengelendiği zaman elde edilir (Cronbach,1990). MTK'de ise puan kavramı yerini yetenek (θ) kavramına bırakmış ve yeteneği belirlemek için çeşitli matematiksel modeller geliştirilmiştir (Hambleton ve Swaminathan, 1985). Bu modeller hesaplama yükü açısından çok daha zahmetli olsa da KTK'nin sınırlılıklarını ortadan kaldırmasıyla çok büyük avantaj sağlamaktadır. Bu sınırlılıklardan belki de en büyüğü test ve birey karakteristiklerinin birbirlerinden bağımsız olmamasıdır. Yetenek kavramı KTK'de gerçek puanla ifade edildiğinden "zor" bir test alan bireyin yeteneği "kolay" bir test aldığı yeteneğinden düşük çıkacaktır. Aynı zamanda test karakteristikleri de bireylere bağlı olduğundan bazı gruplarda madde ayırt ediciliği, güçlüğü ve test güvenirliği ile geçerliği yeterli olurken bir başka grupta bu değerler çok kötü olabilir. Buna bağlı olarak bir diğer sınırlılık ise iki farklı test alan grupları ya da iki farklı gruba uygulanan maddeleri karşılaştırmanın çok zor olmasıdır.

Test ve birey karakteristikleri bağımsız olamadıklarından KTK'de güvenirlik belirleme yöntemlerinden biri olan bir testin paralel iki formundan elde edilen korelasyon katsayılarına bakılmaktır ancak yukarıda bahsedilen sınırlılıklardan dolayı paralel test formları oluşturmak imkânsız değilse de oldukça zordur (Hambleton, Swaminathan & Rogers, 1991). KTK'ye göre ölçmenin standart hatası ölçmenin keskinliği ile doğru orantılıdır (Cronbach, 1990). Ölçmenin standart hatası hesaplanırken güvenirlik katsayısı olan Cronbach-α katsayısı kullanıldığından ve bir testin güvenirliğini yükseltmek için paralel formlar gibi yöntemler kullanıldığından bu da KTK için bir sınırlılıktır. Bir diğer problem ise test puanlarının güvenirliği ve varyansının bir fonksiyonu olan ölçmenin standart hatasının bütün bireyler için aynı kabul edilmesidir. Ancak yukarıda da belirtildiği gibi farklı yeteneklerdeki bireylerin herhangi bir testtin sonuçları eşit olmayacaktır. KTK'nin son varsayımı ise madde yerine test odaklı olmasıdır. Bunun sonucunda bir adayın bir maddeye nasıl cevapladığı hakkında bilgi vermez (Hambleton, vd. 1991).

Sonuç olarak bu sınırlılıkları aşmak amacıyla araştırmacılar alternatif test kuramları için istenen özellikleri belirlemişlerdir. Bunlar;

a) Gruptan bağımsız olarak belirlenebilecek olan madde karakteristiklerinin elde edilebilmesi,
b) Testten bağımsız olarak bireyin yeterliliğine karşılık gelecek olan puanları elde edilebilmesi,
c) Test yerine madde seviyesinde ifade edilebilen bir model olması,
d) Güvenirliği belirlemek için paralel formlar gibi metotlara ihtiyaç duymaması,
e) Her yetenek düzeyi için hassas ölçümler yapabilen bir model olmasıdır.

Yukarıdaki özelliklerin madde tepki kuramında bulunduğu gösterilmiştir (Hambleton ve Swaminathan, 1985; Wright ve Stone, 1979; Hambleton, vd. 1991).

### 2.1.2. Madde tepki kuramı (MTK)

Madde Tepki Kuramı ya da Örtük Özellikler Kuramı bir bireyin belirli davranışlarının yetenek (özellik) olarak tanımlandığı, bireyin bu özelliklerinin nicel olarak tahmin edildiği ve daha sonra bu özelliklerinin performansının matematiksel modeller ile tahmin edildiği bir kuramdır (Lord ve Novick, 2008).

KTK 20. yüzyılda kullanılan ana test kuramı olsa da gerek yukarıda bahsedilen sınırlılıklar gerekse Lord ve Novick (2008)'in tanıtımını yaptığı ve KTK'nin sınırlılıklarını ortadan kaldıran, KTK'ye göre birçok avantaja sahip olan madde tepki kuramının ortaya çıkması MTK'yi alanyazında ana test kuramı olarak kabul ettirmiştir (Embretson ve Reise, 2000). Bu avantajlardan birkaçını; a) ölçülen aynı gizil özellik uygulanan testlerden bağımsız şekilde kestirilebilir, b) madde istatistikleri testin uygulandığı gruptan bağımsız şekilde kestirilebilir ve c) bireylerin kestirilen yeteneklerinin ne kadar isabetli olduğuna dair bir istatistik vermesi şeklinde sıralayabiliriz. Ancak bu avantajlar için geniş bir madde havuzu kullanan test ve geniş bir gruba sahip olma varsayımı dikkate alınmalıdır (Hambleton ve Swaminathan, 1985; Crocker ve Algina, 2008). MTK'nin KTK'ye göre daha avantajlı olduğu açıktır ancak sağlanması oldukça güç olan bazı varsayımlara da sahiptir (Embretson ve Reise, 2000). Bu varsayımlar tek boyutlu MTK modelleri için tek boyutluluk, yerel bağımsızlık, tek boyutluluğu bozabildiği için testin hız testi olmaması ve normal ogive modeller için normalliktir (DeMars, 2010; Lord ve Novick, 2008; Allen ve Yen, 1979). Tek boyutluluk ve yerel bağımsızlık kavramları birbirleriyle ilişkili kavramlardır. Tek boyutluluk varsayımı bir maddenin yalnızca bir yeteneği ölçebileceğine dayanmaktadır. Yerel bağımsızlık ise bireylerin yetenekleri sabit tutulduğunda herhangi bir maddeye verdikleri yanıtların istatistiksel olarak bağımsız olmasıdır. Bir başka deyişle aynı yetenek düzeyindeki bireylerin bir maddelere verdikleri tepkiler arasında ilişkinin olmaması ya da bireyin cevapladığı bir maddenin başka bir madde için ipucu teşkil etmemesidir (Hambleton, vd. 1991; Reckase, 2009). Tek boyutluluk varsayımı sağlandığında yerel bağımsızlık varsayımının da sağlandığı varsayılır. Ancak bu ifade tek boyutluluk ile yerel bağımsızlığın aynı varsayımlar olduğunu göstermez. Bir testin boyut sayısı yerel bağımsızlığı sağlayan örtük özellik sayısı kadardır (Hambleton, vd. 1991; Crocker ve

Algina, 2008; de Ayala, 2009). Bu yüzden tek boyutluluk varsayımı sağlandığında yerel bağımsızlık varsayımı da sağlanmış kabul edilir ancak tersi için bu geçerli değildir. Tüm bunlara ek olarak tek boyutluluk varsayımının sağlanamadığı durumlarda uygulanan çok boyutlu madde tepki kuramı modellerinin de giderek yükselen bir kullanım alanının olduğu bilinmelidir (Shultz ve Whitney, 2005).

MTK'nin bir diğer varsayımı ise normal-ogive modeller için ölçümlerin normal dağılmasıdır (Allen ve Yen, 1979). Normallik varsayımını kontrol etmenin çeşitli yolları vardır. Histogram, gövde yaprak diyagramı, Q-Q grafiği, P-P grafiği, kutu grafiği gibi görsel yöntemlerin yanında Kolmogorov-Smirnov ve Shapiro-Wilk gibi normallik testleri de normalliğin testi için kullanılabilir (Kilmen, 2015; Field, 2009). Bununla birlikte çarpıklık katsayısının (ÇK) hesaplanmasıyla da normallik test edilebilir. ÇK pozitif ya da negatif değerler alabilir ve sıfıra ne kadar yakınsa dağılımın o kadar normal olduğu söylenebilir (Köklü, vd. 2006; Baykul, 2000). Ne var ki gerçek verilerde ÇK'nin sıfır olduğu durumlar oldukça nadirdir ve bu nedenle çarpıklık katsayısının ±1 sınırları içinde bulunması normallikten aşırı bir sapma göstermediği şeklinde yorumlanabilir (Köklü, vd. 2006). ÇK'nin sıfırdan uzaklığı göreceli bir konu olduğundan bu katsayıyı standart z değerlerine dönüştürülerek de normallik varsayımı hakkında yorumda bulunulabilir (Kilmen, 2015).

ÇK, bir dağılımda verilerin daha çok nerede yığıldıklarını göstermektedir. ÇK pozitif değerler aldıkça dağılım sağa, negatif değerler aldıkça dağılımın sola çarpık olduğunu göstermekle beraber sıfıra yakın değerler aldığında da simetrik bir özellik gösterir. Bu özelliğe ise dağılımın çarpıklığı denmektedir. Çarpıklık katsayısı;

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

eşitliği ile hesaplanabilir (Baykul, 2000). Bu formülde;

n: testi alan birey sayısı,

Xi: i. öğrencinin gözlenen test puanını,

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

S: ilgili testin standart sapmasını ifade etmektedir.

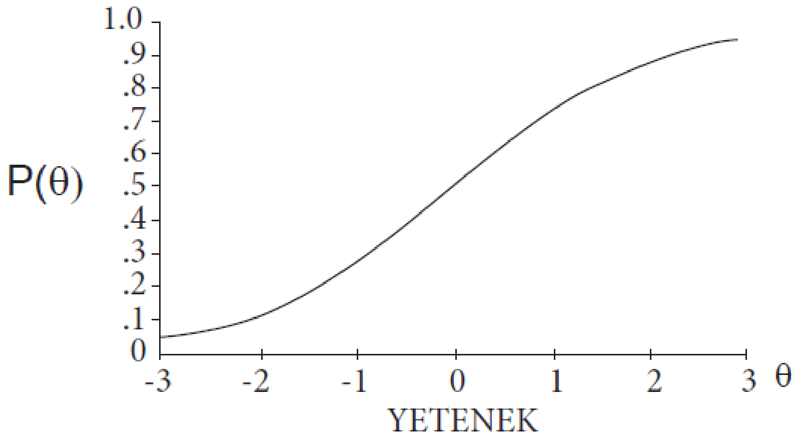
Bunun dışında ÇK;

$$ÇK = \frac{3(\overline{X} - Ortanca)}{S}$$

$$ÇK = \frac{\sum(X - \overline{X})^3/n}{S^3}$$

formülleriyle de hesaplanabilmektedir (Köklü, vd. 2006).

Ölçme işlemini gerçekleştirmek için bir değişkene ihtiyaç vardır. Zekâ, okuma yeteneği, aritmetik yetenek gibi kolay tanımlanabilen, özelliklerinin sıralanması kolay olan ancak ağırlık, uzunluk gibi doğrudan ölçülmesi mümkün olmayan değişkenler eğitim ve psikoloji alanında sıkça kullanılır. Bu tür değişkenlere MTK'de *"yetenek (ability)"* ya da *"örtük özellik"* adı verilmekte ve θ (theta) sembolü ile gösterilmektedir. Herhangi bir bireyin ölçülmesi hedeflenen yeteneğe ne kadar sahip olduğunu belirlemek için bir ölçeğe ihtiyaç vardır. Hedeflenen bu yetenek ne olursa olsun -∞ ile +∞ arasında uzanan orta noktası "sıfır" ve ölçme birimi "bir" olan ölçek ile sahip olunan yetenek miktarının belirlenebileceği varsayılır. Bir test maddesini doğru yanıtlayan bireyin o maddenin ölçtüğü yeteneğe bir miktar sahip olduğu söylenebilir. Buna göre her bireyin yetenek ölçeği üzerinde kendine ait bir puana sahip olacağı düşünülebilir. Her bir yetenek düzeyinde, bu yetenek düzeyindeki bireyin maddeyi doğru cevaplaması üzerine belirli bir olasılık değeri yer alır. Bu olasılık değeri P(θ) ile gösterilir. P(θ) 0 ile 1 arasında değerler alır ve P(θ) değeri 0'a yaklaştıkça bireyin yeteneği, dolaylı olarak da maddeyi doğru cevaplama ihtimali düşerken 1'e yaklaştığı durumda bireyin yeteneği, dolaylı olarak da maddeyi doğru cevaplama ihtimali artar. Eğer P(θ) değerleri, yetenek düzeyinin bir fonksiyonu olarak çizilirse Şekil 1.1'deki gibi S şeklinde bir eğri elde edilecektir. Bu eğri yetenek ölçeği ile maddeyi doğru cevaplama olasılığı arasındaki ilişkiyi ifade etmektedir ve her madde için bu eğri elde edilir (Baker, 2001). MTK'de bu eğriye "madde karakteristik eğrisi", "madde tepki fonksiyonu" ya da "madde karakteristik fonksiyonu" da denilmektedir (Lord ve Novick, 2008; DeMars, 2010; Baker, 2001). Bu çalışmada bu fonksiyonun adı madde karakteristik eğrisi (MKE) olarak geçecektir.




**Şekil 1.1.** Madde karakteristik eğrisi.

Seçilen modele bağlı olarak madde karakteristik eğrisi 3 farklı bilgi barındırabilir (Hambleton, vd. 1991; DeMars, 2010; Finch ve French, 2019). Bunlardan ilki *"madde güçlüğü"* adı verilen ve çeşitli farklılıklar bulundurmakla beraber KTK'deki madde güçlüğüne (p) benzeyen parametredir. b parametresi olarak da adlandırılan bu parametre doğru cevap verme olasılığının 0,5 olduğu noktanın karşılık geldiği θ parametresinin değerine eşittir. Gruplardan elde edilen θ değişkenleri ortalamaları 0 ve standart sapmaları 1 olacak şekilde dönüştürüldüğünde b parametresi genelde -2,00 ile +2,00 arasında değerler almaktadır (alanyazında bu değerin -3,00 ile +3,00 arasında değiştiğini söyleyen kaynaklar da vardır) ve aldığı değerler büyüdükçe madde zorlaşır, küçüldükçe de madde kolaylaşır (Hambleton, vd. 1991; DeMars, 2010). b parametresine aynı zamanda MKE'nin ölçek üzerindeki yerini belirlediğinden *"lokasyon parametresi"* de denilmektedir (Price, 2017).

MTK'nin ikinci parametresi ise *"madde ayırt ediciliği"* dir ve a parametresi olarak da bilinmektedir. Madde ayırt ediciliği b noktasındaki MKE'nin eğimiyle doğru orantılıdır. b parametresinin bulunduğu noktanın eğimi ne kadar yüksek olursa yani MKE ne kadar dik olursa madde, bireyleri o kadar iyi ayırt edecektir. Madde ayırt edicilik parametresi teorik olarak -∞ ile +∞ arasında değer almaktadır ancak negatif değer aldığı durumlarda maddenin ters çalıştığına işaret ettiğinden bu gibi maddeler testten çıkarılmaktadır. Aynı zamanda +2,00'den büyük değerler elde etmek de olağan dışı olduğundan a parametresinin ranjı genellikle 0 ile 2 arasında olmaktadır (Hambleton, vd. 1991; DeMars, 2010).

MTK'nin son parametresi ise *"şans parametresi"* dir ve *"tahmin parametresi"* olarak da adlandırılmakla birlikte c ile gösterilmektedir. Bu parametre MKE'nin düşük asimptot değerinin sıfır olmayacağını gösterir ve düşük θ düzeyindeki bireylerin maddeyi nasıl doğru cevapladığını betimler. c parametresi teorik olarak 0 ile 1 arasında değerler alır ancak genelde 0 ile 0,35 arasında kabul edilmektedir. (Hambleton, vd. 1991; DeMars, 2010; Baker, 2001).

#### 2.1.2.1. Madde Tepki Kuramı modelleri

MTK, gözlenebilen değişkenler ile (maddelere verilen tepki) gözlenemeyenler (okuma yeteneği, matematik yeteneği vb.) arasındaki ilişkinin matematiksel fonksiyonlarla tanımlandığı bu yüzden matematiksel modellerden oluşan bir kuramdır. Fonksiyonlar değiştiğinde ortaya yeni bir MTK modeli de çıkar. Bu nedenle sonsuz sayıda MTK modeli geliştirilebilir (Hambleton ve Swaminathan, 1985). Sonsuz sayıda oluşturulabilecek modellerin sınıflandırılmaları bilim insanlarının yaptıkları çalışmalarda değişiklik gösterebilmektedir.

Bu sınıflandırmalardan biri ise madde yanıtlarının puanlama sistemine dayalı olarak; ikili puanlanan, çoklu puanlanan ve sürekli puanlanan şeklinde yapılabilir. Yıllar boyunca ikili puanlanan çoktan seçmeli testler ise eğitim alanında en çok kullanılan yöntemler olagelmiştir (Hambleton ve Swaminathan, 1985).

MTK'de kullanılan veriye göre uyumlu olan modeli kullanmak büyük önem taşımaktadır. Bu uygunluğu belirlemede ise birçok ölçüt kullanılmaktadır. Bunlar;

a) Maddelerin puanlanmasındaki ağırlıklandırma (eşit ya da eşit olmayan),
b) Ölçme için istenen ölçek özellikleri,
c) Model-veri uyum indeksleri,
d) Parametre kestirimindeki amaç

olarak sıralanabilir. Örneğin maddeler eşit olarak ağırlıklandırılmış ve ölçek özellikleri için güçlü bir doğrulama isteniyorsa bir parametreli modelin kullanılması uygun olacaktır. Ancak maddelerin uyumunun yüksek ya da yüksek kesinlikte parametre tahminlerine ihtiyaç duyuluyorsa bu kez iki veya üç parametreli modeller kullanılabilir. Eğer esas amaç bireylerin yeteneklerini en doğru biçimde kestirmekse model uyumları ki kare istatistiği yardımıyla karşılaştırılarak en uygun model belirlenebilir (Embretson ve Reise, 2000). Bunlara ek olarak elde edilmek istenen bilgi miktarı da model seçiminde etkilidir. Örneğin madde ayırt edicilik ve güçlük parametreleri hakkında bilgi isteniyorsa iki parametreli model kullanılmalıdır (Finch ve French, 2019).

#### 2.1.2.1.1. Normal ogive modeller

Ortaya çıkan ilk model olmasa da normal-ogive model MTK'nin erken dönemlerinde sıklıkla kullanılmıştır. Lord (1952) tarafından geliştirilen bu modelde kullanılan varsayımlar bugün lojistik modellerde kullandığımız varsayımları da içermektedir. Bunun bir istisnası ise normal-ogive modelin formülü gereği doğru cevap verme olasılıklarını standart normal dağılım altında bulmasından kaynaklanan yeteneklerin normal dağılım gösterdiği varsayımıdır.

Normal-ogive modelin formülü gereği;

$$Pi(\theta) = \int_{-\infty}^{w} f(z)\, dz$$

$P_i(\theta)$: Bireyin i maddesini doğru cevaplama olasılığı (madde karakteristik eğrisi),

w: $a_i(\theta - b_i)$,

f(z): $\frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ Normal dağılım fonksiyonu,

$\int_{-\infty}^{w} dz$: -∞ ile w arasındaki dağılım alanı için integral notasyonu.

f(z) fonksiyonu Gauss dağılımı olarak da bilinen normal dağılım eğrisini veren formüldür (Crocker ve Algina, 2008). Normal-ogive modelde formül gereği MKE integral yardımıyla standart normal dağılımda alan hesaplanarak elde edilir. Ancak bu, matematiksel hesaplamalar açısından lojistik modele göre çok güç bir işlemdir (Crocker ve Algina, 2008; Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1985; van der Linden ve Hambleton, 1997; Birnbaum, 1967; Berkson, 1951; Lord ve Novick, 2008; Allen ve Yen, 1979).

#### 2.1.2.1.2. Tek boyutlu lojistik modeller

Daha önce belirtildiği gibi MTK'de birçok model çeşidi bulunmaktadır. Bunlardan bir tanesi ise tek boyutluluk varsayımına uyan modellerdir. Erken dönem MTK çalışmalarında yoğunlukla normal-ogive modeller kullanılmış olsa da lojistik modellerin gelişmesi ve görece daha kolay matematiksel işlemler gerektirmesiyle günümüzde normal-ogive modelin yerine tercih edilmektedir (Crocker ve Algina, 2008).

Birnbaum tarafından geliştirilen bu model basit bir matematiksel modelle doğru cevap verme olasılığını bulan lojistik (kümülatif) dağılıma dayanmaktadır (Lord ve Novick, 2008; Birnbaum, 1967; Embretson ve Reise, 2000).

Aşağıdaki bölümlerde detaylandırılmak üzere lojistik modellerin temel formülüne değinecek olursak;

$$Pi(\theta) = \frac{e^{x}}{1 + e^{x}} = \frac{1}{1 + e^{-x}}$$

$P_i$ : i maddesine doğru cevap verilme olasılığı,

e : Euler sayısı 2.718… değerindedir ve sonsuza gider,

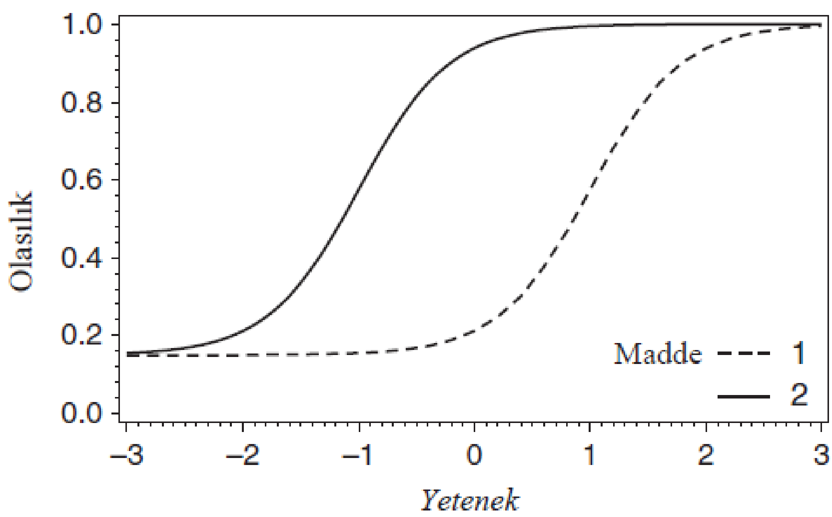
x : $Da_i\ (\theta - b_i)$,

D : 1.702 değerinde, lojistik modellerin normal-ogive modele yaklaşık değerler üretmesini sağlayan sabit sayıdır.

Haley (1952) tarafından geliştirilen D sabiti sayesinde normal-ogive model ve lojistik model çok daha benzer değerler almaya başlamış ve lojistik modeller normal-ogive modellere kıyasla daha çok tercih edilmeye başlamıştır (Allen ve Yen, 1979).

2.1.2.1.2.1.Rasch modeli (1 parametreli lojistik model (1PL))

Rasch modeli madde ayırt edicilik indeksinin (a) 1 değerine sabitlendiği ve yalnızca madde güçlük parametresinin (b) serbest bırakıldığı, maddelerin b parametresine göre değişim gösterdiği bir modeldir (Embretson ve Reise, 2000). Buna ek olarak Rasch modelinde şans parametresinin (c) minimal değerde olduğu ya da bireylerin tahmin yoluyla maddelere cevap veremeyecekleri varsayılır (Hambleton ve Swaminathan, 1985; Hambleton, vd. 1991). Araştırmacılar arasında sıklıkla tercih edilen bu modelin diğer modellere kıyasla birkaç avantajlı özelliği bulunmaktadır. Bunlardan ilki diğer modellerden daha az parametre içermesi ve böylelikle veri üzerinde daha kolay çalışılabilmesidir. İkincisi ise yine daha az parametreye sahip olmasından kaynaklı parametre tahmin işlemlerinin daha keskin yapılabilmesidir (Hambleton ve Swaminathan, 1985; van der Linden, 2016).




**Şekil 2.1.** Madde güçlük (b) katsayıları farklı iki madde.

Yukarıdaki Şekil 2.1'de 2 adet maddenin MKE'si görülmektedir. a parametreleri sabit olduğundan eğimleri aynıdır. Ancak b parametreleri değişim gösterdiği için farklı bölgelerde bulunmaktadırlar.

Rasch, modelini MTK çalışmalarından bağımsız olarak geliştirmiş olsa da matematiksel olarak 1PL model ile birbirlerine eşdeğerdir (de Ayala, 2009; Hambleton ve Swaminathan, 1985; Allen ve Yen, 1979). Bu modellerin formülüne değinmek gerekirse;

$$Pi(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

şeklinde ifade edilebiliriz. Burada i maddesine ait olan ayırt edicilik parametresi ($a_i$ ) 1 değerine sabitlenmiştir. Ancak her ne kadar Rasch modeli ile 1PL model birbirlerine eşdeğer de olsa aralarında farklılıklar da bulunmaktadır. Rasch modelinde a parametresi 1 değerine sabitlenirken 1PL modelde a parametresi belli bir değere sabitlenmek yerine modelin verilere en iyi uyum gösterdiği değere sabitlenir (de Ayala, 2009; Price, 2017). Bu farklılıkla beraber Rasch modelinin formülü şu şekilde de yazılabilir;

$$Pi(\theta) = \frac{e^{(\theta_i^* - b_i^*)}}{1 + e^{(\theta_i^* - b_i^*)}}$$

$\theta^*$: D sabiti ile θ'nın çarpımı,

$b_i^*$ : D sabiti ile i maddesinin b parametre değerinin çarpımını ifade etmektedir (Crocker ve Algina, 2008; Wright ve Stone, 1979).

2.1.2.1.2.2. İki parametreli lojistik model

Bu modelin kurucusu olan Birnbaum, MTK alanındaki çalışmalarına her ne kadar 1950'li yılların sonlarında başlamış olsa da modellerinin yaygınlaşması Lord ve Novick'in 1968 yılındaki çalışmasından sonra olmuştur. Rasch'tan farklı olarak amacı bir test modeli ortaya koymak değil, Lord (1952) tarafından başlatılan normal-ogive model çalışmalarının istatistiksel olarak çürütülebilir olduğunu göstermekti. Böylelikle 2PL model MKE için kullanılan ilk lojistik model olarak karşımıza çıktı (van der Linden ve Hambleton, 1997; Baker, 2001).

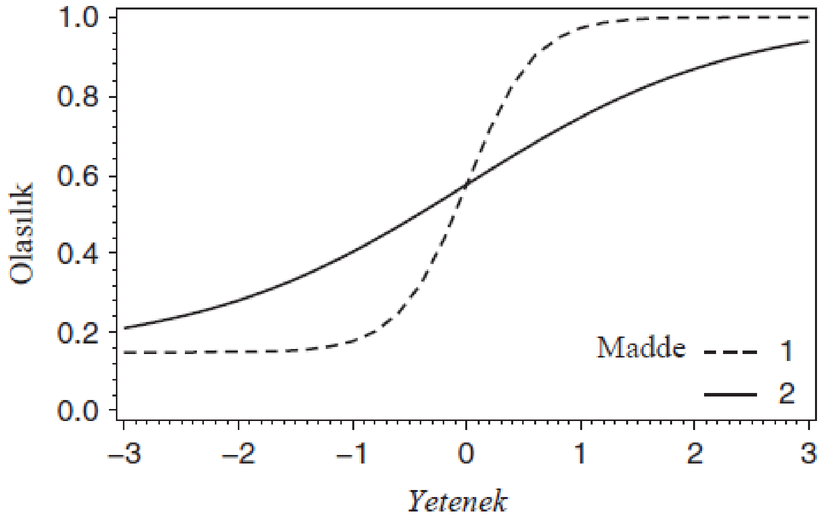
1PL model ve Rasch modelin sağladığı kolaylıklara rağmen madde ayırt edicilik parametresinin belli bir değere sabitlenmiş olma varsayımı çoğunlukla sağlanamayacak kadar zordur. Bu varsayım sağlanamadığında ise araştırmacılar genelde 2PL modeli kullanmayı tercih etmektedir (Finch ve French, 2019). Birnbaum tarafından geliştirilen bu modelin formülüne bakmak gerekirse;

$$Pi(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

şeklindedir. Kullanılan katsayılar 1PL model ve Rasch modeli için kullanılanlar ile aynıdır. Bundan dolayı tanımlamaya gidilmeyecektir.

Görülebileceği üzere 2PL modelin Rasch veya 1PL modelden bir farkı yoktur. 2PL modeli tanıtılan diğer modellerden farklı kılan ise a parametresinin kestirimini serbest bırakmasıdır. Diğer bir deyişle a parametresi her madde için farklı değer alacaktır (Price, 2017; Finch ve French, 2019; Embretson ve Reise, 2000; de Ayala, 2009). 2PL modelin tanıtılan diğer modellerden ayrılan bir diğer özelliği veri ile modelin uyumluluğunun diğer modellere kıyasla daha iyi değerler almasıdır (de Ayala, 2009). Tüm bunlara ek olarak 2PL model de 1PL model ve Rasch modelinde olduğu gibi şans parametresinin (c) 0 olduğunu varsaymaktadır (Hambleton, vd. 1991; de Ayala, 2009).

Daha önce değinildiği gibi maddede yüksek ayırt edicilik istenen değerlerdir. Yüksek ayırt edicilik değerleri maddenin farklı θ düzeyindeki bireyleri iyi derecede ayırt ettiğini gösterirken bu bireyler içinde yüksek miktarda bilgi sağlar. Ayrıca daha düşük standart hata değerleri verir ve ölçümün keskinliğini arttırır (Finch ve French, 2019).




**Şekil 2.2.** Madde ayırt edicilik (a) katsayıları farklı iki madde.

Şekil 2.2'de b parametreleri aynı olan ancak a parametreleri farklı olan iki madde gösterilmiştir. Eğimin arttıkça ayırt ediciliğin de ne kadar arttığı görülmektedir.

2.1.2.1.2.3. Üç parametreli lojistik model

Kullandığımız testler özellikle çoktan seçmeli maddeler içeriyor ve bireyin herhangi bir alandaki başarısıyla ilgili ölçümler yapıyorsa şans başarısının diğer bir deyişle maddeleri şansla ya da tahminle doğru cevap verme durumunun görmezden gelinmesi çok zordur.

Çoktan seçmeli testlerde birey bir maddeye tamamen rastgele doğru bir cevap verecekse bu onun 1/seçenek sayısı kadar şansının olduğu anlamına gelmektedir. Örneğin 5 seçenekli bir maddeyi rastgele doğru cevaplama olasılığı 1/5 olduğu için yüzde 20'dir. Ancak önceki bölümlerde bahsettiğimiz şans parametresi (c) nadiren bu oranlarda değerler almaktadır (Finch ve French, 2019).

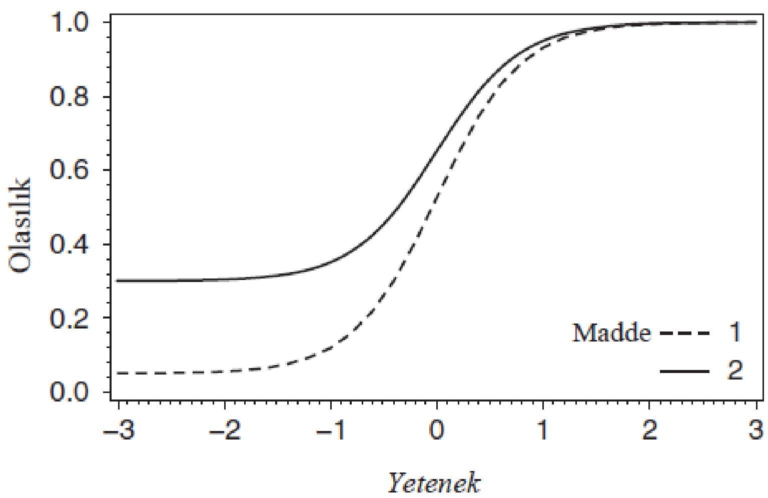
Bu durumun sebeplerine eğilecek olursak en temelde maddeleri hazırlayan uzman kişilerin çeldirici seçenekleri soruyu doğru yanıtlamaya yetecek kadar yeteneği olmayan bireylere çekici gelecek şekilde hazırlama becerilerine dayanmaktadır (Finch ve

French, 2019; Lord, 1974; Embretson ve Reise, 2000) . Yani iyi hazırlanmış bir maddenin çeldirici seçeneği düşük θ düzeyindeki bireyleri kendine daha çok çekebilir. Tam aksine kötü hazırlanmış bir maddede de düşük θ düzeyindeki birey bile seçenekleri kolayca eleyerek maddeyi doğru cevaplayabilir ya da doğru cevap verme şansını arttırabilir. Bu ve benzeri durumlarda bir maddenin şansla doğru cevaplanma olasılığının her zaman 1/seçenek sayısı olamayacağı son derece açıktır. Bireyin risk almaya yatkınlığı, yorgunluk derecesi, maddenin çok zor ya da kolay bir madde olması vb. sebepler de c parametresini etkileyen unsurlar içinde sayılabilir (de Ayala, 2009; Lord, 1974). Yukarıda c parametresini etkileyen durumlardan yola çıkarak aslında c parametresinin madde parametresinden çok birey parametresi olmaya daha yakın olduğu görülecektir (de Ayala, 2009).

3PL model 2PL modele c parametresinin eklenmesiyle oluşturulan bir formüle sahiptir (Hambleton, vd. 1991). Bu matematiksel eşitliği göstermek gerekirse;

$$Pi(\theta) = c_i + (1 - c_i)\,\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

şeklinde ifade edebiliriz. 2PL modelden temel farkı görülebileceği gibi c parametresinin eklemesidir. Ancak formülde yapılan bu değişiklik 3PL modelin bazı çok yararlı matematiksel yararlarını götürmüştür. Bundan dolayı aslında 3PL model artık bir lojistik model değildir (Baker, 2001).




**Şekil 2.3.** Madde şans parametreleri (c) farklı iki madde.

Şekil 2.3'te c parametreleri farklı iki madde görülmektedir. Anlaşılabileceği üzere Madde 2'nin şans ya da tahmin yoluyla doğru cevaplanması daha ihtimal dâhilindedir.

c parametresi de aslında diğer madde parametreleri gibi her madde için ayrı ayrı tespit edilmelidir. Ancak bunu yapmak kestirim işlemlerinde birçok karmaşaya yol açacağından tüm bir madde grubu için ya da bir gruba uygulanan denk maddeler için ortak bir c parametre değeri hesaplanmaktadır. Diğer bir deyişle yüksek θ düzeyindeki birey ile düşük θ düzeyindeki bireyin ortak bir maddeye şansla doğru cevap verme olasılığı eşittir (Embretson ve Reise, 2000).

Tüm bu dezavantajlarına rağmen 3PL modeli kullanmak eğer ölçme aracında bulunan maddelere şans yoluyla doğru cevap verme olasılığı mümkünse bir avantaj sağlayacaktır (Price, 2017).

### 2.1.3. Parametre kestiriminde keskinlik (Ölçme keskinliği)

Bir pazarcı, bir kuruyemişçi ürünlerini tartmak için eşit kollu terazi kullanır ancak çoğu zaman müşterinin üründen istediği miktarı tam olarak tutturamaz. Benzer şekilde bir bireyin matematik ya da okuduğunu anlama yeterliliği de çoğu zaman doğru biçimde tespit edilemez. Yukarıda örnek verdiğimiz durumlar sırasıyla doğrudan ve dolaylı ölçme olarak adlandırılmaktadır. Kısaca her türlü ölçme işleminde az ya da çok bir miktar hata bulunacaktır (Turgut ve Baykul, 2015).

Psikolojik yapıların ölçümü için dolaylı ölçme yöntemlerinin kullanıldığına ve KTK'de bir ölçümün hatasız yapılması için sonsuz kez tekrarlanması gerektiğine değinmiştik. Ancak ölçme işlemlerinde, pratikte bir ölçme işleminin sonsuz kez uygulanamayacağı gerçeğinden hareketle gerçek durum ya da puan ile gözlenen durum ya da puan arasındaki farkın olabilecek en düşük değerde tutulması esas alınmaktadır. Bu farklara *"hata"* denilmekle beraber ölçme işleminin kalitesinin yani keskinliğinin yorumlanmasında yardımcı olur.

MTK’de gerek parametre kestirim yöntemlerinin verimliliğini gerekse çeşitli koşullarda madde ve birey parametrelerinin ne derece hassaslıkta kestirildiğini öğrenmek için çeşitli simülasyon çalışmaları yapılmaktadır (Feinberg ve Rubright, 2016). Bu simülasyon çalışmaları ise şu birkaç adımda gerçekleştirilmektedir;

1- Bir MTK modeli seçilir.
2- Modelin madde ile birey parametreleri test uzunluğu, örneklem büyüklüğü gibi farklı koşullar altında üretilir.
3- Modelin parametre değerlerini kestirmek için bir ya da daha fazla sayıda kestirim yöntemi belirlenir ve kestirim işlemi yapılır.
4- Üretilen parametre değerleri ile kestirim sonucu elde edilen parametre değerleri karşılaştırılarak sonuca varılır (Luecht ve Ackerman, 2018).

Yukarıdaki adımlar incelendiğinde aslında 2. maddenin bir ölçme işlemindeki gerçek durumu ya da puanı 3. maddenin ise gözlenen durumu ya da puanı temsil ettiği ve 4. maddede ise gerçek puan ile gözlenen puan arasındaki farkın incelendiği yani *“hata”* değerlerine bakıldığı anlaşılabilecektir. 4. maddede yapılacak olan karşılaştırma işlemini yorumlamak için MTK’de *“hata kareleri ortalamasının karekökü (root mean squared error (RMSE))”* ve *“yanlılık (Bias)”* değerleri kullanılabilir. Yanlılık değerini bulmak için;

$$Bias = \frac{\sum_{i=1}^{K}(\hat{X}_i - X_i)}{K}$$

formülü kullanılabilir. Burada K toplam madde sayısını, i madde numarasını, $\widehat{Xi}$ kestirilen i. madde parametre değerini, $X_i$ ise üretilen yani gerçek madde parametre değerini göstermektedir. RMSE değerini bulmak içinse;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{K}(\hat{X}_i - X_i)^2}{K}}$$

formülü kullanılabilir. Bu formüllerden ne kadar küçük değerler elde edilirse ölçme işleminin ya da parametre geri kazanımının o kadar hatasız olduğunu söylenebilir, işlemin keskinliği yorumlanabilir.

Bir veri setini bir modelle uyumlamanın belki de en önemli adımı madde ve birey parametrelerinin veri setinden kestirildiği andır. Genellikle bu kestirimlerin keskinliği MTK modelinin varsayımları ihlal edilmesine, kullanılan kestirim yöntemine, kestirim yönteminin varsayımlarının ihlal edilmesine, kullanılan bilgisayar programına vb. birkaç nedene dayandırılabilir (Kirişçi vd., 2001; Yıldırım, 2015).

### 2.1.4. Monte Carlo simülasyonu çalışma prensipleri

Simülasyon çalışmaları, bilginin gerçek hayatta elde edilmesinin çok zor ya da imkânsız olduğu durumlarda çok kullanışlıdır. Bu simülasyon çalışmalarından biri olan Monte Carlo (MC) yöntemi ise birçok araştırma sorusuna cevap aramak için kullanılabilir. Bu durumlar normal dağılımdan üretilmiş bir veri setinden parametre kestirimlerini karşılaştırmak kadar kolay, çok boyutlu MTK modellerinde veri üretmek kadar zor olabilirken MTK ya da tahmin yöntemi varsayımlarının ihlal edildiği durumları inceleme gibi çeşitli olabilir (Feinberg ve Rubright, 2016; Harwell vd., 1996).

Bir MC çalışması oluşturmak için atılması gereken birkaç adım vardır. Bunlar;

1. Bir ya da daha fazla araştırma sorusu belirlenir. Bu “ 2PL modelde örneklem büyüklüğü değişiminin parametre kestiriminin keskinliğine etkisi nedir?” gibi basit bir soru olabilir.
2. Madde sayısı, örneklem büyüklüğü, dağılımların çarpıklığı gibi ilgilenilen bağımsız değişkenler belirlenir. Eğer bu simülasyon faktörleri tamamen çaprazsa faktörlerin olası her birleşimi için bir veri setine ihtiyaç duyulacaktır. Örneğin bir çalışmada üç farklı madde sayısı, üç farklı örneklem büyüklüğü ve üç farklı dağılım kullanılacaksa ve bu faktörler tamamen çaprazsa toplamda 3x3x3=27 veri seti üretilmelidir. Ancak faktörlerin kısmen çapraz olduğu varsayılmışsa yalnızca ilgi faktörlerin birleşimi kadar veri seti üretilmelidir.

3. Üretilecek olan verilerin doğası hakkında varsayımlar yapılır. Buradaki amaç simülasyon çalışmasının gerçek hayattan kopuk olmamasını sağlamaktır. Örneğin bir araştırmacı çalışmasında 300 maddelik tek boyutlu bir veri seti kullanmak isteyebilir

ancak bu kadar çok madde içeren bir veri setinin gerçek hayatta uygulanması çok da mümkün değildir.

4. Yukarıda bahsedilen simülasyon faktörleri ve yapılan varsayımlar doğrultusunda her koşul için birçok veri seti üretilir. *"Replikasyon"* olarak da bilinen bu işlem MC simülasyon çalışmalarında kestirilen parametrelerin dağılımını inceleme, tek bir veri setinden elde edilebilecek absürt sonuçları elimine etme, yapılan varsayımlar doğrultusunda üretilen parametreler üzerinde yeniden örnekleme yapma imkânına sahip olma amaçları için yapılmaktadır.

5. Simüle edilen veri setlerinde istatistiksel analizler yapılarak ilgilenilen parametre kestirim değerleri kaydedilir.

6. Son olarak kestirilen parametreler istatistiksel yöntemlerle (RMSE, Bias, korelasyon vb.) değerlendirilir ve uygun şekilde raporlaştırılır (Harwell vd.,1996; Bulut ve Sünbül, 2017).

Bu adımlara ek olarak araştırmacılar bir MC çalışmasında şu üç kurala dikkat etmelidir;

1. *Gerçeklik:* Yukarıda 3. maddede bahsettiğimiz gibi simülasyon faktörleri gerçek hayattan kopuk olmamalıdır.
2. *Uygulanabilirlik:* Bir simülasyon çalışmasında ne kadar bağımsız değişken yani faktör varsa çalışma o kadar karmaşık olacak ve uygulanması da bir o kadar zahmetli olacaktır. Örneğin bir araştırmacı 2PL model için 4 farklı örneklem büyüklüğü (100, 200, 500, 1000, 2000), 5 farklı madde sayısı (10, 15, 20, 40, 60), 3 farklı yetenek (θ) dağılımı (normal dağılım, sağa çarpık dağılım, sola çarpık dağılım) belirlemiş olsun. Bu deneyin deseninde toplam 4x5x3= 60 veri seti olacağı anlamına gelmektedir. Eğer araştırmacı bu veri setine 10,000 replikasyon uygularsa toplamda 60,000 veri seti elde etmiş olur ki çok iyi donanımsal ve yazılımsal imkanlara sahip olsa bile simülasyon çalışmasını tamamlaması hatrı sayılır miktarda zaman alacaktır.
3. *Tekrar Edilebilirlik:* MC çalışmalarında tekrar edilebilirlik önceden yapılmış bir simülasyon çalışmasının yeniden aynı koşullarla yapılarak aynı sonuçların elde

edilebilmesidir. Ancak unutulmamalıdır ki kullanılan donanım ve yazılımlar muhtemel farklılıklara yol açabilir (Bulut ve Sünbül, 2017).

## 2.2. İlgili Alanyazın

Bu bölümde bu çalışma ile ilgili yapılan araştırmalara özet olarak yer verilmiştir.

Swaminathan ve Gifford (1979), gerçekleştirdikleri araştırmada 3PL model altında test uzunluklarının (10, 15, 20 ve 80 madde), örneklem büyüklüklerinin (N=50, 200 ve 1,000) ve θ parametrelerinin dağılımlarının (normal, çarpık, uniform) farklı kestirim yöntemleri (Urry ve maximum likelihood-ML) kullanıldığı durumlarının parametre kestirim keskinliğine olan etkilerini incelemeyi amaçlamışlardır. Yapay verileri üretmek için DATGEN yazılımını kullanan araştırmacılar yetenek (θ) parametrelerini uniform dağılım için ortalaması 0 ve standart sapması 1 olan dağılım için -1,73 ve 1,73 aralığında parametre değerleri üreten uniform dağılımından, çarpık dağılım için α=5 ve β=1,5 olan beta dağılımından üretmişlerdir. Madde parametrelerini üretmek içinse a parametrelerinde 0,6 ve 2,0 arasında değerler üreten uniform dağılım, b parametrelerinde -2,0 ve 2,0 arasında değerler üreten uniform dağılım kullanmışlardır. Araştırmacılar c parametrelerinin tümünü 0,25 değerine sabitlemişlerdir. Elde ettikleri veri setlerinden parametreleri kestirmek için ANCILLES yazılımında Urry kestirim yöntemini, LOGIST yazılımında ise ML kestirim yöntemini kullanmışlardır. Kestirim yöntemlerinden elde ettikleri sonuçları kestirilen ve üretilen parametrelerin ortalamalarını ve korelasyonlarını karşılaştırarak değerlendirmişlerdir. Araştırmanın sonucunda ML kestirim yöntemi genellikle tüm madde (a, b, c) ve yetenek (θ) parametre kestiriminde Urry yöntemine üstünlük sağlamıştır.

Hulin, Lissak ve Drasgow (1982), yaptıkları MC çalışmasında 2PL ve 3PL modellerde örneklem büyüklüğünün ve madde sayısının değişken olduğu durumların parametre kestiriminin hassasiyetinde ne gibi değişikliklere yol açtığını gözlemlemeyi amaçlamışlardır. Bu amaç doğrultusunda 200, 500, 1000 ve 2000 kişilik örneklem büyüklüğü; 15, 30 ve 60 maddeden oluşan test uzunluğu kullanmışlardır. a, b ve c parametrelerini uniform dağılımdan (sırasıyla min= 0,3, max= 1,4; min= -3, max= 3; min= 0,11, max= 0,33), θ parametrelerini ise normal dağılımdan FORTRAN yazılımında SPECTRUM IV paketinde üretmişlerdir. Parametre kestirimi için LOGIST yazılımında Maximum Likelihood (ML) kestirim yöntemi kullanmışlardır. Madde ve θ parametrelerini değerlendirmek için gerçek parametreler ile kestirilen parametrelerin korelasyon değerlerini incelemenin yanında üretilen madde parametrelerinin madde karakteristik eğrileri (MKE) ile kestirilen madde parametrelerinin MKE'si arasındaki RMSE değerlerini incelemiş ayrıca üretilen ve kestirilen θ parametrelerinin RMSE değerlerini ele almışlardır. Yaptıkları analizleri incelediklerinde madde karakteristik eğrilerinden elde ettikleri RMSE istatistikleri hem 2PL model hem de 3PL model için 30 ve 60 maddeden oluşan test uzunluklarında 1000 ve 2000 kişilik örneklemlerde son derece iyi değerlerde yer almıştır. 2PL modelde 30 ve 60 maddelik test uzunluklarında 500 kişilik örneklemlerde nispeten yine iyi sonuçlar elde etmişler ancak 15 maddelik ve 200 kişilik çalışma grubunda son derece kötü değerler elde etmişlerdir. 3PL modelde ise 60 maddelik 500 ve 200 kişilik örneklemlerde nispeten küçük RMSE değerlerine ulaşırken madde sayısı ve örneklem büyüklüğü azaldıkça daha büyük RMSE değerleri elde etmişlerdir. θ parametreleri için elde edilen sonuçlara baktıklarında 2PL model için test uzunluğu sabit tutulduğunda örneklem büyüklüklerinin kestirimlere etkisinin çok düşük düzeyde olduğunu tespit etmişlerdir. Ancak 3PL model için korelasyon değerlerine baktıklarında örneklem büyüklüğünün kestirim sonuçlarına net bir etkisinin olduğunu gözlemlemişlerdir.

Yen (1987), yaptığı çalışmasında BILOG ve LOGIST yazılımlarından elde edilen sonuçların keskinliğini karşılaştırmalı olarak araştırmayı amaçlamıştır. Bu amaçla Yen (1984) 3PL modelde 1000 kişilik bir örneklem büyüklüğü için 1 adet 10 maddelik, 4 adet 20 maddelik ve 4 adet 40 maddelik toplam 9 veri seti üretmiştir. Ayrıca 1000 kişilik örneklem için üç farklı dağılım türü (pozitif çarpık (ÇK = +0,4, Basıklık Katsayısı (BK) = -0,1), negatif çarpık (ÇK = -0,4, BK = -0,1) ve basık dağılım (ÇK = +0,1, BK = -0,4)) kullanmıştır. Bu dağılımları elde etmek için iki normal dağılımı birleştirmiştir. Bütün analizlerini IBM 370 bilgisayarında DOS/VSE işletim sisteminde VS FORTRAN derleyicisi kullanarak BILOG ve LOGIST yazılımlarında gerçekleştirmiştir. Elde ettiği sonuçlara göre BILOG yazılımı, Maximum Likelihood Estimation (MLE) kestirim yöntemi kullanıldığı durumlarda 20 ve 40 maddelik testler için LOGIST yazılımına göre

%25 daha hızlı işlem gerçekleştirmiştir. Bayesian expected a priori (EAP) kestirim yöntemi θ'ları kestirmek için kullanıldığında 20 maddelik testlerde iki yazılımda hemen hemen aynı işlem sürelerini kullanırken 40 maddelik testler için BILOG LOGIST yazılımına göre %40 daha hızlı işlem gerçekleştirmiştir. Ancak kısmi ihlallere izin verildiği durumlarda LOGIST yazılımı BILOG'a göre daha iyi performans göstermiştir. Yen (1987) çalışmasının devamında elde ettiği parametre kestirim keskinliği değerlerinden genel olarak BILOG yazılımının LOGIST yazılımına göre daha hassas tahminler yaptığı, MLE kestirim yöntemi kullanıldığında MKE'lerin BILOG ve LOGIST yazılımlarında neredeyse aynı olduğunu, ihlaller analiz edildiğinde işlem süresi dışında sonuçların her iki yazılım için de hemen hemen aynı olduğu sonuçlarına ulaşmıştır.

Drasgow (1989), gerçekleştirdiği çalışmasında MML kestirim yönteminin 2PL modelde çeşitli koşullar altında (madde sayısı = 5,10,15,25, örneklem büyüklüğü = 200, 300, 500, 1000.) parametre kestirimlerinin hassasiyetini gözlemlemeyi amaçlamıştır. Bu amaç doğrultusunda madde ve θ parametrelerini Drasgow ve Hulin (1987)'in çalışmalarından elde etmiştir. Madde parametrelerinin kestiriminde kullandığı MML yöntemine ek olarak Joint Maximum Likelihood (JML) yöntemini de MML ile karşılaştırma amacı güderek kullanmıştır. Her bir faktör için 10 replikasyon gerçekleştirmiştir ve LOGIST yazılımında kestirim işlemi için en fazla 40 iterasyona izin vermiştir. Analizlerden elde ettiği sonuçları değerlendirmek içinse Bias istatistiğine ve gerçek MKE ile kestirilen parametre değerlerinden elde edilen madde karakteristik eğrileri arasındaki ortalama uzaklık derecelerini kullanmıştır. Çalışmasının sonunda MML kestirim yönteminin JML'den daha hassas sonuçlar verdiği, aşırı değerlerdeki madde parametrelerinin kullanıldığı durumlarda kestirimlerin de daha yüksek yanlılık ve standart hata (SH) ürettiği ve simülasyon çalışmalarında gözlenmek istenen keskinlik derecesinin madde parametrelerinin değerlerinden etkilendiği sonuçlarına ulaşmıştır.

Seong (1990), çalışmasında önsel (prior) θ dağılımlarının gerçek θ dağılımlarıyla uyuşmadığında MML kestirim yönteminin madde ve θ parametrelerinin kestirimindeki hassasiyetini öğrenmeye çalışmıştır. Bu amaçla 100 ve 1000 örneklem büyüklüklerinde, normal dağılıma sahip bir θ dağılımının yanında ÇK -1,0 ve 1,0 olan 3 farklı θ dağılımı kullanarak kırk beşer maddeden oluşan 30 adet veri seti oluşturmuştur. Oluşturulan bu veri setlerini 2 parametreli normal ogive model altında PC-BILOG 1.1 yazılımı ile analiz etmiştir. Yapılan analizleri değerlendirmek içinse RMSE ve ortalama mutlak farklılık (AAD) istatistikleri kullanmıştır. Elde edilen sonuçlarda Seong (1990) a, b ve θ parametreleri için genellikle eşleşen dağılımların eşleşmeye dağılımlara göre daha küçük RMSE ve Bias değerlerine sahip olduğunu, örneklem büyüklüğü arttıkça RMSE ve Bias değerlerinin küçüldüğünü gözlemlemiştir.

Zwinderman ve Wollenberg (1990), gerçekleştirmiş oldukları çalışmada Rasch modelinde θ dağılımının yanlış belirlendiği durumlarda parametre kestiriminde kullanılan MML yönteminin hassasiyetini belirlemeye çalışmışlardır. MML yönteminin performansını gözlemleyebilmek için aynı verileri Conditional Maximum Likelihood (CML) yöntemi ile de kestirime tabii tutmuşlar ve sonuçları karşılaştırmışlardır. Bu amaçla 5, 10 ve 15 maddelik testler hazırlamışlardır ve örneklem büyüklüğünü 4,000 olarak belirlemişlerdir. Daha sonra araştırmacılar exponential (üssel) dağılım kullanarak ortalamaları sırasıyla 1, 5 ve 10 olan θ parametrelerinin dağılımlarını üretmişlerdir. Bu parametreler ÇK=-2,01, ÇK=-0,80 ve ÇK=-0,15 olmak üzere sırasıyla aşırı, orta ve hafif çarpık olarak adlandırmışlardır. Sonuç olarak θ dağılımı gerçekte çarpıkken eğer normal dağıldığı varsayılarak analizler gerçekleştirilirse MML yönteminin kestirim hassasiyetini kaybettiğini ve CML yönteminin bu gibi durumlarda daha iyi sonuçlar verdiğini tespit etmişlerdir.

Stone (1992), yapmış olduğu simülasyon çalışmasında 2PL model altında MML kestirim yöntemi kullanarak MULTILOG yazılımını değerlendirmeye çalışmıştır. Bu amaç doğrultusunda test uzunluğu (10, 20, 40 madde), örneklem büyüklüğü (250, 500, 1000) ve θ dağılımı (normal dağılım, pozitif çarpık (ÇK = 0,75, BK = 0,0), basık (ÇK = 0, BK = -1,0) değişkenlerini manipüle ederek parametre kestirim keskinliğini incelemiştir. 27 simülasyon faktörünün her biri için GENIRV yazılımını kullanarak 100 replikasyon yapılıp bir MC çalışması elde etmiştir. Veri üretimi için 2PL modele göre madde parametreleri kalibre edilmiş 20 maddeden oluşan bir matematik başarı testi kullanmış, bunlardan çift numaraya sahip maddeler seçilerek 10 maddelik bir test oluşturmuş ve 40 maddelik testi oluşturmak için 20 maddelik test tekrar kullanmıştır. Çarpık ve basık θ dağılımlarını elde etmek içinse Fleishman (1978)'ın güç yöntemini kullanmıştır. Parametre kestirimlerinin keskinliğini değerlendirmek amacıyla Stone (1992) çalışmasında RMSE ve Bias istatistiklerinden yararlanmış ve iterasyon sayısını (kestirim tekrar sayısı) raporlamıştır. Çalışmanın sonucunda θ'nın normal dağılmadığı durumlarda iterasyon sayısının küçük miktarlarda arttığını gözlemlemiştir. a parametresi için dağılımın çarpık ve basık olduğu durumlarda test uzunluğu arttıkça Bias ve RMSE değerlerinin küçüldüğünü, b parametresi için Bias istatistiğinin test uzunluğunun ve örneklem büyüklüğünün değerlerine bakılmaksızın neredeyse sıfıra yakın olduğu ancak çarpık θ dağılımının normal ve basık dağılımlarla karşılaştırıldığında hatırı sayılır miktarda daha yüksek değerlere sahip olduğunu gözlemlemiştir.

Abdel-fattah (1994), araştırmasında 3PL model üzerinde BILOG ve LOGIST yazılımlarını karşılaştırmayı amaçlamıştır. Bu amaç doğrultusunda 20 ve 60 maddelik test uzunluklarının yanında N=250 ve N=1,000 örneklem büyüklüklerini kullanan araştırmacı madde ayırt edicilik (a) parametrelerini lognormal, madde güçlük (b) parametrelerini normal ve şans (c) parametrelerini ise beta dağılımından üretmiştir. Araştırmacı daha sonra θ parametrelerini normal, truncated normal ve beta dağılımlarından üretmiştir. Böylece test uzunluklarını, örneklem büyüklüklerini ve θ parametrelerinin dağılımlarını manipüle eden araştırmacı araştırmasının amacı doğrultusunda JML kestirim yöntemini LOGIST yazılımında, MML ve marginal Bayesian (MB) kestirim yöntemlerini BILOG yazılımında kullanarak elde ettiği sonuçları sapmanın ortalama karesi (mean square deviation-MSD) istatistiği yardımıyla değerlendirmiştir. Çalışmanın sonucunda MB kestirim yönteminin θ parametrelerinin normal dağılıma sahip olduğu veri setlerinde özellikle küçük test uzunlukları ve örneklem büyüklüklerinde daha keskin sonuçlar verdiğini belirten araştırmacı, genellikle MB kestirim yönteminin madde parametre kestirim işleminde daha keskin sonuçlar verdiğini gözlemlemiştir.

Yoes (1995), araştırmasında 3PL model üzerinde XCALIBRE yazılımından elde LOGIST) karşılaştırmayı amaçlamıştır. Bu amaç doğrultusunda araştırmacı verilerini Wickman ve Hill (1982)’in belirlediği prosedürlere göre gerçekleştirmiştir. Araştırmada 250, 500, 1,000 ve 2,000 örneklem büyüklükleriyle beraber 15, 25, 50, 75 ve 100 maddeden oluşan test uzunlukları kullanılmıştır. Araştırmacı θ parametrelerinin dağılımını standart normal dağılımda sabit bırakmıştır. Test koşullarını da incelemek isteyen Yoes (1995) bu amaçla iki ayrı koşul oluşturmuştur. İlk koşulda b parametrelerini standart normal dağılımdan, c parametrelerini ortalaması 0,25 ve standart sapması 0,5 olan normal dağılımdan, a parametrelerini ise büyük bir üniversiteden Psikolojiye Giriş adlı kursunda uygulanan sınavdan ortalaması 0,75 ve standart sapması 0,1 olacak şekilde elde etmişti. Daha sonra a parametrelerinin bu dağılımını *orta ayırt edici* olarak adlandırmıştır. İkinci koşulda ise b ve c parametreleri ilk koşulla aynı olmak üzere a parametrelerini Silahlı Hizmetler Mesleki Yetenek Bataryasından bir Genel Bilim testinden ortalaması 1,50 ve standart sapması 0,2 olacak şekilde elde etmiştir. Parametre kestirimi içim MML kestirim yöntemini kullanan araştırmacı analizlerinden elde ettiği sonuçları RMSE istatistiği ve üretilen parametreler ile kestirilen parametreler arasındaki korelasyon katsayıları yardımıyla değerlendirmiştir. Ayrıca araştırmacı değerlendirme üzerinde gereksiz bir etki bırakmaması için aşırı uç kestirim sonuçlarını veri setinden çıkarmıştır. Araştırmacı çalışmasının sonucunda XCALIBRE yazılımının küçük örneklem büyüklükleri ve kısa test uzunluklarında BILOG kadar iyi sonuçlar verdiğini ve bu gibi durumlarda kullanılabilecek en iyi yazılımlar olduğunu belirtmiştir. Örneklem büyüklüğü 500'ü ve test uzunluğu 50 maddeyi geçtiğinde ise yazılımlar arasında çok ufak farklar meydana geldiğini gözlemlemekle beraber XCALIBRE yazılımının BILOG yazılımına alternatif olabileceğini belirtmiştir.

Boulet (1996), parametre kestiriminde full-information (FI) ve limited-information (LI) metotlarının kullanıldığı durumlarda θ parametrelerinin çarpık dağılımının parametre kestirim keskinliğine etkisini incelemek istemiştir. Bu amaç doğrultusunda 2PL model üzerinde çalışan araştırmacı FI metodu için TESTFACT yazılımında MML kestirim yöntemini, LI metodu içinse NOHARM II yazılımında Unweighted least-squares (ULS) kestirim yöntemini kullanmıştır. Araştırmada 15, 30, 45 ve 60 maddelik testlerin yanında 250, 500, 1,000 ve 10,000 örneklem büyüklükleri kullanılmıştır. Araştırmacı θ parametrelerini M2PLGEN yazılımından standart normal dağılım ve serbestlik derecesi sırasıyla 3 ve 8 olan $\chi^2$ dağılımından üretmiştir. Tüm dağılımların ortalamasını 0 ve standart sapmasını 1 olarak düzenlemiştir. Araştırmacı serbestlik derecesi 8 olan dağılımda ÇK= 1,00, 3 olan dağılımda ise ÇK= 1,75 olarak tespit etmiştir. Daha sonra madde ayırt edicilik (a) parametrelerini üreten araştırmacı bu parametreleri 0,5, 1,0 ve 1,5 olarak ayarlamış ve testin barındırdığı madde sayısına eşit olarak dağıtmıştır. Madde güçlük parametrelerini (b) -2, -1, 0, 1, 2 olarak ayarlayan araştırmacı benzer dağıtım işlemini b parametreleri için de uygulamıştır. Ayrıca araştırmacı b parametrelerinin düzeylerine göre de kestirim keskinliğini değerlendirmiştir. Araştırmacı her bir koşul için 100 replikasyon işlemi gerçekleştirmiştir. Elde ettiği sonuçları değerlendirmek içinse RMSE istatistiğinden yararlanmıştır. Boulet (1996), analizler sonucunda θ dağılımının normal olduğu durumlarda b parametrelerinin kestirim keskinliğinin LI metodunda daha iyi sonuçlar ürettiği ancak çarpık dağılımın söz konusu olduğu durumlarda ise FI metodunda daha iyi sonuçlar elde ettiğini ifade etmiştir. Ayrıca test uzunluğunu sabit tuttuğunda örneklem büyüklüğü artışının RMSE değerlerinde küçülmeyi sağladığını tespit etmiştir. a parametreleri içinse b parametrelerinden elde edilen sonuçlara benzer sonuçlar elde etmiştir.

Kirişçi, Hsu ve Yu (2001), yaptıkları çalışmada MTK'nin tek boyutluluk ve normallik varsayımlarının karşılanamadığı durumlar için MTK alanında kullanılan BILOG, MULTILOG ve XCALIBRE kestirim yazılımlarının hassasiyetinin gözlemlenmesini amaçlamıştır. Bu amaç için 3PL model üzerinden madde ve yetenek parametrelerini üretmişlerdir. Madde parametrelerini, a parametreleri için 0,4 ile 2,0, b parametreleri için -2,0 ile 2,0 ve c parametreleri için ise 0,0 ile 0,3 arasında olmak üzere tümünde uniform dağılım kullanarak üretmişlerdir. θ parametrelerini ise ortalaması 0,00 ve standart sapması (SS) 1,00 olmak üzere normal dağılım, pozitif çarpık (ÇK = 0,75, BK = 0,00) dağılım ve basık (ÇK = 0,00, BK = -1,00) dağılım olacak şekilde kullanmış, çarpık ve basık dağılımları Fleishman'ın güç yöntemini ile oluşturmuşlardır. Çalışmalarında tüm veri setlerini örneklem sayısı 1000 ve madde sayısını 40 olacak şekilde sabitlemişlerdir. Parametre kestirim işlemi için MML yöntemi kullanmışlar, parametre kestirim keskinliğini değerlendirmek içinse RMSE değerlerini ve RMSE değerlerinin logaritmik dönüşümü yapılıp üç yönlü varyans analizinden (ANOVA) geçirilmesiyle elde edilen sonuçları kullanmışlardır. Bu sonuçların ışığında θ dağılımının ve dağılımın diğer değişkenlerle etkileşiminin parametre kestirim keskinliğine etkisinin $p<0.001$, $\eta^2 >0.09$ anlamlılık düzeylerinde manidar olmadığı sonucuna ulaşmışlardır.

Sass, Schmitt ve Walker (2008), yaptıkları araştırmada 2PL model altında çarpık dağılan θ ve zorluk parametrelerinin parametre kestirim keskinliğine nasıl etki ettiğini bulmak için otuzar maddeden oluşan 3 veri seti oluşturarak BILOG-MG yazılımında çalışmıştır. Bu veri setlerinde b parametrelerinin ortalamaları ve SS'si 1'e sabitlenerek normal ve gamma dağılımlarından rastgele üretilmiştir. Negatif çarpık b parametrelerini üretmek içinse gamma dağılımından üretilen b parametreleri -1 ile çarpılarak işaretleri değiştirilmiştir. a parametresi ise her 3 veri seti için de sabit olmak üzere uniform dağılımdan rastgele üretilmiştir ve orta düzeyde bir ayırt edicilik elde etmek için minimum 0,9 ile maksimum 2,1 değerleri arasındadır. θ parametrelerinde ise ortalama 0 ve standart sapma 1 olmak üzere normal dağılımın yanında ortalama 0, SS 1, ÇK 1,6 ve basıklık katsayısı (BK) 4,1 olan gamma dağılımından rastgele üretilen pozitif çarpık bir dağılım kullanılmıştır. Bu veri setlerinin kestiriminde MML, Maximum a Posteriori (MAP), Version 1 of Expected A Posteriori (EAP1) ve Version 2 EAP (EAP2) olmak üzere 4 farklı yöntem kullanılmıştır. Parametre kestirim keskinliğini değerlendirmek içinse RMSE ve AAD istatistikleri kullanılmıştır. Yapılan analizlerin sonucunda a parametresinin b parametresine göre özellikle büyük ayırt edicilik değerlerinde çarpık dağılımlardan önemli ölçüde etkilendiği, θ parametresinin kestirimde ise kullanılan kestirim yöntemlerinden bağımsız olarak üretilen madde parametreleri de kullanılsa kestirilen madde parametreleri de kullanılsa θ dağılımı çarpık olduğu sürece parametre kestiriminin normal dağılımda daha sağlıklı olduğu tespit edilmiştir.

Köse (2010), araştırmasında tek ve çok boyutlu MTK modelleri altında farklı örneklem büyüklükleri ve test uzunluklarında hangi modelin daha iyi ölçme keskinliği değerleri vereceğini incelemeyi amaçlamıştır. Çalışmasında gerçek verilerden yararlanan araştırmacı kullandığı verileri 1516 öğrencinin cevapladığı ilköğretim 8. sınıf öğrencileri için hazırlanmış 24 maddelik bir Türkçe testinden elde etmiştir. Araştırmacı çalışmanın amacı doğrultusunda 12 ve 24 maddelik test uzunlukları ve 500, 1,000 ve 1,500 örneklem büyüklüklerin kullanmıştır. Tek boyutlu MTK modellerinden 2PL modeli, çok boyutlu MTK modellerinden ise telafisel tipte çok boyutlu MTK modelini kullanan araştırmacı verilerini SPSS 13.0 ve Excel yazılımlarında analiz ederken madde parametrelerinin kestirimi için NOHARM, yetenek parametrelerinin kestirimi içinse TESTFACT yazılımını kullanmıştır. Köse (2010) elde ettiği sonuçları değerlendirmek içinse artık varyansların kareleri toplamı (sum of squares of residuals (SSR)), artıkların ortalama karelerinin karekökü (root mean square of residuals (RMSR)) ve Tanaka uyum indeksi istatistiklerinden yararlanmıştır. Araştırmacı çalışmasının sonunda tek boyutluluk varsayımının ihlal edildiği durumlarda çok boyutlu MTK modelinin daha iyi keskinlik değerleri ürettiği, farklı test uzunluklarında çok boyutlu MTK modellerinin daha iyi keskinlik değerleri ürettiği, örneklem büyüklüğündeki değişimin tek boyutlu MTK modelinde belirgin bir etkisinin görülmediği ancak çok boyutlu MTK modelinde örneklem büyüklüğündeki artışın daha iyi keskinlik değerleri verdiği sonuçlarına ulaşmıştır.

Xu ve Jia (2011) çalışmalarında EM algoritmasının normal dağılmayan verilerde uygulanabilirliği, farklı θ dağılımlarında madde parametre kestirimlerinin keskinliğini, farklı θ dağılımlarında betimsel istatistiklerin tutarlılığını incelemeyi amaçlamışlardır. Bu amaç doğrultusunda araştırmacılar Rasch ve 2 PL modelde θ parametrelerinin generalized skew normal (GSN) ve normal dağılıma sahip olduğu durumları incelemişlerdir. Test uzunluğunu 30 madde, örneklem büyüklüğünün ise 1,000 olarak belirleyen araştırmacılar madde güçlük (b) parametrelerini standart normal dağılımdan, madde ayırt edicilik parametrelerini (a) ise uniform dağılımdan minimum 0,5 ve maximum 2,5 olacak şekilde üretmişlerdir. θ parametreleri ise normal dağılımda ortalaması -0,5, standart sapması ise 1 olacak şekilde ayarlanmış; GSN dağılımı kullanılırken ise lokasyon (location), ölçek (scale) ve çarpıklık (skewness) parametreleri sırasıyla SN(-0,5, 1, -1), SN(-0,5, 1, 3) şeklinde ayarlanarak yetenek parametreleri üretilmiştir. Araştırmacılar her bir veri seti için 60 replikasyon uygulamışlardır. Çalışmanın sonucunda araştırmacılar Rasch modelinde θ parametrelerinin dağılımı yanlış tanımlansa bile kestirimlerin bu durumdan çok fazla etkilenmediğini, 2PL modelde de benzer durumlar olduğunu ancak burada ÇK arttıkça kestirimlerin daha fazla etkilendiğini bulmuşlardır.

Kieftenbeld ve Natesan (2012), çalışmalarında Dereceli Tepki Modelinde (Graded Response Model (GRM)) θ parametrelerinin dağılımı, test uzunluğu ve örneklem büyüklüğü değişkenlerini manipüle ederek MML/EAP ve Markov Chain Monte Carlo (MCMC) kestirim yöntemlerinin keskinliğini karşılaştırmayı ve OpenBUGS yazılımının performansını gözlemlemeyi amaçlamışlardır. Bu doğrultuda θ parametreleri için 3 farklı dağılım türü (normal, skew-normal (ÇK= 1,25, BK= 1,5) ve uniform), 5 farklı örneklem büyüklüğü (75, 150, 300, 500 ve 1,000) ve 4 farklı test uzunluğu (5, 10, 15, 20) kullanmışlardır. Araştırmacılar θ parametrelerini standart normal dağılım $N(0,1)$, -3 ve 3 aralığında veriler üreten uniform dağılım ve Fleishman'ın metodunu uygulayarak ÇK= 1,25, BK= 1,5 olan skew-normal dağılımdan elde etmişlerdir. Her veri setinde ortak olarak kullanılan madde parametrelerini ise Harwell, vd. (1996) çalışmasından almışlardır. Çalışmada sonuçları değerlendirmek içinse RMSE ve Bias istatistiklerindeki varyans gözlemlenmiştir. Araştırmacılar MML kestirim yönteminden sonuç almak için MULTILOG yazılımını, MCMC kestirim yönteminden sonuç almak için OpenBUGS yazılımını kullanmışlar ayrıca θ parametrelerinin kestirimini ekstra olarak MML/EAP yöntemini Wolfram (2011)'ın Mathemtica kodu yoluyla da incelemişlerdir.

Araştırmacılar madde parametre kestiriminde örneklem büyüklüğünün parametre kestirim keskinliğine etki eden en önemli değişken olduğu ancak örneklem büyüklüğünün belli bir boyuta ulaştıktan sonra çok az etki gösterdiği, test uzunluğunun madde parametre kestirim keskinliğine önemli bir etkisinin olmadığı, θ parametrelerinin dağılım türlerinin madde parametre kestiriminin keskinliğine ufak negatif etkilerde bulunduğu, N= 300 örneklem büyüklüğü ve yukarısında MML/EAP ve MCMC kestirim yöntemlerinin benzer sonuçlar ürettiği ve daha küçük örneklem büyüklüklerinde bazı parametrelerde MCMC kestirim yönteminin daha iyi kestirim değerleri verdiği sonuçlarına ulaşmışlardır. θ parametrelerinin kestirimlerinde ise keskinliğe en büyük etkiyi gösteren değişkenin test uzunluğu olduğunu tespit etmekle beraber MML/EAP ve MCMC kestirim yönteminin MULTILOG yazılımından elde edilen sonuçlardan daha iyi olduğunu belirtmişlerdir.

Akour ve AL-Omari (2013), araştırmalarında örneklem büyüklüğü ve test uzunluğunun 3PL modelde madde parametre kestirim keskinliğine etkisini gözlemlemeyi amaçlamışlardır. Bu amaç doğrultusunda araştırmacılar Ürdün Eğitim Bakanlığının 8. sınıflar için uyguladığı bir matematik sınavından elde ettikleri gerçek verileri kullanmışlardır. Matematik testi 60 maddeden oluşmaktadır ve bu testi 40,000 birey cevaplamıştır. Araştırmacılar bu veri setinden 3 farklı test uzunluğu (15, 30, 60) ve 6 farklı örneklem büyüklüğü (200, 500, 1,000, 5,000, 10,000, 20,000) elde etmişlerdir. Araştırmacılar gerçek verilerin parametre değerlerini BILOG programından elde ettikten sonra veri setlerini istedikleri şekilde organize ederek tahmin edilen parametreler için de BILOG programını kullanmışlardır. Analiz sonuçlarını root mean square loss (RMSL) istatistiğinden yararlanarak değerlendirmişler ayrıca her bir testin kestirilen ve gerçek durumlarından elde edilen test bilgi fonksiyonunu birbirlerine bölerek göreceli verimlilik (relative efficiency (RE)) katsayısını da değerlendirmeye almışlardır. Çalışmanın sonucuna göre araştırmacılar örneklem büyüklüğündeki değişimin en çok b parametresini etkilediğini, örneklem büyüklüğündeki ve test uzunluğundaki artışın parametre kestirim keskinliğini arttırdığını bulmuşlardır.

Olmuş, Nazman ve Erbaş (2016), gerçekleştirdikleri çalışmalarında çeşitli koşullar altında (örneklem büyüklüğü, madde sayısı; yüksek, düşük ve normal θ düzeylerine sahip dağılımlar, yüksek ve düşük düzeydeki a parametre değerleri) 2PL modelde JML kestirim yönteminin kullanılmasının parametre kestirim hassasiyetinde ne gibi etkilere sahip olduğunu gözlemlemeyi amaçlamışlardır. Bu amaç doğrultusunda araştırmacılar 20, 40, 60, 90, 120 ve 150 maddeden oluşan veri setlerinin yanı sıra 100, 250, 500, 1000 ve 5000 kişilik örneklem grupları oluşturmuşlardır. b parametresini normal dağılımdan (N(0,1)); yüksek ayırt edicilik düzeyine sahip a parametrelerini ortalaması 0, SS = 2 olan uniform dağılımdan; düşük ayırt edicilik düzeyine sahip a parametrelerini ortalaması 0, SS = 0,2 olan uniform dağılımdan üretmişlerdir. θ dağılımları normal dağılımdan üretilmiş ancak θ düzeylerini ayarlayabilmek için dağılımların ortalamalarını değiştirmişlerdir. Normal yetenek düzeyi için N(0,1) kullanmışlarken yüksek θ düzeyi için N(1,1), düşük θ düzeyi içinse N(-1,1) değerlerini kullanmışlardır. JML kestirim yöntemiyle alınan sonuçlar test bilgi fonksiyonundan (TBF) elde edilen bilgi miktarları ile nokta çift serili korelasyon yöntemi kullanılarak kestirilen parametreler ve gerçek parametrelerin korelasyonuna bakarak değerlendirmişlerdir. Yapılan analizler ışığında önceden belirlenmiş yüksek a parametrelerinin olduğu düşük örneklem büyüklüklerinde madde sayısı arttıkça TIF'in de artış gösterdiği, düşük a parametrelerinde ise bu durumun tamamen zıt bir sonuç verdiği sonuçlarına ulaşmışlardır. Ayrıca önceden belirlenen düşük a parametrelerinde madde sayısı azalırken kestirilmiş ve gerçek parametreler arasındaki korelasyonun artma eğiliminde olduğunu tespit etmişlerdir. Araştırmacılar, önceden belirlenmiş yüksek a parametrelerinde θ dağılımı düşük düzeyli bireylerden oluştuğunda tüm örneklem büyüklüğü ve madde sayıları için b parametreleri en yüksek değerlere sahip olduğunu gözlemlemişlerdir.

Karadavut (2017), çalışmasında Rasch, 2PL ve 3PL model altında θ uniform dağılıma sahipse parametre kestirim keskinliğinin nasıl etkilendiğini incelemeyi amaçlamıştır. Bu amaç doğrultusunda θ parametrelerini normal dağılım ve uniform dağılım gösterecek şekilde R yazılımında üretmiştir. Madde parametrelerini ise a ve b parametresi normal dağılımdan c parametresi ise beta dağılımından üretmiştir. Bununla beraber çalışmasında 15 ve 30 maddeden oluşan iki test uzunluğunda, 600 ve 2,000 kişilik iki örneklem büyüklüğünde çalışmasını sürdürmüştür. Üretilen parametrelerle birlikte kullanılan modellerden parametre kestirimi yapmak için OpenBUGS yazılımında MCMC kestirim yöntemini, elde ettiği parametre kestirimlerinin keskinliğini değerlendirmek için Bias, RMSE, ortalama mutlak hata (MAE) ve ortalama karesel hata (MSE) istatistiklerini kullanmıştır. Her bir simülasyon faktörü için 25 veri setini simüle etmiştir ve parametre kestiriminde toplam 30000 iterasyon kullanmıştır. Çalışmanın sonuçlarına göre 2PL model için θ dağılımı uniform varsayıldığında asıl dağılımın ne olduğuna bakılmaksızın diğer kombinasyonlardan daha keskin sonuçlar vermekte, 3PL model için θ dağılımı uniform varsayıldığında asıl dağılım da uniform ise diğer kombinasyonlara göre daha keskin sonuçlar vermekte ancak Rasch modeli için ayırt edici herhangi bir sonuç elde edilemediğini söylemektedir.

Kim ve Lee (2017) araştırmalarında üç faktörün (θ dağılımları, sayısal integrasyon (numerical integration) ve madde parametrelerinin kestirimi için referans çerçevesi (θ dağılımları için önsel (prior) ve ardsal (sonsal, posterior) dağılım) ) ) 3PL modelde madde parametre kestirimlerinin keskinliğine etkisini incelemişlerdir. Bu amaca ek olarak beş farklı madde kalibrasyon yönteminden (Normal-Midpoint-Prior (NMPr), Normal-Hermit-Prior (NHPr), Normal-Midpoint-Posterior (NMPo), Normal-Hermit-Posterior (NHPo) ve Empirical-Midpoint-Prior (EMPr) ) ve dört farklı yazılımdan (BILOG-MG, PARSCALE, flexMIRT ve ICL) elde ettikleri sonuçları da karşılaştırmışlardır. Araştırmacılar bu amaç doğrultusunda yapay verileri üretmek için bir lise matematik sınavından elde ettikleri 40 maddeden oluşan ve 1,000 örneklem büyüklüğü bulunan veri setini kullanmışlardır. Bu veri setinin yardımıyla simülasyon çalışmalarında N= 500 ve N= 3000 örneklem büyüklüğünde veri setlerini kullanmışlardır. θ parametrelerinin dağılımlarını ise standart normal dağılımdan ve skew-normal dağılımdan ( SN(0, 1, -2) ve SN(0, 1, -5) ) üretmişlerdir. Çarpık dağılımların ÇK'leri ise sırasıyla -0,45 ve -0,85'tir. Araştırmacılar elde edilen sonuçları değerlendirmek içinse ortalama Bias ve RMSE istatistiklerini kullanmışlardır. Araştırmanın sonunda θ dağılımı standart normal dağılıma sahipken bütün kalibrasyon yöntemlerinin düşük Bias değerleri verdiği, EMPr kalibrasyon yöntemi hariç diğer yöntemlerde θ parametreleri normal dağılımdan saptıkça madde parametrelerinin kestirimlerinin keskinliğinin düştüğünü ve ÇK arttıkça bu yöndeki eğilimin de arttığı, EMPr kalibrasyon yönteminin ise θ parametrelerinin normal dağılmadığı durumlarda diğer dört kalibrasyon yönteminden daha iyi değerler verdiği sonuçlarına ulaşmışlardır.

Svetina, Valdiva, Underhill, Dai ve Wang (2017) çalışmalarında çok boyutlu MTK modellerinde maddelerin karmaşıklık derecesinin ve θ dağılımlarının normal olmadığı durumların madde parametre kestirim keskinliğinin nasıl etkilediği sorusuna cevap aramışlardır. Bu amaç doğrultusunda araştırmacılar Ulusal Eğitim İlerleme Değerlendirilimesi (NAEP) kapsamında 4. sınıflar için hazırlanmış 62 dikotom (1-0 şeklinde puanlanan) maddeden rastgele biçimde 30 madde seçmişlerdir. Çalışma için 2,000 örneklem büyüklüğü belirleyen araştırmacılar analizlerini ikişer boyuta sahip olan 2 ve 3 parametreli normal ogive modelleri üzerinde gerçekleştirmişlerdir. Araştırmacılar boyutlar arasındaki korelasyonları da sırasıyla 0,00, 0,40, 0,70 olmak üzere 3 farklı düzeyde incelemişlerdir. Maddelerin karmaşıklığını 5 farklı düzeyde inceleyen araştırmacılar θ dağılımlarını incelemek içinse θ çiftlerinin dağılımını normal-normal dağılım, normal-çarpık dağılım ve çarpık-çarpık dağılım olarak manipüle etmişlerdir. Araştırmacılar aynı zamanda madde ayırt edicilik parametrelerini (a) iki gruba ayırmış ve bu gruplardan birini *dengeli* diğerini ise *dengesiz* olarak adlandırmışlar ve madde güçlük parametrelerini (b) ise ortalaması -0,169, standart sapması 0,774 olacak şekilde üretmişlerdir. b parametrelerinin minimum değerini -1,318, maksimum değerini ise 1,204 olarak tespit etmişlerdir. Araştırmacılar analizlerini yapmak için R programlama dilinde *mirt* paketinden faydalanmış, kestirim için EM algoritmasını kullanmışlardır. Her bir koşul içinse 500 replikasyon gerçekleştirmişlerdir. Elde ettikleri sonuçları değerlendirmek içinse Bias ve RMSE istatistiklerinden yararlanmışlardır. Araştırmanın sonunda genellikle θ çiftlerinin çarpıklığı arttıkça, a parametreleri *dengesiz* konumda oldukça ve karmaşıklık arttıkça parametre kestirim keskinliğini azalma eğiliminde olduğu sonucuna ulaşmışlardır.

Şahin ve Anıl (2017) çalışmalarında test uzunluğunun ve örneklem büyüklüğünün madde parametre kestirim keskinliğine etkisini incelemek istemişlerdir. Bu amaç doğrultusunda bir dil testinden elde ettikleri gerçek verileri kullanmışlardır. Kullanmış oldukları test 50 maddeden oluşmaktadır ve bu testi 6,288 birey cevaplamıştır. Araştırmacılar bu verilerden yararlanarak istenilen özellikte veri setleri oluşturmuşlardır. Bu veri setleri 10, 20 ve 30 maddeden oluşan testlerden ve 150, 250, 350, 500, 750, 1,000, 2,000, 3,000 ve 5,000 örneklem büyüklüklerinden oluşmaktadır. Çalışmada sırasıyla tek boyutlu dikotom 1PL, 2PL ve 3PL modeller kullanılmıştır. Verileri analiz etmek için Xcalibre 4.1 yazılımından yararlanan araştırmacılar parametre kestirimi için MMLE kestirim yöntemini kullanmışlardır. Araştırmacılar kestirim işleminden elde edilen sonuçları değerlendirmek içinse gerçek parametre değerleri ile kestirilen parametreler arasındaki korelasyon katsayılarından ve root-mean-squared difference (RMSD) istatistiğinden yararlanmışlardır. Araştırmanın sonunda 1PL modelde kestirilen b parametreleri gerçek parametreler ile son derece yüksek pozitif korelasyon gösterdiğini, genellikle örneklem büyüklüğü ve test uzunluğu arttıkça korelasyon katsayısının arttığını bulmuşlardır. Elde edilen RMSD istatistiklerine baktıklarında ise sonuçlar dalgalı bir grafik çizmekte olduğunu ancak genel olarak örneklem büyüklüğünün ve test uzunluğunun artmasının RMSD değerlerini düşürdüğünü tespit etmişlerdir. 2PL modelde a parametrelerinin kestiriminde test uzunluğunun düşük olması korelasyon katsayısının da düşük olmasına sebep olduğu ancak örneklem büyüklüğü arttıkça test uzunluğunun etkisini yitirdiğini yine a parametreleri için RMSD değerlerine baktıklarında grafiğin yer yer dalgalandığını buna rağmen genellikle test uzunluğunun ve örneklem büyüklüğünün artmasının RMSD değerlerinin de düşmesini sağladığını bulmuşlardır. 2PL modelde b parametrelerinin kestirimlerine baktıklarında her test uzunluğu ve örneklem büyüklüğünde son derece yüksek pozitif korelasyon katsayıları elde etmişlerdir. RMSD değerlerine bakıldığında yine dalgalı bir grafik elde etmişlerdir ancak test uzunluğunun artmasının bu dalgalanmayı azalttığını görmüşlerdir. Tüm bu sonuçlara rağmen yine genellikle test uzunluğu ve örneklem büyüklüğü artışının RMSD değerlerini düşürdüğü sonucunu elde etmişlerdir. 3PL modelde ise a ve c parametrelerinde en küçük test uzunluğunda korelasyonun dalgalı olduğunu, b parametresinde son derece yüksek pozitif korelasyon katsayıları elde edildiğini yine test uzunluğu ve örneklem büyüklüğü artışının korelasyon katsayılarında pozitif yönde artış sağladığını gözlemlemişlerdir. RMSD istatistik değerlerine bakıldığında a ve b parametrelerinde test uzunluğunu en küçük olduğu durumda dalgalı bir grafiğin ortaya çıktığını ancak c parametrelerinde son derece stabil bir grafik görmüşlerdir. Tüm parametreler için genellikle örneklem büyüklüğünde ve test uzunluğundaki artışın RMSD değerlerinde azalma eğilimine yol açtığı sonucuna ulaşmışlardır.

Şahin ve Yıldırım (2018), çalışmalarında normal dağılmayan b parametrelerinin, test uzunluğunun, örneklem büyüklüğünün ve normal dağılmayan θ parametrelerinin θ parametre kestirim keskinliğine olan etkilerini incelemişlerdir. Bu amaç doğrultusunda θ parametrelerini 2012 yılında yapılan Seviye Belirleme Sınavından (SBS) elde eden araştırmacılar araştırmanın amacına uygun şekilde dağılımları sağa ve sola çarpık olacak şekilde organize etmişlerdir. Sağa çarpık dağılımın ÇK’sini 1, sola çarpık dağılımın ÇK’sini -1 olarak belirleyen araştırmacılar örneklem büyüklüklerini ise 500, 1,000, 2,500, 5,000 ve 10,000 olarak ayarlamışlardır. Madde güçlük parametrelerinin (b) dağılımı için sırasıyla normal dağılım, uniform dağılım, sağa çarpık ve sola çarpık dağılımları kullanmışlardır. Madde güçlük parametreleri (b) üretmek için WinGen 3 yazılımını kullanan araştırmacılar katsayıları normal dağılımda ortalama=0 ve standart sapma=1 olarak, uniform dağılım için minimum değeri -3 ve maksimum değeri 3 olarak, sağa çarpık dağılımları üretmek için $\alpha=2$ ve $\beta=8$ olarak, sola çarpık dağılımları üretmek içinse $\alpha=8$ ve $\beta=2$ olarak belirlemişlerdir. Test uzunluğunun etkisini görmek içinse 20 ve 30 madde içeren iki test belirlemişlerdir. Araştırmacılar analizlerini MULTILOG 7.03 yazılımında gerçekleştirmişler ve her bir veri seti için 25 replikasyon uygulamışlardır. Analizlerinde Maximum Likelihood Estimation (MLE) kestirim yöntemini kullanan araştırmacılar elde ettikleri sonuçları değerlendirmek içinse AAD ve RMSE istatistiklerinden yararlanmışlardır. Çalışmanın sonucunda araştırmacılar b parametrelerini tüm dağılımları için $\theta$ parametrelerinin sağa ya da sola çarpık olduğu durumlarda örneklem büyüklüğünün RMSE ve AAD değerlerinde çok fazla bir değişime yol açmadığını ancak test uzunluğu arttıkça RMSE ve AAD değerlerinde azalma gözlemlendiğini belirtmişlerdir. Yine $\theta$ dağılımlarının, test uzunluklarının ve örneklem büyüklüklerinin tümünde en iyi kestirim keskinliği sonuçlarının b parametreleri normal dağılırken elde edildiğini bulmuşlardır.

# BÖLÜM III

## 3. Yöntem

Bu bölümde araştırmanın modeli, çalışma grubu, veri toplama aracı, verilerin toplanması ve verilerin analizi başlıkları bulunacaktır.

### 3.1. Araştırmanın Modeli

Bu araştırma, farklı örneklem büyüklüklerinin ve farklı θ dağılımlarının kullanılarak 2PL model altında üretilen madde parametrelerini gerçek parametre değerleri olarak ele alıp daha sonra yapılan kestirim işlemleri sonunda ortaya çıkan madde parametrelerini RMSE ve Bias istatistikleri yardımıyla karşılaştırarak ölçme keskinliği hakkında fikir sahibi olmak amacıyla gerçekleştirilmiştir.

Bu amaç doğrultusunda izlenen adımlar (Meyers ve Grossen, 1974);

1. Bir durumun belli özelliklerinin manipüle edilmesi.
2. Bu manipüle işleminin kontrollü olarak gerçekleştirilmesi.
3. Manipüle işlemi sonunda elde edilen sonuçlar gözlemlenebilmesi ve hakkında veri toplanabilmesi.

olduğundan yapılan araştırma bir betimsel modeldir.

## 3.2. Veri Toplama Deseni

Bu çalışmada veri setleri simülasyon yoluyla oluşturulmuştur. Simülasyon çalışmaları sosyal bilimlerde bir durumun incelenmesinin etik olmadığı, hedeflenen evrenin örneklemlerine ulaşmanın zor, maliyetli ya da imkânsız olduğu gibi durumlarda son derece avantajlıdır. Bilgisayarlar veri analizi alanında 1970'lerde bir devrim yaratarak 1990'larda sosyal bilimler için tamamen yeni bir bakış açısı kazandırmışlardır. Günümüzde ise simülasyon yoluyla çeşitli sosyal modelleri oluşturup farklı durumları test etmek mümkündür (Gilbert, 1999). Bu simülasyon çalışmasında ise 2PL model üzerinde gerçekleştirilen parametre kestirimlerinde örneklem büyüklüğü ve θ dağılımının etkisi sınanmıştır.

Araştırmanın verilerini üretmek için alt problemlere uygun olarak madde ve birey parametreleri oluşturulmuştur. Bu amaç doğrultusunda 2PL modele uygun olarak kullanılan madde sayısı kadar a ve b parametre değerleri, değişen örneklem büyüklüğüne bağlı olarak da θ parametre değerleri üretilmiştir. Yapılan bir simülasyon çalışmasında avantaj gerçek parametrelerin bilinmesi ve kestirim sonucunda elde edilen parametreler ile karşılaştırılabilmesidir (Way vd., 1988). Ancak MC simülasyon çalışmasının sağlıklı bir şekilde gerçekleştirilebilmesi için daha önce bahsedilen şartları sağlamak gerekir.

Bu çalışmada kullanılan MC simülasyon çalışmasının şartlarını yerine getirmek amacıyla madde parametreleri üretilirken gerçek durumların yansıtılmasına dikkat edilmiştir. Bu nedenle a parametresi ortalaması 0 standart sapması 0,5 olan rastgele log-normal, b parametresi ise rastgele standart normal dağılımdan üretilmiştir (Abdel-fattah, 1994; Eser ve Gelbal, 2015; Uysal vd., 2019; Bulut ve Sünbül, 2017; Feinberg ve Rubright, 2016; Riley ve Carle, 2012).

## 3.3. Simülasyon Faktörleri ve Koşulları

Bu araştırmada, çeşitli koşulların parametre kestirimi üzerindeki etkisini belirlemek amacıyla MC simülasyon çalışması yapılmıştır. Hulin vd., 1982; Karadavut, 2017; Stone, 1992; Olmuş vd., 2016; Seong, 1990 araştırmalarında örneklem büyüklüğünün ve θ dağılımının parametre kestirimini etkilediği görülmüş olup bu faktörler simülasyon çalışmasında manipüle edilecek faktörler olarak belirlenmiştir.

MTK'de kullanılan birçok model bulunmaktadır ancak daha önce de bahsedildiği gibi bu modellerin birbirlerine avantajları olduğu kadar dezavantajları da bulunmaktadır. Bu araştırmada 2PL modelin seçilme sebebi ise alanyazında ve bilimsel çalışmalarda en geniş kullanım alanına sahip olan modellerden biri olmasıdır (Hulin vd., 1982).

Araştırmada manipüle edilen değişkenlerden ilki θ dağılımlarıdır. θ dağılımlarının neden normal dağılım dışındaki durumlarda gözlemlenebildiği konusuna daha önce değinilmiştir. Buna ek olarak θ dağılımlarının parametre kestirimlerine etkisi net değildir (Kirişçi, 2001). Ayrıca R programlama dili ile gerçekleştirilen çalışmalar yeterince kapsamlı değildir. Bu nedenlerle araştırmada θ dağılımlarının çarpıklık katsayısı daha önceki çalışmalarda kullanılan dağılımlar da göz önünde bulundurularak geniş bir alanı kapsamak için -2,00, -1,00, 1,00, 2,00 ve standart normal dağılım olmak üzere 5 düzey olarak belirlenmiştir (Stone, 1992; Sass vd., 2008; Yıldırım, 2015; Swaminathan ve Gifford, 1979*;* Reise ve Yu, 1990). Daha sonra her bir durum birbirleriyle karşılaştırılmış böylece 10 farklı durum kontrol edilmiştir.

İkinci faktör olarak ise örneklem büyüklüğü seçilmiştir. Örneklem büyüklüğü daha önce yapılan ölçme keskinliği çalışmalarında belirleyici bir etken olarak öne çıkmıştır ve örneklem büyüklüğü arttıkça ölçme keskinliğinin arttığı görülmüştür. Bu araştırmalardan yola çıkarak 250, 500, 1,000 ve 2,000 kişilik 4 örneklem büyüklüğü seçilmiştir (Drasgow, 1989; Seong, 1990; Hulin vd., 1982*;* Lord, 1968; Swaminathan ve Gifford, 1979*)*.

Sabit tutulan faktörlerden biri olan replikasyon sayısı, MC çalışmalarında manipüle edilen değişkenlerin gözlemlenmek istenen durumlar üzerindeki etkisini daha net görebilmek için yüksek sayılarda tutulması gereklidir. Çalışmanın amacına göre değişebilen bu sayı örneklem büyüklüğünün ve test uzunluğunun yeterli olduğu miktarlarda 100’den az durumlarda iyi sonuçlar verebilmektedir (Harwell vd., 1996). Aynı zamanda çalışmanın karmaşıklığı, kullanılan donanım ve yazılım imkanları da replikasyon sayısı üzerinde etkili olmaktadır (Feinberg ve Rubright., 2016). Elde edilen bulguların ışığında bu araştırmada 100 replikasyon kullanılması uygun görülmüştür.

Sabit olan bir diğer faktör ise test uzunluğu yani bir veri setinde kullanılan madde sayısıdır. İncelenen çalışmalarda farklı test uzunluklarının kullanıldığı görülmüştür. Genel olarak bir testte bulunan madde sayısı eğitimde ve psikolojide sıkça kullanılan testlerde bulunan 20 ile 40 madde arasında değişmektedir. (Hulin vd., 1982; Lord, 1968*;* Yen, 1987; Abdel-fattah, 1994; Stone, 1992; Swaminathan ve Gifford, 1979). Test uzunluğunun çalışmayı etkilememesi ve yalnızca manipüle edilen değişkenlerin gözlemlenmek istenen durumlar üzerindeki etkisinin bilinmek istenmesi sebepleriyle bu çalışmada alanyazından elde edilen veriler ışığında bir parametre keskinliği çalışmasında 2PL model altında 30 maddenin kullanılması uygun görülmüştür (Hulin vd., 1982).

Araştırmada 2PL model kullanılmıştır. Bu modelin seçilme amacı daha önce bahsedildiği gibi diğer modellerden daha tutarlı kestirimler vermesi ve bununla birlikte gerçek durumlara olabildiğince yakın olmasıdır.

Sonuç olarak 5 adet θ dağılımı ve 4 adet örneklem büyüklüğünün kullanıldığı bu çalışmada toplam 5x4=20 koşul incelenmiştir. Araştırma 100 replikasyon sayısıyla gerçekleştirilmiştir. Yani bu 20 koşulun her biri 100 kez tekrarlanmıştır ve bu işlemlerin sonucunda toplam 20x100=2,000 veri seti elde edilmiştir. Tablo 3.1’de simülasyon koşulları gösterilmiştir.

**Tablo 3.1.** Araştırmaya dair manipüle edilen simülasyon koşulları.

| Örneklem Büyüklüğü (N) | Yetenek (θ) Dağılımı (ÇK) |
|---|---|
| N= 250 | -2,00 |
| | -1,00 |
| | 0,00 |
| | 1,00 |
| | 2,00 |
| N= 500 | -2,00 |
| | -1,00 |
| | 0,00 |
| | 1,00 |
| | 2,00 |
| N= 1,000 | -2,00 |
| | -1,00 |
| | 0,00 |
| | 1,00 |
| | 2,00 |
| N= 2,000 | -2,00 |
| | -1,00 |
| | 0,00 |
| | 1,00 |
| | 2,00 |

Not: Tüm koşullarda test uzunluğu 30 maddedir.

## 3.4. Verilerin Üretilmesi

Araştırmada manipüle edilen koşulların parametre kestirimini nasıl etkilediği gözlemlenmek istenmiştir ve bu amaç doğrultusunda çalışma yapay veriler yardımıyla gerçekleştirilmiştir. Örneklem büyüklüğü, test uzunluğu, madde ve θ parametrelerinin dağılımı, kestirim yöntemleri, MTK modelleri ve replikasyon büyüklüğü gibi değişkenler arasından örneklem büyüklüğü ve θ parametrelerinin farklı dağılımlara sahip olduğu durumlar seçilerek manipüle edilmiştir.

Bir MC çalışmasının prensiplerinden daha önce bahsedilmiştir ve bu prensiplerden biri de çalışmanın tekrar edilebilmesidir. Bu prensibi gerçekleştirmek amacıyla simülasyon çalışmasının yapıldığı yazılımda *"seed"* değeri girilmektedir. Bu değer, çalışma tekrar edilmek istendiğinde çalışmanın ilk kez gerçekleştirildiği zamandaki üretilen parametre değerlerinin tekrar üretilmesine olanak verir. Ancak unutulmamalıdır ki farklı bilgisayarlarda rastgele numara üreticisi (RNG) farklı değerler üretebilir ya da yazılımsal olarak farklılıklar gözlemlenebilir (Bulut ve Sünbül, 2017).

Çalışmada ilk olarak madde parametreleri üretilmiştir. Bu parametreler üretilirken daha önce yapılan çalışmalar incelenmiş ve gerçek durumlara uygun olabilecek dağılımlar kullanılarak üretim işlemi gerçekleştirilmiştir. Bu amaç doğrultusunda a parametresi log-normal dağılımdan rastgele a~lnN(0,0, 0,5), b parametresi standart normal dağılımdan rastgele b~N(0,0, 1,0) üretilmiştir (Seong, 1990; Feinberg ve Rubright, 2016; Bulut ve Sünbül, 2017; Riley ve Carle, 2012). Madde parametrelerini üretmek için RStudio yazılımında R programlama dili kullanılmıştır ve üretim için kullanılan kodlar Şekil 3.1'de verilmiştir.

```
# kodların çalışması için mirt ve moments paketlerinin yüklü olması gerekmektedir.
itemrecovery<- function(nitem, samplesize, seed){
  set.seed(seed)


  item_disc<- as.matrix(round(rlnorm(nitem, meanlog = 0.0, sdlog = 0.5), 3), ncol=1) # madde ayırt edicilik parametrelerini üretir
  item_diff<- as.matrix(round(rnorm(nitem, mean = 0,sd=1 ),3), ncol=1) # madde güçlük parametrelerini üretir
```

**Şekil 3.1.** Sırasıyla madde ayırt edicilik (a) ve madde güçlük (b) parametrelerini üreten kod parçaları.

Burada *itemrecovery* ifadesi çalışma için bir fonksiyondur ve içinde *nitem, samplesize* ve *seed* ifadelerini barındırmaktadır. Bu ifadeler sırasıyla madde sayısı, örneklem büyüklüğü ve çalışmanın tekrar edilebilmesi için girilmesi gereken seed sabit sayısını ifade etmektedir. Bu ifadelere girilecek olan değerler kodun en sonunda yer aldığından burada gösterilemememektedir. *item_disc* ifadesi ise madde ayırt edicilik (a) parametresini göstermektedir ve *rlnorm* fonksiyonu ile üretilmektedir. *item_diff* ise madde güçlük (b) parametresini göstermektedir ve *rnorm* fonksiyonu ile üretilmektedir. Her iki kod parçasında da görülen *as.matrix* ifadesi üretilen değerleri tek sütunlu bir matrise toplamaktadır ve *round* ise üretilen parametre değerlerini virgülden sonra 3 basamak kalacak şekilde yuvarlar (Bulut ve Sünbül, 2017).

Nihayetinde madde parametrelerinin üretimi için kullanılan parametreler Tablo 3.2'de görülebilir.

**Tablo 3.2.** Madde parametrelerini (a, b) üretmek için kullanılan parametre değerleri.

| **Madde Parametreleri** | **Kullanılan Parametre Değerleri** | |
|---|---|---|
| | meanlog(ortalama) | sdlog(standart sapma) |
| Madde Ayırt Edicilik (a) | 0 | 0,5 |
| | Ortalama | Standart Sapma |
| Madde Güçlük (b) | 0 | 1 |

Araştırmada daha sonra θ parametreleri üretilmiştir. Bu parametreler de üretilirken daha önce yapılmış olan çalışmalar incelenmiş ve gerçek durumlara en uygun olabilecek dağılımlar seçilmeye çalışılmıştır. Bu amaç doğrultusunda θ parametreleri beta dağılımından rastgele üretilmiştir. İstenilen çarpıklık değerlerinde veri üretmek için negatif çarpıklık katsayısından (-2,00, -1,00) pozitif çarpıklık katsayısına (1,00, 2,00) doğru sırasıyla θ~BETA(21,0, 0,8; 9,0, 1,45; 1,45, 9,0; 0,8, 21,0) katsayıları kullanılmıştır. Ardından istenilen θ değerlerine ulaşmak amacıyla üretilen her bir değerden ÇK= -2,00 durumunda 0,965 çıkarılmış ve elde edilen sonuç 25,7 ile çarpılmıştır bu işlem diğer çarpıklık katsayılarının üretildiği durumlarda ÇK= -1,00 için 0,86 değerinin çıkarılıp 10,7 ile çarpılarak, ÇK= 1,00 için 0,14 değeri çıkarılıp 10,7 ile çarpılarak, ÇK= 2,00 için 0,035 çıkarılıp 25,7 ile çarpılarak devam etmiştir (DeMars, 2003; Abdel-fattah, 1994; Swaminathan ve Gifford, 1979). Bu işlemlerin ardından elde edilen dağılımların yazılan kodlarla standart sapmalarının 1'e, ortalamalarının 0'a sabit olduğundan ve istenilen çarpıklık katsayılarına sahip olduğundan emin olunmuştur.

Yetenek (θ) parametrelerini üretmek için RStudio yazılımında R programlama dili kullanılmıştır ve üretim için kullanılan kodlar Şekil 3.2’de verilmiştir.

```
repeat{

  ability<- as.matrix(round(rbeta(samplesize,0.8,21),3),ncol=1)
  #istenen çarpık dağılımı üretir ancak beta dağılımı bunu 0 ve 1 sayıları arasında yaptığından ranjı genişletmek gerekmektedir.

  sability<- matrix(ncol = 1)

  # ranjı genişletilmiş dağılım

  for(i in 1:samplesize){

    sability[i]<- as.matrix(round(((ability[i]-0.035)*25.7),3),ncol=1)
    #ranjı genişletilmiş dağılım burada üretilmektedir.

    sability<- as.matrix(sability,ncol=1)
    # verileri analiz için matrix içinde tek sütunda toplar
  }

  if(round(skewness(sability),2)== 2 && round(sd(sability),2)== 1 && round(mean(sability),1)== 0){
    # istenilen özellikleri sağladığında kod döngüyü durdurur ve bir sonraki aşamaya geçer aksi takdirde başa döner
    break
  }else{
    next
  }


}
```

**Şekil 3.2.** Yetenek (θ) parametrelerinin üretildiği ve istenilen özelliklerin dağılıma verildiği kod parçaları.

Burada görülen *rbeta* fonksiyonu θ parametrelerini üretmektedir. Daha sonra gelene *sability* ise üretilen değerlerin θ haline getirilmiş durumlarının aktarıldığı ifadedir. *round(skewness(sability),2)* ifadesi elde edilen dağılımın çarpıklık katsayısının virgülden sonraki 2 basamağa kadar yuvarlayıp istenilen değerde elde edilmesini sağlarken *round(sd(sability),2)* benzer işlemi dağılımın standart sapmasının 1’e eşit olması için *round(mean(sability),1)* ise dağılımın ortalamasının 0’a eşit olması için yapmaktadır.

Standart normal dağılıma N(0,1) sahip θ parametrelerinin üretimi de RStudio yazılımında R programlama dilinde yapılmıştır ve Şekil 3.3’te görülebilir.

```
ability<- as.matrix(round(rnorm(samplesize, mean = 0,sd=1),3), ncol=1)
```

**Şekil 3.3.** Standart normal dağılıma sahip yetenek (θ) parametrelerini üreten kod parçası.

Sonuç olarak kullanılan çarpıklık dağılımlar, bu dağılımları üretmek için kullanılan katsayılar ve istenilen özelliklere sahip olması için gerekli olan çıkarma ve çarpma işlemlerinde kullanılan değerler Tablo 3.3'te görülebilir.

**Tablo 3.3.** Çarpıklık düzeylerine göre kullanılan parametre değerleri.

| **Çarpıklık Düzeyi** | **Kullanılan Parametre Değerleri** | | **Üretilen her bir sayıdan çıkarılan değer** | **Çıkarma işleminden sonra elde edilen her bir sayının çarpıldığı değer** |
|---|---|---|---|---|
| | Ortalama (α) | Standart Sapma (β) | | |
| ÇK= 0,00 | 0 | 1 | | |
| ÇK= 2,00 | 0,8 | 21 | 0,035 | 25,7 |
| ÇK= 1,00 | 1,45 | 9 | 0,14 | 10,7 |
| ÇK= -1,00 | 9 | 1,45 | 0,86 | 10,7 |
| ÇK= -2,00 | 21 | 0,8 | 0,965 | 25,7 |

## 3.5. Verilerin Analizi

Araştırmaya dair tüm işlemler RStudio v1.2.1335 programında R v3.6.1 yazılımı kullanılarak i7-4700MQ ve 8GB RAM özelliklerine sahip bir bilgisayarda gerçekleştirilmiştir. Madde ve yetenek (θ) parametrelerinin üretimi R yazılımına ait çekirdek paket programlarla gerçekleştirilmiştir. Madde parametrelerinin tahmini için *"mirt"* paket programı kullanılmıştır. Bu paket açımlayıcı ve doğrulayıcı modeller için çok boyutlu madde tepki kuramının parametrelerini maximum likelihood yöntemleri ile tahmin etmesi amacıyla oluşturulmuştur (Chalmers, 2012). Bu amacının yanında tek boyutlu lojistik modeller için de tahmin işlemleri gerçekleştirilebilir. mirt paketi varsayılan biçimde Marginal Maximum Likelihood (MML) kestirim yöntemini kullanmaktadır (Paek ve Cole, 2019). Ayrıca kestirim işleminde çeşitli durumlara göre kullanılmak üzere Metropolis-Hastings Robbins-Monro (MH-RM) ve Expectaion-Maximization (EM) algoritmaları da bulunmaktadır. Bu çalışma için MML kestirim yöntemi ve EM algoritması kullanılmıştır.

Veri üretiminden analizine geçişteki ilk adımsa üretilen parametreler doğrultusunda her bir birey için 0 ve 1'lerden oluşan bir cevap matrisi oluşturmaktadır. Bu işlem Şekil 3.4'te görülebilir.

```
dat<- as.matrix(simdata(a=item_disc,d=item_diff,N=samplesize,itemtype = "dich",Theta = sability),ncol=nitem)
#mirt paketi üzerinden üretilmiş olan parametreleri kullanarak 0-1 cevap matrixine çevirir.
```

**Şekil 3.4.** Üretilen madde ve yetenek (θ) parametreleri kullanılarak her birey için 0 ve 1'lerden oluşan bir cevap matrisi üreten kod parçası.

Burada *"dat"* ifadesi bireylerin cevap matrislerinin tutulduğu fonksiyondur. *"simdat"* ise *"mirt"* paketinin parametre değerlerini kullanarak cevap matrisleri oluşturduğu fonksiyondur. *"a"* ve *"d"* ifadeleri madde parametrelerinin ve *"N"* ise örneklem büyüklüğünün girildiği ifadelerdir. *"itemtype"* ise üretilecek olan cevap matrislerinin nasıl puanlandığının belirlendiği ifadedir ve *"dich"* ifadesi ikili puanlama yani doğru cevabın 1 yanlış cevabın 0 olduğu puanlama yöntemini ifade etmektedir. *"Theta"* ise yetenek parametrelerinin girildiği ifadedir (Chalmers, 2012; Bulut ve Sünbül, 2017).

Daha sonra oluşan cevap matrisleri kullanılarak parametre tahmini gerçekleştirilmiştir. Yapılan kestirim işlemi sonucunda parametre değerleri elde edilmiş ve bir fonksiyona toplanmıştır. Bu işlem Şekil 3.5'te görülebilir.

```
model2pl<- mirt(data = dat,model=1,itemtype = "2PL",SE=TRUE,verbose = FALSE) # üretilen cevap matrixinden parametre kestirimi yapar.

parameters <- as.data.frame(coef(model2pl, simplify=TRUE)$items)# kestirilen parametre değerleri buraya aktarılır.
```

**Şekil 3.5.** Oluşturulan cevap matrislerinin kullanılarak madde parametre kestirimlerini yapıldığı kod parçası.

Burada *"model2pl"* madde parametre kestirimlerinin gerçekleştiği kod parçasıdır ve kestirimlere dair bilgiler buraya aktarılır. *"mirt"* ifadesi kestirim işlemini gerçekleştiren fonksiyondur. *"model"* kestirimde kullanılan faktör sayısını belirtir bu sayı çok boyutlu MTK modellerinde kullanılan boyut sayısı kadar a parametre kestirimi yapar. *"item type"* kestirimin hangi MTK modeline göre yapılacağını belirtirken *"SE"* standart hata kestirimlerinin yapılıp yapılmayacağını belirtir bu değer *"TRUE"* olduğunda standart hata kestirimi yapar. *"verbose"* ise kestirim sonunda elde edilen değerlerin yazdırılıp yazdırılmayacağını belirtir bu değer *"FALSE"* iken değerler yazdırılmaz. *"parameters"* kestirilen parametrelerin aktarıldığı fonksiyondur. *"coef"* ise katsayıların modellerden dışarı aktarılmasını sağlayan fonksiyondur (Chalmers, 2012; Bulut ve Sünbül, 2017).

## 3.6. Ölçme Keskinliğinin (Parametre Kestirim Keskinliği) Değerlendirilmesi

Araştırmada üretilen (gerçek) parametreler ile kestirim sonucunda elde edilen parametreler arasındaki farkın ölçülmesi için Bias ve RMSE istatistiklerinin kullanılacağı önceden belirtilmiştir. Bu amaçla yazılan kod parçaları Şekil 3.6'da görülebilir.

```
bias.item_disc<- round(mean(parameters[,1]-item_disc), 3)
# her replikasyonda madde ayırt edicilik (a) parametrelerinin üretilen değerleri ile kestirilen değerleri arasında Bias değeleri hesaplanır.
bias.item_diff<- round(mean(parameters[,2]-item_diff), 3)
# her replikasyonda madde güçlük (b) parametrelerinin üretilen değerleri ile kestirilen değerleri arasında Bias değerleri hesaplanır.
rmse.item_disc<- round(sqrt(mean((parameters[,1]-item_disc)^2)), 3)
# her replikasyonda madde ayırt edicilik (a) parametrelerinin üretilen değerleri ile kestirilen değerleri arasında RMSE değerleri hesaplanır.
rmse.item_diff<- round(sqrt(mean((parameters[,2]-item_diff)^2)), 3)
# her replikasyonda madde güçlük (b) parametrelerinin üretilen değerleri ile kestirilen değerleri arasında RMSE hesaplanır.
```

**Şekil 3.6.** Her replikasyonda üretilen (gerçek) parametreler ile kestirilen parametreler arasındaki Bias ve RMSE istatistiklerinin hesaplandığı kod parçası.

Burada *"bias.item_disc"* ifadesi gerçekleştirilen her replikasyonda üretilen (gerçek) madde ayırt edicilik (a) parametreleri ile kestirilen a parametreleri arasındaki Bias istatistik değerlerini hesaplayarak içinde tutmaktadır. Aynı işlemi madde güçlük (b) parametresi için *"bias.item_diff"* ifadesi yapar. *"rmse.item_disc"* ifadesi ise yine her replikasyonda üretilen (gerçek) a parametreleri ile kestirilen a parametreleri arasında RMSE istatistiğini hesaplayarak içinde tutmaktadır. *"rmse.item_diff"* ise aynı işlemi b parametreleri için yapmaktadır (Chalmers, 2012; Bulut ve Sünbül, 2017).

```
results<- data.frame(sample.size=samplesize, nitem=nitem, bias.item_disc=bias.item_disc, bias.item_diff=bias.item_diff,
                                                        rmse.item_disc=rmse.item_disc, rmse.item_diff=rmse.item_diff)
# sonuçları buraya aktarmak üzere tanımlanmıştır.


return(results)
}
```

**Şekil 3.7.** Replikasyonlar sonucunda elde edilen Bias ve RMSE istatistiklerinin aktarıldığı kod parçası.

Şekil 3.7'de de görülebileceği gibi her bir replikasyon sonunda madde parametrelerinden elde edilen Bias ve RMSE istatistikleri *"results"* ifadesi altında toplanarak *"itemrecovery"* fonksiyonu sonlanır (Bulut ve Sünbül, 2017). Son olarak MC çalışmasını gerçekleştirmek için simülasyon koşulları yazılan kodda ilgili yerlere girilmelidir. Bu yerler Şekil 3.8'de görülebilir.

```
myseed <- sample.int(n = 1000000, size = 100) # seed ve iterasyonun belirlendiği kısım buradan değiştirilebilir.

write.csv(myseed, "simulation seeds.txt", row.names = FALSE)
# seedi dışarı aktarır.


results <- data.frame(samplesize=0, nitem=0, bias.item_disc=0, bias.item_diff=0, rmse.item_disc=0, rmse.item_diff=0)
# sonuçları aktarmak için tanımlanmıştır.


for (i in 1:length(myseed)) {

  results[i,] <- itemrecovery(nitem = 30, samplesize = 2000, seed = myseed[i])

  #analiz ve sonuçların aktarılması işi burada gerçekleştirilir madde sayısı ve örneklem büyüklüğü buradan gerçekleştirilmektedir.

}
```

**Şekil 3.8.** Simülasyon çalışması için koşulların girildiği kod parçaları.

Burada *"myseed"* ifadesi yapılan simülasyon çalışmasının tekrar gerçekleştirilmesi için gerekli olan seed değerlerinin tutulduğu fonksiyondur. *"sample.int"* ifadesi rastgele tam sayı üreten bir fonksiyondur ve içinde bulundurduğu *"n"* ifadesi kaç adet tam sayı üretileceğini belirlerken *"size"* ifadesi üretilen bu tamsayılardan rastgele kaç adet seçileceğini belirtir. Bu sayede seçilen tamsayı kadar replikasyon yapma imkânı elde edilir. *"write.csv"* ise seed değerleri olmak üzere rastgele seçilen tamsayıları bir belgeye kaydeder böylece kayıtlı seed değerleri ilgili alanlara girilerek çalışmadan elde edilen değerlere tekrar ulaşılabilir. *"results"* ifadesi sonuçların aktarıldığı ifadenin önceden tanımlanmış durumudur. *"for (i in 1:length(myseed))"* ifadesi çalışmanın istenene replikasyon sayısı kadar tekrar edilmesini sağlayan fonksiyondur. *"results[i,]"* ifadesi ise simülasyon çalışmasında istenen madde sayısı, örneklem büyüklüğü ve hangi seed değerinin kullanılacağının belirlendiği fonksiyondur (Bulut ve Sünbül, 2017).

# BÖLÜM IV

## 4. Bulgular ve Tartışma

Bu bölümde araştırmadan elde edilen bulguların alt problemler çerçevesinde incelenmesine, RMSE ile Bias istatistiklerinden yararlanarak değerlendirilmesine ve bu bulgulara ilişkin tartışmaya yer verilmiştir.

### 4.1. Bulgular

#### 4.1.1. Birinci alt probleme ilişkin bulgular

Araştırmanın birinci alt problemi " Tek boyutlu 2PL modelde madde parametrelerinin kestiriminde θ parametrelerinin normal dağılmadığı durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 1,00, 2,00) hata kareleri ortalamasının karekökü (RMSE) değerlerini nasıl etkiler?" olarak ifade edilmiştir. Bu alt problem iki alt başlık altında incelenecektir.

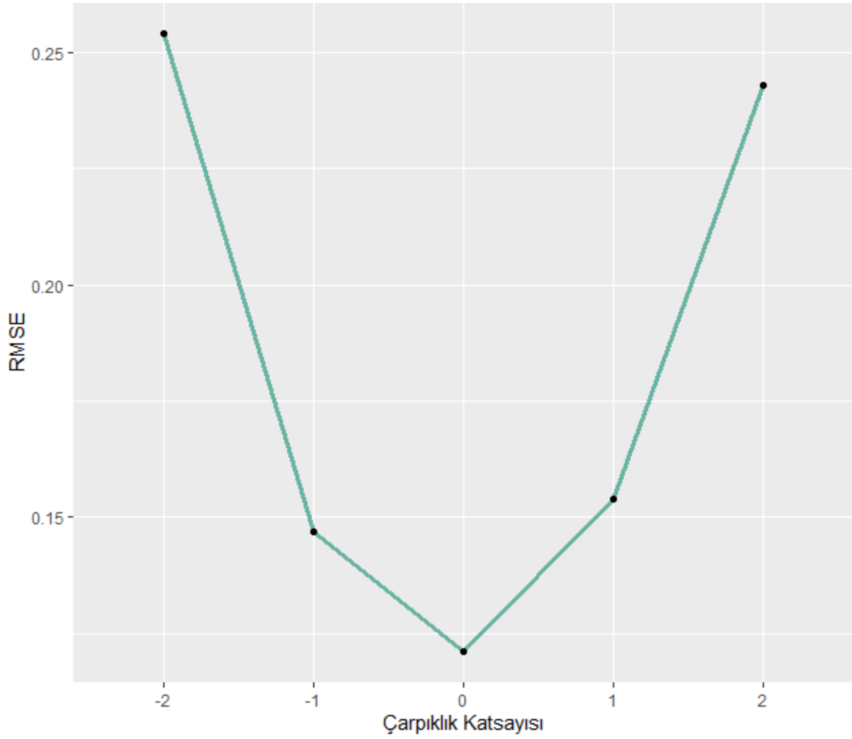
##### 4.1.1.1. 1.1. Alt probleme ilişkin bulgular

"Tek boyutlu 2PL modelde a parametresinin kestirimde θ parametrelerinin normal dağılmadığı durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 1,00, 2,00) hata kareleri ortalamasının karekökü (RMSE) değerlerini nasıl etkiler?" sorusuna cevap arama amacıyla yapılan analizler sonucunda elde edilen veriler Tablo 4.1'de görülebilir. Bu alt problemde Lord (1968) ve Drasgow (1989)'un önerileri doğrultusunda örneklem büyüklüğü 1000 olan veri setleri sunulmuştur.

**Tablo 4.1.** Farklı dağılımlar için a parametresinin ortalama RMSE değerleri.

| **Örneklem Büyüklüğü (N)** | **Test Uzunluğu** | **Çarpıklık Katsayısı (ÇK)** | **Ortalama RMSE** |
|---|---|---|---|
| | | ÇK= 2,00 | 0,243 |
| | | ÇK= -2,00 | 0,254 |
| N=1000 | 30 | ÇK= 1,00 | 0,154 |
| | | ÇK= -1,00 | 0,147 |
| | | ÇK= 0,00 | 0,121 |

Tablo 4.1 incelendiğinde en düşük RMSE değeri (0,121) ÇK= 0,00 yani θ dağılımı standart normal dağılımken en yüksek RMSE değeri (0,254) ÇK= -2,00 olan dağılımdan elde edilmiştir. Bu sonuçlara bakarak kestirim keskinliği için a parametresinden elde edilen ortalama RMSE değerlerinin oldukça yüksek olduğu söylenebilir. Bunlara ek olarak ÇK'nin negatif ya da pozitif olması yani dağılımın çarpıklık yönünün RMSE değerleri arasında önemli bir fark yaratmadığı da gözlemlenebilir. Ayrıca a parametresi için θ parametre dağılımlarının ÇK'si arttıkça RMSE değerlerinin de artma eğiliminde olduğu Grafik 4.1'de görülebilir.




**Grafik 4.1.** Çarpıklık katsayısındaki değişimin a parametresine ait ortalama RMSE değerlerine etkisi.

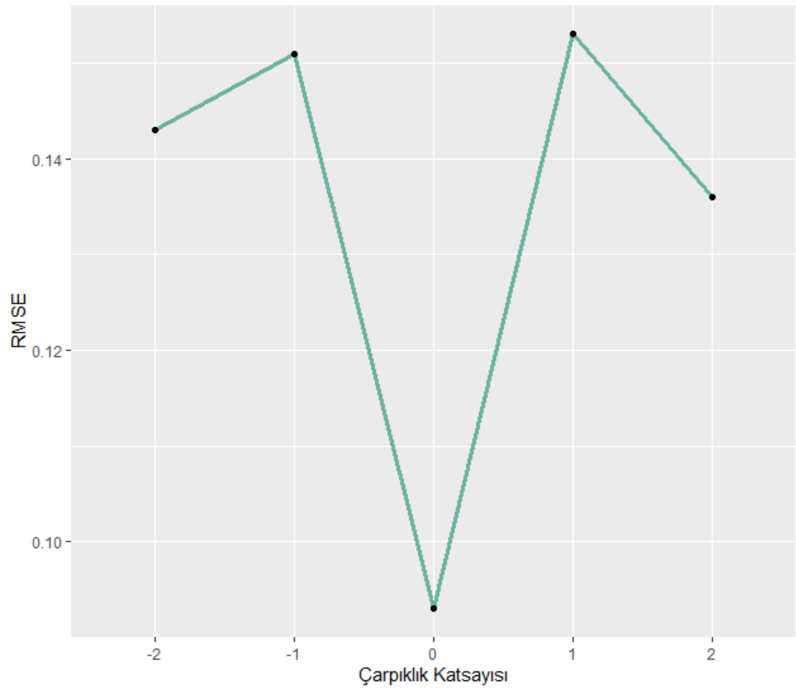
#### 4.1.1.2. 1.2. Alt probleme ilişkin bulgular

"Tek boyutlu 2PL modelde b parametresinin kestirimde θ parametrelerinin normal dağılmadığı durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 1,00, 2,00) hata kareleri ortalamasının karekökü (RMSE) değerlerini nasıl etkiler?" sorusuna cevap arama amacıyla yapılan analizler sonucunda elde edilen veriler Tablo 4.2'de görülebilir. Bu alt problemde Lord (1968) ve Drasgow (1989)'un önerileri doğrultusunda örneklem büyüklüğü 1000 olan veri setleri sunulmuştur.

**Tablo 4.2.** Farklı dağılımlar için b parametrelerinin ortalama RMSE değerleri.

| Örneklem Büyüklüğü (N) | Test Uzunluğu | Çarpıklık Katsayısı (ÇK) | Ortalama RMSE |
|---|---|---|---|
| N=1000 | 30 | ÇK= 2,00 | 0,136 |
| | | ÇK= -2,00 | 0,143 |
| | | ÇK= 1,00 | 0,153 |
| | | ÇK= -1,00 | 0,151 |
| | | ÇK= 0,00 | 0,093 |

Tablo 4.2 incelendiğinde en düşük RMSE değerinin (0,093) θ parametreleri standart normal dağılıma sahipken en yüksek RMSE değerinin (0,153) θ parametreleri ÇK= 1,00 olan dağılıma aitken elde edildiği görülebilir. Bu sonuçlara bakarak kestirim keskinliği için b parametresinden elde edilen ortalama RMSE değerlerinin yalnızca θ parametreleri standart normal dağılıma sahipken kabul edilebilir düzeyde olduğu, diğer dağılım türlerinde RMSE değerlerinin oldukça yüksek olduğu söylenebilir. Bunlara ek olarak ÇK'nin pozitif ya da negatif olmasının ortalama RMSE değerleri arasında önemli bir fark yaratmadığı da söylenebilir. Ayrıca ilginç bir şekilde θ parametrelerinin dağılımı ÇK= 1,00 ve ÇK= -1,00 iken çok yüksek değerler almaktayken dağılımların ÇK'si arttıkça bu değerlerde azalma görülmektedir. Bu durum Grafik 4.2'de görülebilir.




**Grafik 4.2.** Çarpıklık katsayısındaki değişimin b parametrelerine ait ortalama RMSE değerlerine etkisi.

### 4.1.2. İkinci alt probleme ilişkin bulgular

Araştırmanın ikinci alt problemi "Tek boyutlu 2PL modelde madde parametrelerinin kestiriminde θ parametrelerinin normal dağılmadığı durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 1,00, 2,00) Bias değerlerini nasıl etkiler?" olarak ifade edilmiştir. Bu problem iki alt başlık altında incelenecektir.

#### 4.1.2.1. 2.1 Alt probleme ilişkin bulgular

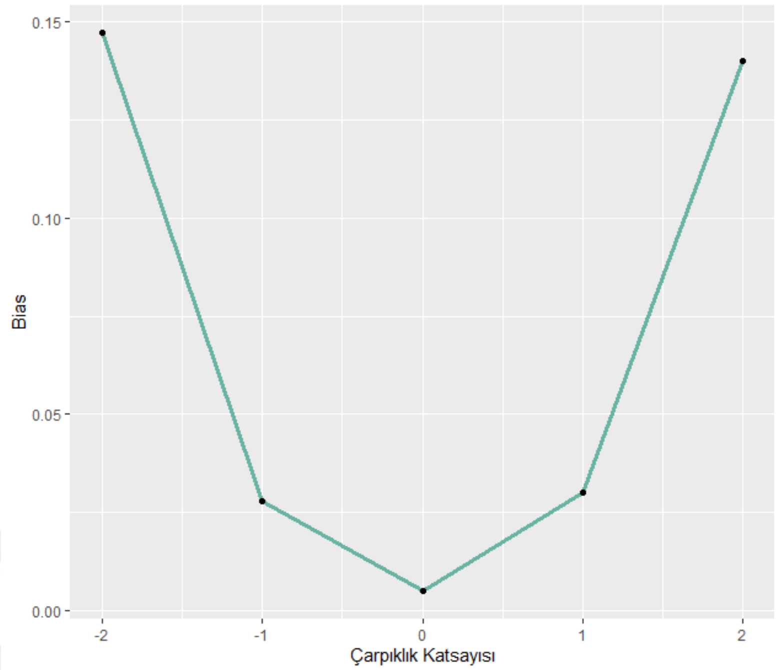
"Tek boyutlu 2PL modelde a parametresinin kestirimde θ parametrelerinin normal dağılmadığı durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 1,00, 2,00) Bias değerlerini nasıl etkiler?" sorusuna cevap arama amacıyla yapılan analizler sonucunda elde edilen veriler Tablo 4.3'te görülebilir. Bu alt problemde Lord (1968) ve Drasgow

(1989)'un önerileri doğrultusunda örneklem büyüklüğü 1000 olan veri setleri sunulmuştur.

**Tablo 4.3.** Farklı dağılımlar için a parametrelerinin ortalama Bias değerleri.

| Örneklem Büyüklüğü (N) | Test Uzunluğu | Çarpıklık Katsayısı (ÇK) | Ortalama Bias |
|---|---|---|---|
| | | ÇK= 2,00 | 0,140 |
| | | ÇK= -2,00 | -0,147 |
| N=1000 | 30 | ÇK= 1,00 | 0,030 |
| | | ÇK= -1,00 | 0,028 |
| | | ÇK= 0,00 | 0,005 |

Tablo 4.3 incelendiğinde en düşük Bias değerinin (0,005) θ parametreleri standart normal dağılıma sahipken en yüksek Bias değerinin (-0,147) ise θ parametreleri ÇK= -2,00 olan dağılıma ait olduğu görülebilir. Bias değerleri 0'a ne kadar yakınlarsa o kadar iyi sonuç alındığına işaret eder böylece en iyi sonucun standart normal dağılımdan alındığı ve dağılımın ÇK'si büyüdükçe değerlerinde kötüleştiği söylenebilir. Bu bilgilere ek olarak ÇK'nin pozitif ya da negatif olması ortalama Bias değerlerini önemli derecede etkilemediği söylenebilir. Ayrıca θ parametrelerinin ÇK'sinin arttıkça b parametrelerine ait ortalama Bias değerlerinin de artma eğiliminde olduğu Grafik 4.3'te görülebilir.




**Grafik 4.3.** Çarpıklık katsayısındaki değişimin a parametrelerine ait ortalama mutlak Bias değerlerine etkisi.

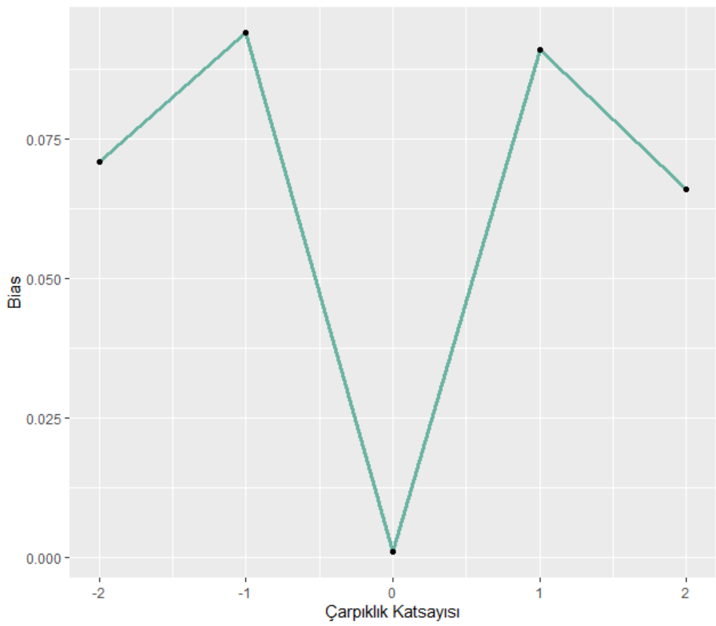
4.1.2.2. 2.2 Alt probleme ilişkin bulgular

"Tek boyutlu 2PL modelde b parametresinin kestirimde θ parametrelerinin normal dağılmadığı durumlar (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 1,00, 2,00) Bias değerlerini nasıl etkiler?" sorusuna cevap arama amacıyla yapılan analizler sonucunda elde edilen veriler Tablo 4.4'te görülebilir. Bu alt problemde Lord (1968) ve Drasgow (1989)'un önerileri doğrultusunda örneklem büyüklüğü 1000 olan veri setleri sunulmuştur.

**Tablo 4.4.** Farklı dağılımlar için b parametrelerinin ortalama Bias değerleri.

| Örneklem Büyüklüğü (N) | Test Uzunluğu | Çarpıklık Katsayısı (ÇK) | Ortalama Bias |
|---|---|---|---|
| | | ÇK= 2,00 | -0,066 |
| | | ÇK= -2,00 | 0,071 |
| N=1000 | 30 | ÇK= 1,00 | -0,091 |
| | | ÇK= -1,00 | 0,094 |
| | | ÇK= 0,00 | -0,001 |

Tablo 4.4 incelendiğinde en düşük Bias değerinin (-0,001) θ parametreleri standart normal dağılıma sahipken en yüksek Bias değerinin ise (0,094) θ parametreleri ÇK= -1,00 olan dağılıma sahipken elde edildiği görülebilir. Bunlara ek olarak ÇK'nin pozitif ya da negatif yönde olmasının ortalama Bias değerleri üzerinde önemli bir etkiye sahip olmadığı söylenebilir. Ayrıca b parametrelerinin ortalama RMSE değerlerinde görüldüğü üzere Bias parametrelerinde de aynı ilginç durum söz konusudur. Beklenenin aksine ÇK'nin hem pozitif hem de negatif 1 olduğu durumlardan elde edilen ortalama Bias değerleri ÇK'nin hem pozitif hem de negatif 2 olduğu durumlardan büyüktür. Bu durum Grafik 4.4'te görülebilir.




**Grafik 4.4.** Çarpıklık katsayısındaki değişimin b parametrelerine ait ortalama mutlak Bias değerlerine etkisi.

### 4.1.3. Üçüncü alt probleme ilişkin bulgular

Araştırmanın üçüncü alt problemi "Tek boyutlu 2PL modelde madde parametrelerinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olması (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 0,00, 1,00, 2,00) hata kareleri ortalamasının karekökü (RMSE) değerlerini nasıl etkiler?" olarak ifade edilmiştir. Bu problem iki alt başlık altında incelenecektir.

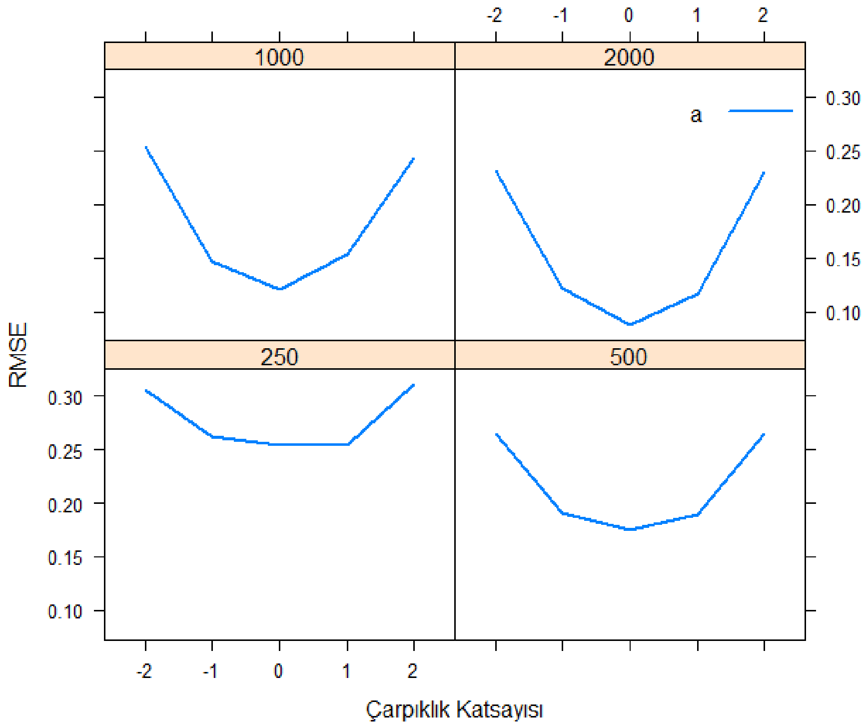
#### 4.1.3.1. 3.1. Alt probleme ilişkin bulgular

"Tek boyutlu 2PL modelde a parametresinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olması (çarpıklık katsayısı (ÇK)= -2,00, -1,00, 0,00, 1,00, 2,00) hata kareleri ortalamasının karekökü (RMSE) değerlerini nasıl etkiler?" sorusuna cevap alma amacıyla yapılan analizler sonucunda elde edilen veriler Tablo 4.5'te görülebilir.

**Tablo 4.5.** Farklı dağılımlar ve örneklem büyüklükleri için a parametrelerinin ortalama RMSE değerleri.

| **Test Uzunluğu** | **Çarpıklık Katsayısı (ÇK)** | **Örneklem Büyüklüğü (N)** | **Ortalama RMSE** |
|---|---|---|---|
| | ÇK= 2,00 | N= 250 | 0,310 |
| | | N= 500 | 0,265 |
| | | N= 1000 | 0,243 |
| | | N= 2000 | 0,230 |
| | ÇK= -2,00 | N= 250 | 0,306 |
| | | N= 500 | 0,264 |
| | | N= 1000 | 0,254 |
| | | N= 2000 | 0,231 |
| 30 | ÇK= 1,00 | N= 250 | 0,254 |
| | | N= 500 | 0,189 |
| | | N= 1000 | 0,154 |
| | | N= 2000 | 0,117 |
| | ÇK= -1,00 | N= 250 | 0,262 |
| | | N= 500 | 0,191 |
| | | N= 1000 | 0,147 |
| | | N= 2000 | 0,122 |
| | ÇK= 0,00 | N= 250 | 0,254 |
| | | N= 500 | 0,175 |
| | | N= 1000 | 0,121 |
| | | N= 2000 | 0,088 |

Tablo 4.5 incelendiğinde bütün dağılım türlerinde örneklem büyüklüğü arttıkça RMSE değerlerinin küçüldüğü, ÇK küçüldükçe RMSE değerlerinin de küçülme eğilimi gösterdiği, ÇK’si mutlak değerde aynı olan dağılımların pozitif ya da negatif olmasının RMSE değerlerinde önemli bir etkiye sahip olmadığı, en iyi sonuçların standart normal dağılıma sahip olan θ parametrelerinde N= 2000 örneklem büyüklüğünde (0,088) elde edildiği, en kötü sonuçların ÇK= 2,00 olan θ parametrelerinde N= 250 örneklem büyüklüğü kullanılması halinde (0,310) elde edildiği görülebilir. Ayrıca ÇK’nin hem pozitif hem de negatif 2,00’den hem pozitif hem de negatif 1,00’e düşmesi durumunda RMSE değerlerinde hatırı sayılır küçülme meydana geldiği ancak aynı etkinin ÇK’nin hem pozitif hem de negatif 1’den standart normal dağılıma düşmesi durumunda gözlemlenemediği Grafik 4.5’te görülebilir.




**Grafik 4.5.** Çarpıklık katsayısındaki ve örneklem büyüklüğündeki değişimin a parametrelerine ait ortalama RMSE değerlerine etkisi.

### 4.1.3.2. 3.2 Alt problemine ilişkin bulgular

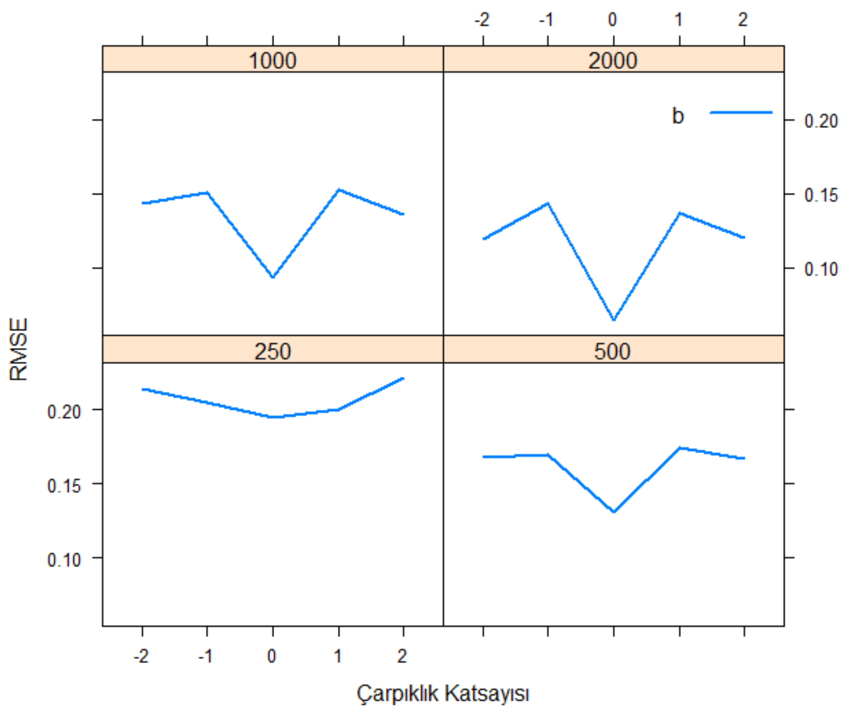
"Tek boyutlu 2PL modelde b parametresinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olması (çarpıklık katsayısı (ÇK)= -2,00, -1,00, 0,00, 1,00, 2,00) hata kareleri ortalamasının karekökü (RMSE) değerlerini nasıl etkiler?" sorusuna cevap alma amacıyla yapılan analizler sonucunda elde edilen veriler Tablo 4-6'te görülebilir.

**Tablo 4.6.** Farklı dağılımlar ve örneklem büyüklükleri için b parametrelerinin ortalama RMSE değerleri.

| Test Uzunluğu | Çarpıklık Katsayısı (ÇK) | Örneklem Büyüklüğü (N) | Ortalama RMSE |
|---|---|---|---|
| | ÇK= 2,00 | N= 250 | 0,221 |
| | | N= 500 | 0,167 |
| | | N= 1000 | 0,136 |
| | | N= 2000 | 0,120 |
| | ÇK= -2,00 | N= 250 | 0,214 |
| | | N= 500 | 0,168 |
| | | N= 1000 | 0,143 |
| | | N= 2000 | 0,119 |
| 30 | ÇK= 1,00 | N= 250 | 0,200 |
| | | N= 500 | 0,174 |
| | | N= 1000 | 0,153 |
| | | N= 2000 | 0,137 |
| | ÇK= -1,00 | N= 250 | 0,205 |
| | | N= 500 | 0,170 |
| | | N= 1000 | 0,151 |
| | | N= 2000 | 0,143 |
| | ÇK= 0,00 | N= 250 | 0,195 |
| | | N= 500 | 0,131 |
| | | N= 1000 | 0,093 |
| | | N= 2000 | 0,065 |

Tablo 4.6 incelendiğinde bütün dağılım türlerinde örneklem büyüklüğü arttıkça b parametrelerine ait ortalama RMSE değerlerinin küçüldüğü görülmüştür. Son derece ilginç bir şekilde ÇK= -1,00, ÇK= 1,00 ile ÇK= -2,00, ÇK= 2,00 karşılaştırıldığında beklenilenin aksine örneklem büyüklüğü arttıkça ÇK= -1,00 ile ÇK= 1,00, ÇK= -2,00 ile ÇK= 2,00'ye göre daha büyük RMSE değerleri verdiği bulgusu elde edilmiştir. b parametrelerinin en iyi ortalama RMSE değerlerinin (0,065) θ parametrelerinin standart normal dağılıma sahip olduğu ve N= 2000 örneklem büyüklüğünün kullanıldığı durumda elde edildiği, en kötü ortalama RMSE değerlerininse (0,221) θ parametrelerinin ÇK= 2,00 olan dağılımdan elde edildiği ve N= 250 örneklem büyüklüğünün kullanıldığı durumda ortaya çıktığı görülebilir. Bu bilgilere ek olarak mutlak değerleri eşit olan çarpıklık katsayılarının pozitif ya da negatif olmalarının RMSE değerleri üzerinde önemli bir

etkilerinin olmadığı da görülebilir. Ayrıca b parametreleri için ÇK'nin hem pozitif hem de negatif 2 değerinin hem pozitif hem de negatif 1 değerine düştüğü durumlarda ortalama RMSE değerlerinde ciddi bir etkinin olmadığı ancak örneklem büyüklüğü arttıkça bu etkinin de ciddi oranlarda arttığı bu durumun özellikle ÇK'nin hem pozitif hem de negatif 1 durumu ile standart normal dağılım durumu karşılaştırıldığında çok daha net şekilde gözlemlenebilir. Bu durumlar Grafik 4.6'da görülebilir.




**Grafik 4.6.** Çarpıklık katsayısında ve örneklem büyüklüğündeki değişimin b parametrelerine ait ortalama RMSE değerlerine etkisi.

### 4.1.4. Dördüncü alt probleme ilişkin bulgular

Araştırmanın üçüncü alt problemi "Tek boyutlu 2PL modelde madde parametrelerinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olması (çarpıklık katsayısı (ÇK) = -

2,00, -1,00, 0,00, 1,00, 2,00) Bias değerlerini nasıl etkiler?" olarak ifade edilmiştir. Bu problem iki alt başlık altında incelenecektir.

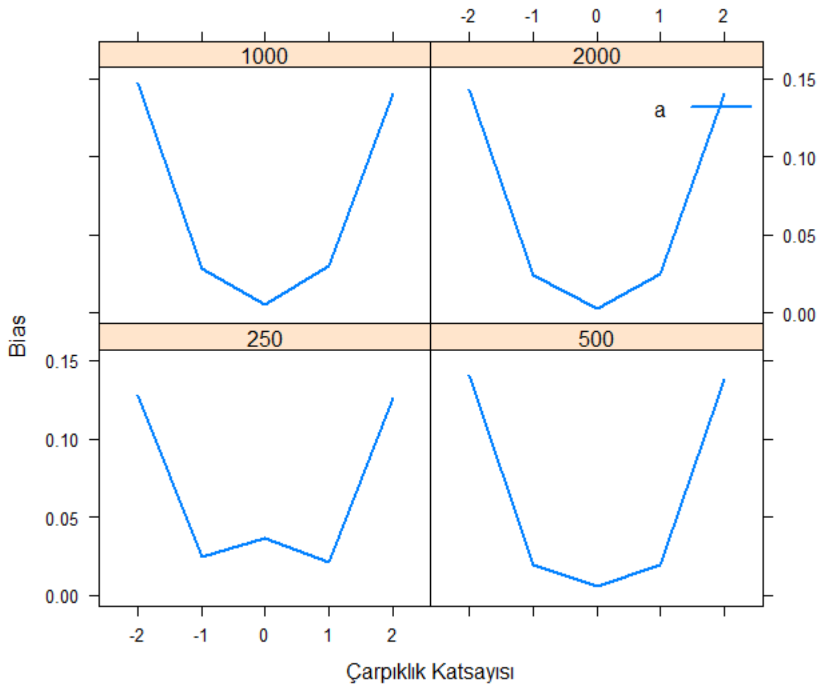
#### 4.1.4.1. 4.1. Alt probleme ilişkin bulgular

"Tek boyutlu 2PL modelde a parametresinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olması (çarpıklık katsayısı (ÇK)= -2,00, -1,00, 0,00, 1,00, 2,00) Bias değerlerini nasıl etkiler?" sorusuna cevap alma amacıyla yapılan analizler sonucunda elde edilen veriler Tablo 4.7'de görülebilir.

**Tablo 4.7.** Farklı dağılımlar ve örneklem büyüklükleri için a parametrelerinin ortalama Bias değerleri.

| Test Uzunluğu | Çarpıklık Katsayısı (ÇK) | Örneklem Büyüklüğü (N) | Ortalama Bias |
|---|---|---|---|
| | ÇK= 2,00 | N= 250 | -0,126 |
| | | N= 500 | -0,138 |
| | | N= 1000 | 0,140 |
| | | N= 2000 | -0,140 |
| | ÇK= -2,00 | N= 250 | -0,128 |
| | | N= 500 | -0,141 |
| | | N= 1000 | -0,147 |
| | | N= 2000 | -0,143 |
| 30 | ÇK= 1,00 | N= 250 | 0,021 |
| | | N= 500 | 0,019 |
| | | N= 1000 | 0,030 |
| | | N= 2000 | 0,025 |
| | ÇK= -1,00 | N= 250 | 0,024 |
| | | N= 500 | 0,019 |
| | | N= 1000 | 0,028 |
| | | N= 2000 | 0,024 |
| | ÇK= 0,00 | N= 250 | 0,036 |
| | | N= 500 | 0,006 |
| | | N= 1000 | 0,005 |
| | | N= 2000 | 0,003 |

Tablo 4.7 incelendiğinde a parametreleri için ortalama Bias değerlerinin tüm dağılım türlerinde örneklem büyüklüğü artışından anlamlı bir etkilenme gerçekleşmediği gözlemlenmiştir. ÇK'nin mutlak değerleri eşit olduğunda ÇK'nin pozitif ya da negatif olmasının ortalama Bias değerlerine büyük bir etkisinin olmadığı da gözlemlenebilir. En iyi ortalama Bias değerlerinin (0,003) θ parametrelerinin standart normal dağılıma sahip olduğu ve N=2000 örneklem büyüklüğünde elde edildiği, en kötü ortalama Bias değerlerininse (-0,147) ÇK= -2,00 olan dağılımda N=1000 örneklem büyüklüğünde elde edildiği görülebilir. Ayrıca a parametreleri için ortalama Bias değerlerinin θ parametrelerinin çarpıklık durumuna daha hassas oldukları ve ÇK standart normal dağılıma yaklaştıkça elde edilen ortalama Bias değerlerinin küçüldüğü Grafik 4.7'de görülebilir.




**Grafik 4.7.** Çarpıklık katsayısında ve örneklem büyüklüğündeki değişimin a parametrelerine ait ortalama mutlak Bias değerlerine etkisi.

#### 4.1.4.2. 4.2 Alt probleme ilişkin bulgular

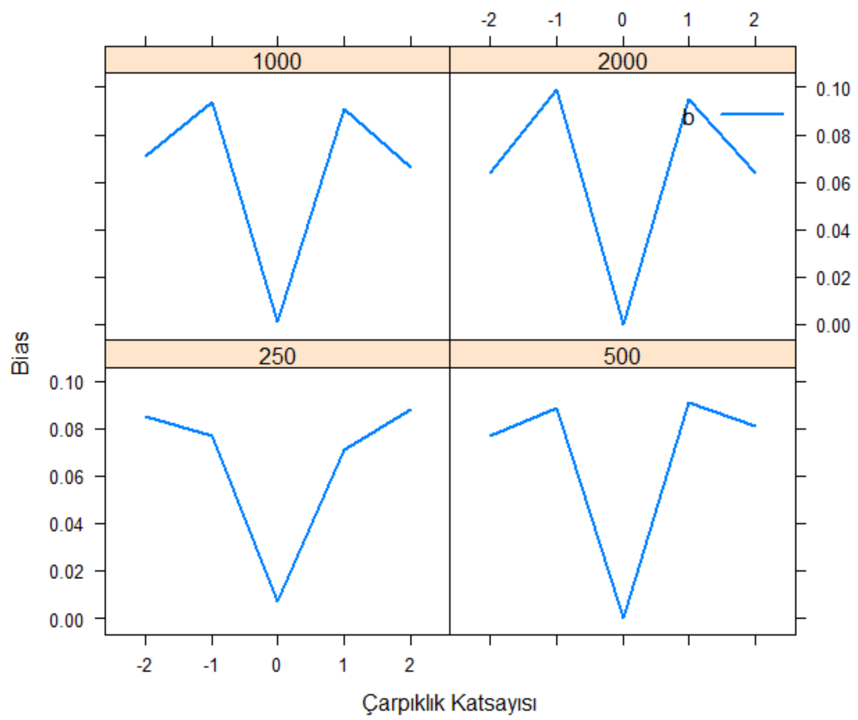
“Tek boyutlu 2PL modelde b parametrelerinin kestiriminde örneklem büyüklüğü (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olması (çarpıklık katsayısı (ÇK)= -2,00, -1,00, 0,00, 1,00, 2,00) Bias değerlerini nasıl etkiler?” sorusuna cevap alma amacıyla yapılan analizler sonucunda elde edilen veriler Tablo 4.8’de görülebilir.

**Tablo 4.8.** Farklı dağılımlar ve örneklem büyüklükleri için b parametrelerinin ortalama Bias değerleri.

| Test Uzunluğu | Çarpıklık Katsayısı (ÇK) | Örneklem Büyüklüğü (N) | Ortalama Bias |
|---|---|---|---|
| | ÇK= 2,00 | N= 250 | -0,088 |
| | | N= 500 | -0,081 |
| | | N= 1000 | -0,066 |
| | | N= 2000 | -0,064 |
| | ÇK= -2,00 | N= 250 | 0,085 |
| | | N= 500 | 0,077 |
| | | N= 1000 | 0,071 |
| | | N= 2000 | 0,064 |
| 30 | ÇK= 1,00 | N= 250 | -0,071 |
| | | N= 500 | -0,091 |
| | | N= 1000 | -0,091 |
| | | N= 2000 | -0,095 |
| | ÇK= -1,00 | N= 250 | 0,077 |
| | | N= 500 | 0,089 |
| | | N= 1000 | 0,094 |
| | | N= 2000 | 0,099 |
| | ÇK= 0,00 | N= 250 | 0,007 |
| | | N= 500 | -0,0001 |
| | | N= 1000 | -0,001 |
| | | N= 2000 | -0,00002 |

Tablo 4.8 incelendiğinde örneklem büyüklüğündeki artışın ÇK= 2,00, ÇK= -2,00 ve standart normal dağılıma sahip durumlarda b parametrelerinin ortalama Bias

değerlerini genelde 0'a yaklaştırdığı görülmüştür. ÇK= 1,00 ve ÇK= -1,00 olan dağılımlarda ise örneklem büyüklüğünde yaşanan artış ortamalama Bias değerlerini 0'dan uzaklaştırmıştır. Mutlak değerleri eşit olan ÇK'lerin pozitif ya da negatif olmasının ortalama Bias değerlerine etkisinin sınırlı olduğu gözlemlenebilir. b parametrelerinin en iyi ortalama Bias değerlerinin (-0,00002) θ parametrelerinin standart normal dağılıma sahip olduğu ve N=2000 örneklem büyüklüğünde elde edildiği, en kötü ortalama Bias değerlerinin ise ÇK= -1,00 olan dağılımda ve N=2000 olan örneklem büyüklüğünden (0,099) elde edildiği görülebilir. Ayrıca en kötü sonuçların ÇK'nin negatif değerlerinde elde edildiği Grafik 4.8'de de gözlemlenebilir.




**Grafik 4.8.** Çarpıklık katsayısında ve örneklem büyüklüğündeki değişimin b parametrelerine ait ortalama mutlak Bias değerlerine etkisi.

## 4.2. Tartışma

Elde edilen sonuçlara göre parametre kestirim keskinliğinin örneklem büyüklüğündeki değişimden ya da θ parametrelerinin dağılımının en az birinden etkilendiği görülmüştür.

Madde ayırt edicilik (a) parametreleri için örneklem büyüklüğü sabit tutulup (N=1,000) yalnızca ÇK manipüle edildiğinde ortalama RMSE değerleri ÇK standart normal dağılıma yaklaştıkça küçülmektedir. Bu sonuç 2 parametreli normal ogive model altında 45 maddelik bir test ve N=1,000 örneklem büyüklüğünü kullanan Seong (1990)'un çalışmasıyla uyuşmamaktadır. Standart normal dağılımın yanında ÇK=-1,00 ve ÇK=1,00 olan dağılımların da kullanıldığı araştırma incelendiğinde önsel (prior) dağılımların eş olduğu durumlarda çarpık dağılımların a parametresi için normal dağılımdan daha küçük RMSE değerleri ürettiği görülmüştür. Bu değerler normal dağılım için 0,072 ve 0,066 iken negatif dağılım için 0,065 ve 0,065, pozitif dağılım için 0,060 ve 0,060'tır. Yine de bulunan değerlere bakıldığında oldukça küçük ve birbirleriyle olan farklarının da oldukça küçük olduğu yani bu değerlerin birbirlerine yakın kabul edilebileceği görülmektedir. Bu sonuçlar araştırmacının kullandığı MTK modeli, kestirim yöntemi, ÇK'ler gibi birçok nedenden kaynaklanıyor olabilir. Stone (1992)'un çalışması incelendiğinde 2PL model altında test uzunluğunu, örneklem büyüklüğünü ve θ parametrelerinin dağılımını manipüle eden araştırmacının elde ettiği bulgular bu çalışmayı desteklemektedir. Araştırmacı a parametreleri için tüm örneklem büyüklükleri ve test uzunluklarında normal dağılımın diğer çarpık dağılımlardan daha küçük RMSE değerleri bulmuştur ve bunu grafikle göstermiştir. Ancak a parametreleri için araştırmacı tüm dağılım koşullarında en yüksek örneklem büyüklüğü (N=1000) ve test uzunluğunda (40 madde) bile 0,1'den küçük RMSE değeri elde edememiştir. Kieftenbeld ve Natesan (2012)'ın çalışmaları incelendiğinde Derecelendirilmiş Tepki Modeli (Graded Response Model) üzerinde θ dağılımlarını, örneklem büyüklüğünü ve test uzunluğunu manipüle eden araştırmacıların elde ettiği bulguların bazıları bu çalışmayı desteklerken bazıları da desteklememektedir. Araştırmada MML kestirim yöntemi altında gerçekleştirilen analizlerin sonuçları incelendiğinde a parametreleri için 20 maddeye sabitlenmiş test uzunluğunda, tüm örneklem büyüklüklerinde (75, 150, 300, 500, 1,000) ÇK=1,75 ve

BK=1,5 olan çarpık dağılımın standart normal dağılımdan daha büyük RMSE değerleri ürettiği görülmektedir ve elde edilen bu sonuç şu anki araştırmayı desteklemektedir. Ancak araştırmacılar θ dağılımları uniform dağılımdan elde edildiğinde standart normal dağılımdan daha küçük RMSE değerleri elde etmiştir. Abdel-fattah (1994)'ın çeşitli kestirim yöntemleri kullandığı (JML, MML, JB, MB) çalışması incelendiğinde 3PL model altında test uzunluğunu, örneklem büyüklüğünü ve θ dağılımlarını (standart normal, Truncated, Beta) manipüle eden araştırmacının a parametreleri için MSD istatistiği kullanarak değerlendirdiği sonuçların bazıları bu çalışmayı desteklerken bazıları da desteklememektedir. Araştırmacının MML kestirim yöntemi kullandığı, N=1,000 örneklem büyüklüğü ve 20 maddelik veri setinden elde ettiği MSD değerleri incelendiğinde normal dağılımdan elde edilen sonucun (0,161) Beta dağılımından elde ettiği sonuca (0,336) göre daha küçük olduğu bu nedenle şu anki çalışmayı desteklediği ancak Truncated dağılımından elde edilen sonuca (0,156) göre daha büyük olduğu bu nedenle şu anki çalışmayı desteklemediği söylenebilir. Yine de bu veri setinden özellikle şu anki çalışmayı desteklemeyen sonucu incelendiğinde aradaki farkın çok küçük olduğu (0,005) bu yüzden aynı kabul edilebileceği söylenebilir. Araştırmacının MML kestirim yöntemi kullandığı, N=1000 ve 60 maddelik veri setinden elde ettiği MSD değerleri incelendiğinde çarpık dağılımların (0,174 ve 0,125) standart normal dağılımdan (0,095) daha büyük MSD değerleri ürettiği bu nedenle sonuçların şu anki çalışmayı tamamen desteklediği söylenebilir. Boulet (1996)'in çalışması incelendiğinde 2PL model altında test uzunluğunu, örneklem büyüklüğünü ve θ dağılımlarını manipüle eden araştırmacının a parametresi için RMSE istatistiğinden elde ettiği sonuçlar tüm test uzunlukları, örneklem büyüklükleri ve kullanılan yazılımlar için bu araştırmayı tamamen desteklemektedir. Sass vd. (2008)'nin çalışmaları incelendiğinde 2PL model altında b parametrelerinin ve θ parametrelerinin dağılımını manipüle eden araştırmacıların RMSE istatistiğinden elde ettikleri sonuçlar bu çalışmayı desteklemektedir. Test uzunluğunu 30 maddeye, örneklem büyüklüğünü ise N=1000'e sabitleyen araştırmacılar her madde için RMSE değeri bulmuşlardır. b parametrelerinin normal dağıldığı ve θ dağılımlarının değişiklik gösterdiği çalışmada RMSE değerlerinin ortalamaları θ parametrelerinin normal dağıldığı durumda (0,13) çarpık dağıldığı duruma (0,23) göre daha küçüktür. Yen (1982)'in çalışması incelendiğinde 3PL model altında test uzunluğunu (10, 20 ve 40 madde), θ parametrelerinin dağılımını (normal, negatif çarpık, pozitif çarpık ve simetrik-basık) ve testlerdeki maddelerin güçlük derecesini (*read.voc.* olarak adlandırılan ortalama madde güçlüğünün 0,23 olduğu test türü, kolay, orta ve zor) manipüle eden araştırmacının a parametresi için θ parametreleri normal dağılırken RMSD istatistiğinden elde ettiği sonuçlar test uzunluğu 20 madde olan ve madde güçlük parametrelerinin ortalaması 0,23 olan *Read.Voc.* testinde LOGIST yazılımı için 0,23, BILOG yazılımı için 0,16'dır. Aynı test türünün ve uzunluğunun kullanıldığı çarpık dağılımlarda (Negatif çarpık (ÇK=-0,4, BK=-0,1), simetrik-basık (ÇK=0,1, BK=-,04), pozitif çarpık(ÇK=0,4, BK=-0,1)) sırasıyla LOGIST yazılımı için 0,16, 0,21 ve 0,16 olarak BILOG yazılımı için sırasıyla 0,18, 0,17 ve 0,17 olarak tespit edilmiştir. Test uzunluğunun 40 olduğu ve aynı test türünün kullanıldığı normal dağılımda elde edilen RMSE sonuçları LOGIST yazılımı için 0,16 ve BILOG yazılımı için 0,21'dir. Aynı test türü ve uzunluğunun kullanıldığı durumda çarpık dağılımlarda sonuçlar yukarıdaki sırayla aynı olmak üzere LOGIST yazılımı için 0,18, 0,18 ve 0,16 olarak BILOG yazılımı içinse 0,18, 0,17 ve 0,15 olarak tespit edilmiştir. Tüm bu sonuçlar değerlendirildiğinde 20 maddelik test uzunluğunun kullanıldığı durumlarda BILOG yazılımından elde edilen sonuçlar şu anki çalışmayı desteklerken 40 maddelik test uzunluğunun kullanıldığı durumlarda LOGIST yazılımından elde edilen sonuçlar şu anki çalışmayı desteklemektedir.

Madde ayırt edicilik (a) parametreleri için örneklem büyüklüğü sabit tutulup (N=1,000) yalnızca ÇK manipüle edildiğinde ortalama Bias değerleri ÇK standart normal dağılıma yaklaştıkça küçülmektedir. Bu sonuç 2 parametreli normal ogive model altında 45 maddelik bir test ve N=1,000 örneklem büyüklüğünü kullanan Seong (1990)'un çalışmasıyla uyuşmamaktadır. Standart normal dağılımın yanında ÇK=-1,00 ve ÇK=1,00 olan dağılımların kullanıldığı araştırma incelendiğinde önsel (prior) dağılımların eş olduğu durumlarda çarpık dağılımların a parametresi için normal dağılıma göre sıfıra daha yakın değerler ürettiği görülmüştür. Bu değerler normal dağılım için 0,059 ve 0,052 iken negatif dağılım için 0,051 ve 0,051, pozitif dağılım için 0,049 ve 0,048'tır. Yine de bulunan değerlere bakıldığında oldukça küçük ve birbirleriyle olan farklarının (farkların maksimum olduğu değer 0,010'dur) da oldukça küçük olduğu yani bu değerlerin birbirlerine yakın kabul edilebileceği görülmektedir. Bu sonuçlar araştırmacının kullandığı MTK modeli, kestirim yöntemi, ÇK'ler gibi birçok nedenden kaynaklanması mümkündür. Stone (1992)'un çalışması incelendiğinde 2PL model altında test uzunluğunu, örneklem büyüklüğünü ve θ parametrelerinin dağılımını manipüle eden araştırmacının a parametrelerinin Bias değerleri için elde ettiği bazı bulgular bu çalışmayı desteklemektedir. Araştırmacının grafik olarak gösterdiği sonuçlarda test uzunluğunun 10 ve 20 olduğu veri setlerinde tüm örneklem büyüklükleri için elde ettiği Bias değerleri bu çalışmayı desteklerken yalnızca test uzunluğu 40 madde olan ve örneklem büyüklükleri N=500 ve N=1000 olan veri setlerinde θ'ların çarpık dağıldığı durumlar standart normal dağılıma sahip olduğu durumlara göre sıfıra daha yakın Bias değerleri üretmişlerdir. Abdel-fattah (1994)'ın çalışması incelendiğinde MML kestirim yönteminin kullanıldığı analizlerde a parametresinin ortalama Bias değerlerinin karesi için test uzunluğunun 20 madde, örneklem büyüklüğünün N=1,000 olduğu veri setinden elde ettiği sonuçlar bu çalışmayı desteklemektedir. Ancak test uzunluğunun 60 madde ve örneklem büyüklüğünün N=1000 olduğu veri setinden elde ettiği sonuçlar bu çalışmayla tamamen çelişmektedir. Bahsedilen veri setinde standart normal dağılım için elde ettiği ortalama Bias'ın karesinin değeri 0,090 iken Truncated dağılım için bu değer 0,064, Beta dağılımı içinse 0,050'dir.

Madde güçlük (b) parametreleri için örneklem büyüklüğü sabit tutulup (N=1,000) yalnızca ÇK manipüle edildiğinde ortalama RMSE değerlerinde en küçük değeri standart normal dağılım verirken onu sırasıyla ÇK=2,00, ÇK=-2,00, ÇK=-1,00 ve ÇK=1,00 izlemiştir. Bu sonuç 2 parametreli normal ogive model altında 45 maddelik bir test ve N=1,000 örneklem büyüklüğünü kullanan Seong (1990)'un çalışmasıyla uyuşmamaktadır. Standart normal dağılımın yanında ÇK=-1,00 ve ÇK=1,00 olan dağılımların da kullanıldığı araştırma incelendiğinde önsel (prior) dağılımların eş olduğu durumlarda b parametresi için çarpık dağılımlardan elde edilen RMSE değerlerinin (negatif dağılım için 0,104 ve 0,103, pozitif dağılım için 0,111 ve 0,110) normal dağılımdan elde edilen değerlerle (0,108 ve 0,101) neredeyse eşit olduğu görülmüştür. Zwinderman ve Wollenberg (1990)'in MML ve CML kestirim yöntemlerini kullandıkları çalışmaları incelendiğinde Rasch modeli altında test uzunluğu, θ dağılımlarını ve b parametrelerinin minimum ve maksimum değerlerini manipüle eden (-3,3 ve -1,1) araştırmacıların RMSD istatistiğinden elde ettikleri sonuçlar tüm test uzunlukları, kestirim yöntemleri, θ dağılımları ve b parametrelerinin tüm maksimum ve minimum değerleri için en çarpık dağılımdan çarpıklığı en küçük dağılıma kadar küçülme eğilimindedir. Elde edilen sonuçlar bu nedenle şu anki çalışmayı desteklememektedir. Stone (1992)'un çalışması incelendiğinde b parametrelerinin ortalama RMSE değerleri θ parametrelerinin farklı dağılımlarında tüm test uzunluklarında ve örneklem büyüklüklerinde birbirlerine oldukça yakındır. Ayrıca Stone (1992) çalışmasından elde ettiği b parametreleri için ortalama RMSE değerleri bu çalışmayla oldukça yakındır. Bu nedenle sonuçların şu anki çalışmayı desteklediği söylenebilir. Boulet (1996)'in çalışması incelendiğinde tüm örneklem büyüklükleri, test uzunlukları ve parametre kestirimi için kullanılan yazılımlarda b parametresi için ortalama RMSE değerleri çarpık dağılımdan standart normal dağılıma doğru küçülmektedir. Bu sonuçlar ışığında Boulet (1996)'in çalışmasının bu araştırmanın sonuçlarını desteklemediği söylenebilir. Sass vd. (2008)'in çalışması incelendiğinde b parametrelerinin kestirimi için θ parametreleri ve b parametreleri standart normal dağılıma sahipken (0,06) θ parametrelerinin normal dağılmadığı ve b parametrelerinin standart normal dağılıma sahip olduğu duruma (0,19) göre daha küçük RMSE değerleri elde edildiği görülebilir. Bu araştırmada da b parametrelerinin kestirimi için θ parametreleri standart normal dağılıma sahipken en küçük RMSE değerlerini verdiğinden Sass vd. (2008)'in çalışmasının bu araştırmayı desteklediği söylenebilir. Abdel-fattah (1994)'ın çalışması incelendiğinde MML kestirim yönteminde test uzunluğunun 20 madde ve örneklem büyüklüğünün N=1,000 olduğu veri setinde standart normal dağılımını (0,143) diğer dağılımlardan (truncated dağılım için 0,432 ve beta dağılımı için 0,249) daha küçük MSD değerleri ürettiği görülmektedir. Test uzunluğunun 60 madde ve örneklem büyüklüğünün N=1000 olduğu veri seti için de aynı durum (standart normal dağılımda 0,139, truncated dağılımda 0,263 ve beta dağılımında 0,291) geçerlidir. Bu sonuçların şu anki çalışmayı en küçük MSD değerinin standart normal dağılımdan elde edilmesi bakımından desteklediği söylenebilir. Yen (1982)'in çalışması incelendiğinde θ parametreleri standart normal dağılıma sahipken test uzunluğunun 20 madde olduğu ve aynı test türünün (*read.voc.*) kullanıldığı durumlarda b parametreleri için ortalama RMSD değeri LOGIST yazılımı için 0,18 olarak BILOG yazılımı içinse 0,11 olarak gözlemlenmiştir. Araştırmacı pozitif çarpık, simetrik-basık ve negatif çarpık olarak belirttiği dağılım türleri için aynı test uzunluğu ve test türünde sırasıyla ortalama RMSD değerlerini LOGIST yazılımında 0,26, 0,19 ve 0,20 olarak BILOG yazılımında ise 0,13, 0,12 ve 0,09 olarak tespit etmiştir. Yen (1982), test uzunluğunun 40 madde olduğu ve test türünün yukarıda belirtilen türle aynı olduğu durumlarda ise θ parametreleri normal dağılırken ortalama RMSD değerleri LOGIST yazılımında 0,12; BILOG yazılımında ise 0,11 olarak bulmuştur. Yukarıdaki sırayla aynı olmak üzere çarpık dağılımlar için aynı test uzunluğu ve türünde LOGIST yazılımı için sırasıyla 0,17, 0,15 ve 0,16 olarak BILOG yazılımı için sırasıyla 0,11, 0,12 ve 0,13 olarak bulunmuştur. Tüm bu sonuçlar değerlendirildiğinde Yen (1987)'in çalışmasının test uzunluğu 20 madde olduğunda normal dağılımdan saptıkça tüm çarpık dağılımlar için RMSD değerleri büyüdüğü için LOGIST değerlerinin şu anki çalışmayı desteklediği söylenebilir. 40 maddelik test uzunluğunda ise aynı nedenden dolayı BILOG sonuçlarının bu çalışmayı desteklediği söylenebilir.

Madde güçlük (b) parametreleri için örneklem büyüklüğü sabit tutulup (N=1,000) yalnızca ÇK manipüle edildiğinde ortalama Bias değerlerinde sıfıra en yakın değeri standart normal dağılım verirken onu sırasıyla ÇK=2,00, ÇK=-2,00, ÇK=1,00 ve ÇK=-1,00 izlemiştir. Seong (1990)'un çalışması incelendiğinde b parametreleri için prior ve posterior dağılımlar aynı ve örneklem büyüklüğü N=1,000 olduğu durumda standart normal dağılımdan sırasıyla 0,085 ve 0,076, negatif dağılımdan sırasıyla 0,079 ve 0,077 ve pozitif dağılımdan sırasıyla 0,086 ve 0,086 ortalama Bias değerleri elde etmiştir. Sonuçlar değerlendirildiğinde birbirlerine çok yakın oldukları bu yüzden aralarındaki farkın göz ardı edilebileceği söylenebilir. Bu sonuçların yalnızca sıfıra en yakın ortalama Bias değerinin standart normal dağılımdan elde edilmesi nedeniyle bu araştırmayı desteklediği söylenebilir. Stone (1992)'un çalışması incelendiğinde test uzunluğunun 40 madde ve örneklem büyüklüğünün N=1,000 olduğu durum dışındaki tüm veri setlerinde standart normal dağılan θ parametrelerinden elde edilen ortalama Bias değerleri diğer dağılım türlerine göre sıfıra daha yakındır. İstisna olan durumda ise basıklığı manipüle edilen dağılım diğerlerine göre sıfıra daha yakın ortalama Bias değerlerine sahiptir. Ancak örneklem büyüklüğünün N=1,000 olduğu durumda bu fark göz ardı edilebilecek kadar küçüktür. Bu nedenle sıfıra en yakın ortalama Bias değerlerinin standart normal dağılımdan elde edilmesi şu anki çalışmayı desteklemektedir. Abdel-fattah (1994)'ın çalışması incelendiğinde MML kestirim yönteminde test uzunluğunun 20 madde ve örneklem büyüklüğünün N=1,000 olduğu veri setinde standart normal dağılımın (0,031) diğer dağılımlardan (truncated dağılım için 0,206 ve beta dağılımı için 0,076) sıfıra daha yakın Bias değerlerinin karesini ürettiği görülmektedir. Test uzunluğunun 60 madde ve örneklem büyüklüğünün N=1,000 olduğu veri seti için de aynı durum (standart normal dağılımda 0,023, truncated dağılımda 0,107 ve beta dağılımında 0,095) geçerlidir. Bu sonuçların sıfıra en yakın Bias değerinin standart normal dağılımdan elde edilmesi nedeniyle şu anki çalışmayı desteklediği söylenebilir.

Madde ayırt edicilik (a) parametreleri için hem örneklem büyüklüğü hem de çarpıklık katsayıları manipüle edildiğinde ortalama RMSE değerleri tüm ÇK'ler için örneklem büyüklüğü arttıkça küçülmüştür. Aynı zamanda ÇK'ler ÇK=0,00'a yaklaştıkça ortalama RMSE değerlerinde yine küçülme gözlemlenmiştir. Seong (1990)'un çalışması incelendiğinde eşleşen prior ve posterior dağılımlarında (Normal/normal, Negatif/negatif, Pozitif/pozitif) tüm θ dağılım türlerinde örneklem büyüklüğü arttıkça (N=100; 1,000) elde edilen ortalama RMSE değerleri de küçülmüştür. Bu sonuçların ışığında Seong (1990)'un çalışmasının şu anki araştırmanın sonuçlarını desteklediği söylenebilir. Aynı zamanda pozitif ve negatif dağılımların tüm örneklem büyüklüklerinde ürettiği ortalama RMSE değerlerine bakıldığında bu araştırmada olduğu gibi dağılımın yönünün herhangi bir etkisinin olmadığı görülmektedir. Stone (1992)'un çalışması incelendiğinde tüm test uzunlukları (10, 20, 40 madde) ve θ parametrelerinin dağılımında (normal, çarpık ve basık) örneklem büyüklüğündeki artışın (N=250, 500, 1000) ortalama RMSE değerlerini düşürdüğü görülmektedir. Aynı zamanda tüm test uzunluklarında θ parametrelerinin standart normal dağılımının diğer çarpık dağılımlardan daha küçük ortalama RMSE değerleri ürettiği gözlemlenmiştir. Bu sonuçların şu anki araştırmadan elde edilen sonuçları tamamen desteklediği söylenebilir. Kieftenbeld ve Natesan (2012)'ın çalışmaları incelendiğinde MML kestirim yöntemi için tüm dağılım türlerinde (normal, çarpık ve uniform) örneklem büyüklüğü arttıkça (N=75, 150, 300, 500, 1000) elde edilen RMSE değerlerinde küçülme gözlenmektedir. Bu nedenle araştırmacıların ulaştığı sonuçların şu anki çalışmadan elde edilen sonuçları desteklediği söylenebilir. Boulet (1996)'in çalışması incelendiğinde tüm test uzunluklarında (15, 30, 45 ve 60 madde), kestirimin uygulandığı yazılım türlerinde (NOHARM ve TESTFACT) ve θ parametrelerinin dağılım türlerinde (Normal ve $\chi^2_{(3)}$ ) örneklem büyüklüğü arttıkça (N=250, 500, 1,000 ve 10,000) a parametreleri için elde edilen RMSE değerlerinde küçülme gözlemlenmiştir. Bu nedenle Boulet (1996)'in çalışmasından elde edilen sonuçlar şu anki çalışmadan elde edilen sonuçlar desteklemektedir. Abdel-fattah (1994)'ın çalışması incelendiğinde elde edilen sonuçların MML kestirim yöntemi için tüm test uzunluklarında (20 ve 60 madde) ve θ parametrelerinin dağılım türlerinde (normal, truncated ve beta) örneklem büyüklüğü arttıkça (N=250, N=1,000) elde edilen

RMSE değerlerinde küçülme gözlemlenmiştir. Bu nedenle araştırmacının çalışmasında ulaştığı sonuçların bu çalışmadan elde edilen sonuçları desteklediği söylenebilir. Yoes (1995)'in XCALIBRE, BILOG, ASCAL ve LOGIST yazılımlarını kullandığı çalışması incelendiğinde 3PL model altında test uzunluklarını (15, 25, 50, 75 ve 100 madde), örneklem büyüklüklerini (N=250, 500, 1,000 ve 2,000) ve madde ayırt edicilik (a) katsayılarının düzeylerini (ortalaması 0,75, standart sapması 0,1 olan Test 1 ve ortalaması 1,50, standart sapması 0,2 olan Test 2) manipüle ettiği görülmüştür. Test 1'de test uzunluğunun 25 maddeden oluştuğu durumlarda LOGIST yazılımı için örneklem büyüklüğü arttıkça sırasıyla 0,64, 0,37, 0,22 ve 0,31 ortalama RMSE değerleri elde edilmiştir. Aynı koşullarda ASCAL yazılımı sırasıyla 0,54, 0,40, 0,34 ve 0,36; BILOG yazılımı 0,30, 0,12, 0,10 ve 0,13; XCALIBRE yazılımı 0,12, 0,08, 0,12 ve 0,09 değerlerini üretmişlerdir. Test 2'de test uzunluğunun 25 maddeden oluştuğu durumlarda LOGIST yazılımı için örneklem büyüklüğü arttıkça sırasıyla 0,68, 0,45, 0,47 ve 0,29 ortalama RMSE değerleri elde edilmiştir. Aynı koşullarda ASCAL yazılımı sırasıyla 0,43, 0,37, 0,40 ve 0,38; BILOG yazılımı 0,51, 0,38, 0,38 ve 0,21; XCALIBRE 0,19, 0,22, 0,18 ve 0,18 değerlerini üretmiştir. Bu çalışmadan elde edilen sonuçlar değerlendirildiğinde test uzunluğunun 25 madde olduğu durumların hiçbir koşulunun şu anki araştırmanın sonuçlarını desteklemediği görülebilir. Akour ve AL-Omari (2013)'nin çalışmaları incelendiğinde 3PL model altında test uzunluklarını (15, 30 ve 60 madde) ve örneklem büyüklüklerini (N=200, 500, 1,000, 5,000, 10,000 ve 20,000) manipüle etmişlerdir. Tüm test uzunluklarında a parametrelerinin ortalama RMSL değerleri örneklem büyüklüğü arttıkça küçülmüştür. Çalışmadan elde edilen bu sonuçların şu anki araştırmayı tamamen desteklediği söylenebilir. Şahin ve Anıl (2017)'ın çalışmaları incelendiğinde araştırmacılar 1PL, 2PL ve 3PL model altında test uzunluklarını (10, 20 ve 30 madde) ve örneklem büyüklüklerini (N=150, 250, 350, 500, 750, 1,000, 2,000, 3,000 ve 5,000) manipüle etmişlerdir. 2PL model altında gerçekleştirilen analizler incelendiğinde a parametrelerinin ortalama RMSD değerleri için test uzunluğu 10 madde olan veri setinde örneklem büyüklüğü N=150'den N=250'ye çıktığında elde edilen ortalama RMSD değerlerinde artış gözlemlenmiştir. N=250 örneklem büyüklüğünden sonra örneklem büyüklüğü arttıkça genellikle ortalama RMSD değerlerinde küçülme görülmüştür. Test uzunluğu 20 maddeden oluşan veri setlerinde örneklem büyüklüğü arttıkça ortalama RMSD değerlerinde küçülme gözlenirken test uzunluğu 30 madde olan veri setleri incelendiğinde örneklem büyüklüğü N=250’den N=350’ye çıktığında ortalama RMSD değerlerinde artış gözlenmiştir. Test uzunluğunun 30 madde olduğu veri setlerinde yukarıdaki durum dışında örneklem büyüklüğü arttıkça ortalama RMSD değerlerinde genellikle küçülme gözlemlenmiştir. Elde edilen bu ilginç sonuçların örneklem büyüklüğünü çok küçük olmasından kaynaklandığı düşünülmektedir. Çalışmadan elde edilen bu sonuçların küçük örneklem büyüklüklerinden elde edilen ilginç sonuçlar göz ardı edildiği takdirde şu anki çalışmayı desteklediği söylenebilir.

Madde ayırt edicilik (a) parametreleri için hem örneklem büyüklüğü hem de çarpıklık katsayıları manipüle edildiğinde ortalama Bias değerleri ÇK=-2,00 ve ÇK=2,00 için örneklem büyüklüğü arttıkça sıfırdan uzaklaşma, ÇK=0,00 için sıfıra yakınlaşma eğilimindedir. ÇK=-1,00 ve ÇK=1,00 için örneklem büyüklüğündeki artış ortalama Bias değerlerinde belirli şekilde sıfıra yakınlaşma ya da sıfırdan uzaklaşma göstermek yerine dalgalı bir seyir izlemiştir. Seong (1990)’un çalışması incelendiğinde prior ve posterior dağılımların eşleştiği durumlarda tüm dağılımlar için örneklem büyüklüğü arttıkça elde edilen ortalama Bias değerleri de sıfıra yakınlaşmıştır. Bu sonuçlar değerlendirildiğinde şu anki çalışmada kullanılan ÇK=0,00 için sonuçlar desteklenirken diğer ÇK’ler için sonuçlar desteklenmemektedir. Stone (1992)’un çalışması incelendiğinde test uzunluğu 10 maddeden oluşan ve θ parametreleri normal ve çarpık dağılıma sahip olan veri setlerinde örneklem büyüklüğü artarken a parametreleri için ortalama Bias değerleri sıfıra yakınlaşmaktadır. Ancak basık dağılımda örneklem büyüklüğü N=500’den N=1,000’e çıktığında ortalama Bias değerleri sıfırdan uzaklaşmıştır. 20 maddeden oluşan test uzunluğunda ve çarpık dağılımda örneklem büyüklüğü arttıkça ortalama Bias değerleri sıfıra yakınlaşırken diğer dağılım türlerinde örneklem büyüklüğü N=500’den N=1,000’e çıktığında ortalama Bias değerlerinde sıfırdan uzaklaşma gözlemlenmiştir. Test uzunluğu 40 madde olan ve normal dağılıma sahip veri setlerinde örneklem büyüklüğü arttıkça ortalama Bias değerleri sıfırdan uzaklaşmıştır. Diğer dağılımlarda ise örneklem büyüklüğü arttıkça ortalama Bias değerleri ya aynı kalmış ya da sıfıra yakınlaşma eğiliminde olmuşlardır. Bu bulgular değerlendirildiğinde şu anki çalışmayla benzer sonuçlara ulaştıkları görülebilir. Bu nedenle Stone (1992)’un çalışması bu araştırmayı desteklemektedir. Abdel-fattah (1994)’ın çalışması incelendiğinde MML kestirim yöntemi için test uzunluğunun 20 maddeden oluştuğu veri setlerinde θ parametreleri normal ve truncated dağılımına sahipken örneklem büyüklüğü arttıkça ortalama Bias değerlerinin karesi sıfıra yakınlaşmıştır. Ancak θ parametreleri beta dağılımına sahipken örneklem büyüklüğü arttıkça ortalama Bias değerlerinin karesi sıfırdan uzaklaşmıştır. Test uzunluğunun 60 maddeden oluştuğu veri setleri incelendiğinde ise tüm dağılımlarda örneklem büyüklüğü arttıkça ortalama Bias değerlerinin karesi sıfıra yakınlaşmıştır. Bu bulgular ışığında şu anki çalışmadan elde edilen sonuçların θ dağılımının türü, madde (a, b) ve yetenek (θ) parametrelerinin dağılımlarının birbirlerini etkilemesi, test uzunluğunun yetersiz kalması gibi nedenlerden kaynaklandığı düşünülmekle beraber Abdel-fattah (1994)'ın bu çalışmasında kullandığı beta dağılımından oluşan veri setleri şu anki araştırmayı desteklemektedir.

Madde güçlük (b) parametreleri için hem örneklem büyüklüğü hem de ÇK manipüle edildiğinde örneklem büyüklüğü arttıkça tüm ÇK'ler için ortalama RMSE değerleri beklendiği gibi küçülme eğilimindedir. Ancak N=250 örneklem büyüklüğü hariç diğerlerinde ÇK=-2,00 ve ÇK=2,00 olan dağılımlar ÇK=-1,00 ve ÇK=1,00 olan dağılımlardan daha küçük ortalama RMSE değerleri üretmiştir. En küçük RMSE değerleri ise tüm örneklem büyüklüklerinde θ parametreleri standart normal dağılıma sahipken elde edilmiştir. Seong (1990)'un çalışması incelendiğinde prior ve posterior dağılımların eşleştiği durumlarda tüm dağılımlarda örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde küçülme gözlemlenmiştir. θ parametreleri normal dağılıma sahip veri seti incelenirken tüm örneklem büyüklüklerinde genellikle diğer dağılımlardan daha büyük ortalama RMSE değerleri elde edilmiştir. Bu bulgular değerlendirildiğinde Seong (1990)'un çalışmasından elde ettiği sonuçların şu anki araştırmayı kısmen desteklediği söylenebilir. Stone (1992)'un çalışması incelendiğinde tüm test uzunlukları ve θ parametrelerinin dağılımlarında örneklem büyüklüğü arttıkça ortalama RMSE değerlerinin küçüldüğü görülmektedir. Ayrıca en küçük ortalama RMSE değerleri test uzunluğunun 40 madde ve örneklem büyüklüğünün N=1,000 olduğu veri setleri hariç her zaman θ parametreleri normal dağıldığında elde edilmiştir. Bu bulgular değerlendirildiğinde elde edilen sonuçların şu anki çalışmayı desteklediği söylenebilir. Abdel-fattah (1994)'ın çalışması incelendiğinde test uzunluğunun 20 madde ve θ parametrelerinin normal ve beta dağılımına sahip olduğu veri setlerinde örneklem büyüklüğü arttıkça ortalama MSD değerlerinde küçülme, θ parametrelerinin truncated dağılıma sahip olduğu durumlarda ise örneklem büyüklüğü arttıkça ortalama MSD değerlerinde de büyüme gözlemlenmiştir. Test uzunluğunun 60 maddeden oluştuğu veri setlerinde tüm dağılım türlerinde örneklem büyüklüğü arttıkça ortalama MSD değerlerinde azalma gözlemlenmiştir. Bu bulgular değerlendirildiğinde test uzunluğunun 20 maddeden oluştuğu ve θ parametrelerinin trucated dağılıma sahip olduğu veri seti dışındaki diğer sonuçların şu anki çalışmayı desteklediği söylenebilir. Kieftenbeld ve Natesan (2012)'ın çalışmaları incelendiğinde MML kestirim yönteminde tüm test uzunlukları ve θ parametrelerinin dağılımları için örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde küçülme gözlemlenmiştir. Bu bulgular değerlendirildiğinde elde edilen sonuçların şu anki çalışmayı desteklediği söylenebilir. Boulet (1996)'in çalışması incelendiğinde kullanılan tüm yazılımlardan elde ettiği sonuçlar için θ parametrelerinin normal dağılıma sahip olduğu veri setlerinde tüm test uzunlukları için örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde küçülme gözlemlenmiştir. TESTFACT yazılımının kullanıldığı analizlerde θ parametrelerinin normal dağılmadığı veri setlerinde tüm test uzunlukları için örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde küçülme gözlemlenmiştir. NOHARM yazılımının kullanıldığı analizlerde θ parametrelerinin normal dağılmadığı veri setlerinde tüm test uzunlukları için örneklem büyüklüğündeki artışın ortalama RMSE değerleri üzerinde belli bir etkisi görülmemiştir. Bu veri setlerinde ortalama RMSE değerleri dalgalı bir seyir izlemiştir. Bu bulgular değerlendirildiğinde θ parametrelerinin normal dağıldığı veri setleri şu anki çalışmanın normal dağılımdan elde ettiği sonuçları desteklemektedir. Çarpık dağılımın kullanıldığı veri setlerinin TESTFACT sonuçları da değerlendirildiğinde bu çalışmayı desteklediği görülmektedir. Yoes (1995)'in çalışması incelendiğinde Test 1 için araştırmacı LOGIST yazılımından elde ettiği sonuçlarda tüm test uzunluklarında örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde küçülme elde etmiştir. Aynı koşullar altında ASCAL yazılımının ürettiği sonuçlar incelendiğinde test uzunluğunun sadece 75 maddeden oluştuğu veri setlerinde örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde küçülme görülmüştür. Aynı koşullar altında XCALIBRE yazılımının ürettiği sonuçlar incelendiğinde test uzunluğunun sadece 50 ve 75 maddeden oluştuğu veri setlerinde örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde küçülme görülmüştür. Test 2 için LOGIST yazılımından elde edilen sonuçlar incelendiğinde test uzunluğunun yalnızca 75 ve 100 maddeden oluştuğu veri setlerinde örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde küçülme görülmüştür. Aynı koşullar altında ASCAL yazılımının ürettiği sonuçlar incelendiğinde test uzunluğunun yalnızca 50 ve 75 maddeden oluştuğu veri setlerinde örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde küçülme görülmüştür. Aynı koşullar altına BILOG yazılımının ürettiği sonuçlar incelendiğinde test uzunluğu yalnızca 25 maddeden oluştuğu veri setlerinde örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde küçülme görülmüştür. Aynı koşullar altında XCALIBRE yazılımının ürettiği sonuçlar incelendiğinde test uzunluğunun yalnızca 50 ve 75 maddeden oluştuğu veri setlerinde örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde küçülme görülmüştür. Geriye kalan tüm durumlarda örneklem büyüklüğü arttıkça ortalama RMSE değerlerinde dalgalı bir seyir gözlemlenmiştir. Bu sonuçlar değerlendirildiğinde yalnızca Test 2'den test uzunluğunun 25 maddeden oluştuğu veri setinin BILOG yardımıyla analiz edilmesi sonucunda elde edilen bulguların şu anki çalışmayı desteklediği söylenebilir. Akour ve AL-Omari (2013)'nin çalışmaları incelendiğinde tüm test uzunlukları için örneklem büyüklüğü arttıkça ortalama RMSL değerlerinin küçülme gösterdiği görülmüştür. Bu bulgular değerlendirildiğinde elde edilen sonuçların şu anki çalışmayı desteklediği söylenebilir. Şahin ve Anıl (2017)'ın çalışmaları incelendiğinde 2PL model altında tüm test uzunluklarında örneklem büyüklüğü arttıkça ortalama RMSD değerlerinde dalgalı bir seyir görülmüştür. Test uzunluğunun 10 maddeden oluştuğu veri setleri için yüksek kabul edilen örneklem büyüklüklerinde (N=500, 750, 1,000, 2,000, 3,000) bile bu dalgalı seyir devam etmektedir. Yalnızca test uzunluğunun 30 maddeden oluştuğu veri setlerinde örneklem büyüklüğü N=1,000'den sonra düzenli olarak ortalama RMSD değerlerinde küçülme görülmektedir. Bu bulgular değerlendirildiğinde elde edilen sonuçların şu anki çalışmayı desteklemediği söylenebilir.

Madde güçlük (b) parametreleri için hem örneklem büyüklüğü hem de ÇK manipüle edildiğinde örneklem büyüklüğü arttıkça ÇK=-1,00 ve ÇK=1,00 dışında ortalama Bias değerleri beklendiği gibi sıfıra yakınlaşma eğilimindedir. ÇK=-1,00 ve ÇK=1,00 olan dağılımlarda ise örneklem büyüklüğü arttıkça ortalama Bias değerleri sıfırdan uzaklaşmaktadır. Seong (1990)'un çalışması incelendiğinde prior ve posterior dağılımların eşleştiği durumlarda θ parametrelerinin tüm dağılım türlerinde örneklem büyüklüğü arttıkça ortalama Bias değerlerinin de sıfıra yaklaştığı görülmektedir. Bu bulgular değerlendirildiğinde elde edilen sonuçların şu anki araştırmayı desteklemediği söylenebilir. Stone (1992)'un çalışması incelendiğinde test uzunluğunun 10 ve 40 maddeden oluştuğu veri setlerinde simetrik-basık dağılım dışında θ parametrelerinin normal ve çarpık dağıldığı veri setlerinden b parametreleri için elde edilen ortalama Bias değerlerinde dalgalanma görülmüştür. Test uzunluğunun 20 madde olduğu veri setlerinde ise tüm dağılımlarda örneklem büyüklüğü arttıkça ortalama Bias değerlerinin sıfıra yaklaştığı gözlemlenmiştir. Bu bulgular değerlendirildiğinde test uzunluğunun 10 ve 40 maddeden oluşan veri setlerinden elde edilen sonuçların şu anki araştırmanın sonuçlarını desteklediği söylenebilir. Abdel-fattah (1994)'ın çalışması incelendiğinde MML kestirim yöntemi altında test uzunluğunun 20 madde olduğu ve truncated dağılım dışında θ parametrelerinin normal ve beta dağılımına sahip olduğu veri setlerinde örneklem büyüklüğü arttıkça ortalama Bias değerlerinin karesi sıfıra yaklaşmaktadır. θ parametrelerinin truncated dağılıma sahip olduğu veri setlerinde ise örneklem büyüklüğü arttıkça ortalama Bias değerlerinin karesi sıfırdan uzaklaşmıştır. Test uzunluğunun 60 madde olduğu veri setlerinde ise tüm dağılımlarda örneklem büyüklüğü arttıkça ortalama Bias değerlerinin karesi sıfıra yaklaşmaktadır. Bu bulgular değerlendirildiğinde elde edilen sonuçların şu anki çalışmayı desteklemediği söylenebilir.

# BÖLÜM V

## 5. Sonuçlar ve Öneriler

Bu bölümde araştırmadan elde edilen sonuçlara ve önerilere yer verilmiştir.

### 5.1. Sonuçlar

Bu araştırmada MTK'ye dayalı tek boyutlu 2PL modelde çeşitli koşulların parametre kestirim keskinliğine olan etkileri incelenmiştir. Araştırmada örneklem büyüklüğü (N) ve çarpıklık katsayısı (ÇK) değişken olarak belirlenmiştir. Belirlenen bu değişkenlerin parametre kestirim keskinliğine olan etkilerini incelemek amacıyla veri üretimi gerçekleştirilmiştir. Veriler üretilirken hem değişken olarak belirlenen koşulların hem de sabit tutulan koşulların gerçek durumlarda elde olmasına özen gösterilmiştir. Parametre kestiriminden elde edilen sonuçlar ile edilebilecek değerlerde üretilen (gerçek) parametreler arasındaki keskinliğin derecesine RMSE ve Bias istatistikleri yardımıyla bakılmıştır. Manipüle edilen değişkenlerin parametre kestirim keskinliğine olan önemli etkileri aşağıda belirtilmektedir.

Örneklem büyüklüğünün sabit tutulup yalnızca ÇK'nin değişim gösterdiği durumlarda a parametrelerinin ortalama RMSE değerlerinin ÇK'nin standart normal dağılıma yaklaştıkça küçüldüğü görülmüştür. Aynı zamanda dağılımın hangi yöne olduğunun ortalama RMSE değerlerinde önemli bir değişikliğe yol açmadığı da gözlemlenmiştir. En küçük ortalama RMSE değerlerinin (0,121) θ parametreleri standart normal dağılıma sahipken alındığı en büyük ortalama RMSE değerlerininse (0,254) θ parametrelerinin ÇK=-2,00 olan dağılıma sahipken elde edildiği görülmüştür. Örneklem büyüklüğünün sabit tutulduğu ve yalnızca ÇK'nin değişim gösterdiği durumlarda a parametrelerinin ortalama Bias değerleri ÇK standart normal dağılıma yaklaştıkça sıfıra yaklaşmıştır. Yine ortalama RMSE değerlerine benzer şekilde ortalama Bias değerlerinin dağılımın yönünden önemli derecede etkilenmediği görülmüştür. Sıfıra en yakın ortalama Bias değerlerinin (0,005) θ parametreleri standart normal dağılıma sahipken elde edildiği, sıfıra en uzak ortalama Bias değerlerinin (-0,147) de θ parametreleri ÇK=-2,00 dağılımına sahipken elde edildiği görülmüştür. Elde edilen bu bulguların ışığında sonuçlardaki artış ve azalış yönü beklendiği gibidir ancak N=1000 örneklem büyüklüğü için ortalama RMSE değerleri oldukça yüksek, ortalama Bias değerleri ise ÇK=-2,00 ve ÇK=2,00 hariç oldukça düşük bulunmuştur. Ortalama RMSE değerlerinin bu kadar yüksek olmasının test uzunluğunun veya örneklem büyüklüğünün yetersiz olmasından kaynaklandığı düşünülmektedir.

Örneklem büyüklüğünün sabit olduğu ve yalnızca ÇK'nin değişim gösterdiği durumlarda b parametrelerinin ortalama RMSE değerlerinin ÇK=-1,00 ve ÇK=1,00 olduğu durumlarda ilginç bir şekilde en büyük değerleri ürettiği görülmüştür. Bu alt problemden elde edilen sonuçlar ışığında en küçük ortalama RMSE değerleri (0,093) standart normal dağılımdan elde edilirken en büyük sonuçlarınsa (0,153) ÇK=1,00 olan dağılımdan elde edildiği görülmüştür. Elde edilen bulgular ışığında ÇK'nin mutlak değerce en büyük olduğu dağılım için en yüksek RMSE değerinin elde edilememesi durumunun a, b ve θ parametrelerinin dağılımlarının birbirleriyle olan etkileşiminden kaynaklandığı düşünülmektedir. Örneklem büyüklüğünün sabit olduğu ve yalnızca ÇK'nin değişim gösterdiği durumlarda b parametrelerinin ortalama Bias değerlerinin ortalama RMSE değerlerinden elde edilen sonuçlarla benzerlik taşıdığı görülmüştür. Mutlak değerleri eşit olan en büyük çarpıklık katsayılarından elde edilen ortalama Bias değerleri, ÇK=1,00 ve ÇK=-1,00 dağılım durumlarından elde edilen ortalama Bias değerlerine nazaran daha küçük değerler almıştır. Yine bu durumun yukarıda açıklanan sebepten kaynaklandığı düşünülmektedir. Bu bilgiler ışığında sıfıra en yakın ortalama Bias değerinin (-0,001) standart normal dağılımdan elde edildiği, sıfıra en uzak ortalama Bias değerininse (0,094) ÇK=-1,00 olan dağılımdan elde edildiği görülmüştür. Elde edilen tüm bu bulguların ışığında RMSE ve Bias değerlerindeki artış yönü b parametreleri için anormaldir. Ortalama RMSE değerleri için ÇK'nin sırasıyla -1,00 ve 1,00 olduğu durumlarda en büyük değeri 0,153 ve ÇK'nin sırasıyla -2,00 ve 2,00 olduğu durumlarda en küçük değeri ise 0,136'dır. Buradan yola çıkıldığında bu beklenmeyen durumun maksimum farkı 0,017’dir. Yani aslında bu durum oldukça küçük bir farkı işaret etmektedir ve dikkate alınmayabilir. Benzer bir durum ortalama Bias değerleri için de geçerlidir. ÇK’nin sırasıyla -1,00 ve 1,00 olduğu durumlarda mutlak değerce en büyük değer 0,094, ÇK’nin sırasıyla -2,00 ve 2,00 olduğu durumlarda mutlak değerce en küçük olduğu değer 0,066’dır. Buradaki fark ise 0,028’dir ve ortalama Bias değerleri için oldukça küçük bir değere denk düşmektedir. Sonuç olarak b parametreleri için ortalama RMSE ve Bias değerlerindeki bu beklenmeyen değerlenme göz ardı edilebilir. En sonunda ise b parametrelerinin ortalama RMSE değerleri oldukça yüksek ancak ortalama Bias değerleri oldukça düşük bulunmuştur.

Hem örneklem büyüklüğünün hem de ÇK’nin serbest bırakıldığı durumlarda a parametrelerinin ortalama RMSE değerlerinin tüm dağılım türlerinde örneklem büyüklüğü arttıkça düştüğü gözlemlenmiştir. Aynı zamanda θ parametrelerinin dağılımı standart normal dağılıma yaklaştıkça elde edilen ortalama RMSE değerlerinde küçülme görülmüştür. Bu bilgilere ek olarak çarpıklık katsayılarının pozitif ya da negatif olmasının ortalama RMSE değerlerine büyük bir etkisinin olmadığı görülmüştür. En küçük ortalama RMSE değerleri (0,088) standart normal dağılımdan N=2,000 örneklem büyüklüğünde elde edilmiştir. En büyük ortalama RMSE değerleri (0,310) ise ÇK=2,00 olan dağılımdan N=250 örneklem büyüklüğünden elde edilmiştir. Elde edilen bulgular ışığında sonuçlar beklendiği gibidir. Çarpıklık katsayıları ve örneklem büyüklükleri birlikte değerlendirildiğinde a parametreleri için yalnızca θ parametreleri standart normal dağılıma sahipken ve N=2,000 örneklem büyüklünde kabul edilebilir sonuçlar elde edilmiştir. Hem örneklem büyüklüğünün hem de ÇK’nin serbest olduğu durumlarda a parametrelerinin ortalama Bias değerlerinin dağılımların çarpıklığından oldukça ciddi biçimde etkilendiği gözlemlenmiştir. Beklendiği üzere θ parametrelerinin dağılımı standart normal dağılıma yaklaştıkça ortalama Bias değerleri de sıfıra yaklaşmaktadır. Ancak dağılımlar kendi içlerinde değerlendirildiğinde örneklem büyüklüğü arttıkça ortalama Bias değerlerinin de sıfıra yaklaştığı tek dağılımın standart normal dağılım olduğu, diğer dağılımların genelde örneklem büyüklüğü arttıkça ortalama Bias değerlerinin de sıfırdan uzaklaşma eğiliminde olduğu görülebilir. Bu oldukça ilginç bir bulgudur ve bu duruma a, b ve θ parametrelerinin dağılımlarının birbirleriyle olan etkileşimlerinin veya madde sayısının yetersiz kalmasının sebep olduğu düşünülmektedir. Ancak yine elde edilen bulgulardan yola çıkarak bu uzaklaşmanın maksimum olduğu noktada farkın 0,019 olduğu görülmüştür. Yani aslında standart normal dağılım haricinde kalan dağılımların kendi içlerinde yakın Bias değerleri ürettiği söylenebilir. Bu bilgilere ek olarak sıfıra en yakın ortalama Bias değerlerinin (0,003) standart normal dağılımdan N=2,000 örneklem büyüklüğünde elde edildiği, sıfıra en uzak ortalama Bias değerlerininse (-0,147) ÇK=-2,00 olan dağılımdan N=1,000 örneklem büyüklüğünden elde edildiği görülmüştür. Tüm bu bilgiler değerlendirildiğinde ÇK=-2,00 ve ÇK=2,00 dağılımları dışında tüm örneklem büyüklükleri için Bias değerlerinin sıfıra oldukça yakın değerler ürettiği söylenebilir.

Hem örneklem büyüklüğünün hem de ÇK'nin manipüle edildiği durumlarda beklenildiği gibi b parametrelerinin ortalama RMSE değerlerinin tüm dağılım türlerinde örneklem büyüklüğü arttıkça azalma eğiliminde olduğu görülmüştür. Ancak ilginç bir şekilde ÇK=-2,00 ve ÇK=2,00 olan dağılımların ÇK=-1,00 ve ÇK=1,00 olan dağılımlarından N=250 örneklem büyüklüğü dışındaki tüm örneklem büyüklüklerinde daha küçük ortalama RMSE değerleri üretmiştir. Ancak normal dağılmayan dağılımlar arasından elde edilen bu ortalama RMSE değerlerine bakıldığında en büyük farkın 0,024 olduğu görülmektedir. Bu nedenle aslında çarpık dağılımların birbirlerine yakın değerler ürettiği söylenebilir. Bu bilgilere ek olarak çarpıklık katsayılarının pozitif ya da negatif olmasının ortalama RMSE değerleri üzerinde büyük bir etkisinin olmadığı da görülmüştür. En küçük ortalama RMSE değerleri (0,065) standart normal dağılımdan N=2,000 örneklem büyüklüğünde elde edilirken en büyük ortalama RMSE değerleri (0,221) ÇK=2,00 olan dağılımdan N=250 örneklem büyüklüğünden elde edilmiştir. Elde edilen sonuçların ışığında RMSE değerlerinde görülen küçülme beklendiği gibidir ancak yalnızca standart normal dağılımda N=1,000 ve N=2,000 örneklem büyüklüklerinde kabul edilebilir RMSE değerleri (sırasıyla 0,093 ve 0,065) elde edilmiştir. Hem örneklem büyüklüğünün hem de ÇK'nin serbest olduğu durumlarda b parametrelerinin ortalama Bias değerleri ÇK=1,00 ve ÇK=-1,00 olan dağılımlar hariç örneklem büyüklüğü arttıkça sıfıra yaklaşma eğiliminde olmuşlardır. ÇK=-1,00 ve ÇK=1,00 olan dağılımlarda ise örneklem büyüklüğü arttıkça ortalama Bias değerleri sıfırdan uzaklaşma eğilimindedirler. ÇK=-1,00 ve ÇK=1,00 dışında sonuçlar beklenildiği gibidir. ÇK=-1,00 ve ÇK=1,00 olan dağılımlarda gözlenen bu ilginç sonuçlara a, b ve θ parametrelerinin dağılımlarının birbirleriyle olan ilişkisinin neden olduğu düşünülmektedir. Ancak tüm örneklem büyüklüklerinde bu dağılımlardan elde edilen sonuçlar incelendiğinde ÇK=-1,00 için ortalama Bias değerlerinin maksimum farkının 0,022, ÇK=1,00 için ortalama Bias değerlerinin maksimum farkının 0,024 olduğu görülmektedir. Bu farklar doğrultusunda aslında ÇK=-1,00 ve ÇK=1,00 dağılımlarından elde edilen ortalama Bias değerlerinin birbirlerine çok yakın olduğu ve elde edilen sonucun ışığında bu artışın göz ardı edilebileceği söylenebilir. Ayrıca tüm dağılım türleri ve örneklem büyüklükleri için elde edilen ortalama Bias değerleri kabul edilebilir sınırlar dâhilindedir. Standart normal dağılım dışında diğer dağılım türleri kendi aralarında karşılaştırıldığında ÇK'nin değişiminin ortalama Bias değerleri üzerinde büyük bir etkisinin olmadığı görülmüştür. Aynı zamanda mutlak değerleri eşit olan çarpıklık katsayılarının pozitifi ya da negatif olma durumlarının da ortalama Bias değerleri üzerinde büyük bir etkilerinin olmadığı gözlemlenmiştir. Sıfıra en yakın ortalama Bias değerleri (-0,00002) standart normal dağılımdan N=2,000 örneklem büyüklüğünden elde edilmiştir. Sıfıra en uzak ortalama Bias değerleri (0,099) ise ÇK=-1,00 olan dağılımda N=2,000 örneklem büyüklüğünde gözlemlenmiştir. Tüm bu bilgiler değerlendirildiğinde b parametrelerinin ortalama Bias değerlerinin çarpıklık katsayılarına ya da örneklem büyüklüğüne bakılmaksızın sıfıra oldukça yakın ve kabul edilebilir eşiklerde olduğu söylenebilir.

Tüm bu sonuçlar değerlendirildiğinde bu simülasyon koşullarında yalnızca θ parametreleri standart normal dağılırken N=2,000 örneklem büyüklüğünde madde parametre kestirim işlemi gerçekleştirmek uygun görünmektedir.

## 5.2. Öneriler

### 5.2.1. Araştırmanın sonuçlarına yönelik öneriler

Araştırmadan elde edilen bulgulara dayanarak R programlama dilinde MML kestirim yöntemi kullanılarak gerçekleştirilecek bir parametre kestirim çalışmasında θ parametrelerinin ÇK'lerinin 0,00 olduğu durumlarda test uzunluğunun en az 30 maddeden oluşması ve örneklem büyüklüğünün ise en az N=2,000 olması önerilebilir.

Aksi durumlarda gerçekleştirilecek tüm çalışmalarda θ parametrelerinin dağılımlarının kontrol edilmesi önerilmektedir.

5.2.2. Gelecek araştırmalara yönelik öneriler

Gelecekte gerçekleştirilebilecek parametre kestirim keskinliği çalışmalarına yönelik aşağıdaki öneriler dikkate alınabilir.

1. Bu araştırmada test uzunluğu sabit tutulmuştur. Gelecekte yapılabilecek çalışmalarda test uzunluğunu manipüle edilebilir.
2. Bu araştırmada a ve b parametreleri tek bir dağılım türünden üretilmiştir. Alanyazın araştırılarak a ve b parametreleri farklı dağılımlardan üretilerek bir araştırma gerçekleştirilebilir. Bunun yanında a ve b parametreleri için birden fazla dağılım kullanılabilir.
3. Bu araştırma tek boyutlu 2PL model altında gerçekleştirilmiştir. Ayrıca veriler ikili puanlama (dichotomous) yöntemiyle puanlanmıştır. Gelecek çalışmalarda tek boyutlu olan farklı lojistik modeller, çok boyutlu MTK modelleri ya da çoklu puanlama (polytomous) yöntemiyle puanlanan veriler için uygun olan MTK modelleri kullanılabilir.
4. R programlama dilinde gerçekleştirilen bu çalışmada *mirt* paketi kullanılmıştır. Gelecek çalışmalarda aynı programlama dilinde *ltm* paket programı ya da farklı yazılımlar (BILOG, LOGIST vs.) kullanılabilir.

Araştırmada MML kestirim kestirim yöntemi kullanılmaktadır. Gelecek araştırmalarda başka kestirim yöntemleri kullanılabilir.

# KAYNAKÇA


Abdel-fattah, A. A. (1994, Nisan). *Comparing BILOG and LOGIST Estimates for Normal, Truncated Normal, and Beta Ability Distributions.* Amerikan Eğitim Araştırmaları Derneği yıllık toplantısında sunulan bildiri, New Orleans, LA. (ERIC Belge Çoğaltma Servisi No. ED374158)

Akour, M. & AL-Omari, H. (2013). Empirical Investigation of the Stability of IRT Item-Parameters Estimation. *International Online Journal of Educational Sciences,* 5(2), 291-301. Retrived from https://eis.hu.edu.jo/deanshipfiles/pub106314725.pdf

Allen, M. J. & Yen, W. M. (1979). *Introduction to Measurement Theory.* Belmont, CA: Wadsworth.

Atılgan, H., Kan, A. & Doğan, N. (2009). *Eğitimde Ölçme ve Değerlendirme* (4. baskı). Ankara: Anı Yayıncılık.

Baker, F. B. (2001). *The Basics of Item Response Theory* (2. baskı). USA: ERIC Clearinghouse on Assessment and Evaluation.

Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması.* Ankara: Ösym Yayınları.

Berkson, J. (1951, Aralık). Why I Prefer Logits to Probits. *Biometrics*, 7(4), 327-339, DOI: 10.2307/3001655.

Birnbaum, A. (1967, Şubat). Statistical Theory for Logistic Mental Test Models with a Prior Distribution of Ability. *Research Bulletin Series 67-12,* Princeton, N. J.: Educational Testing Service, DOI: 10.1016/0022-2496(69)90005-4.

Bond, T. G. & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (3. baskı). New York, NY: Routledge.

Boulet, J. R. (1996). *The Effect of NonNormal Ability Distributions on IRT Parameter Estimation Usin Full-Information and Limited-Information Methods.* Yayımlanmamış doktora tezi, University of Ottowa/Faculty of Education, Ottowa.

Bulut, O. & Sünbül, Ö. (2017). Monte Carlo Simulation Studies in Item Response Theory with the R Programming Language. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi,* 8(3), 266-287, DOI: 10.21031/epod.305821.

Chalmers, R. P. (2012). mirt:A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6), 1-29, DOI: 10.18637/jss.v048.i06.

Crocker, L. & Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. Mason, Ohio: Cengage Learning.

Cronbach, L. J. (1990). *Essentials of Psychological Testing* (5. baskı). New York, NY: Harper & Row.

de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York, NY: The Guilford Press.

DeMars. C. (2003). Sample Size and the Recovery of Nominal Response Model Item Parameters. *Applied Psyhcological Measurement*, 27(4), 275-288, DOI: 10.1177/0146621603027004003.

DeMars, C. (2010). *Item Response Theory*. New York, NY: Oxford University Press.

Drasgow, F. (1989, Mart). An Evaluation of Marginal Maximum Likelihood Estimation for the Two-Parameter Logistic Model. *Applied Psychological Measurement*, 13(1), 77-90, DOI: 10.1177/014662168901300108.

Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of Educational Measurement* (5. baskı). New Delhi: Prentice-Hall.

Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Eser, D. Ç. & Gelbal, S. (2015). Farklı Boyutluluk Özelliklerindeki Basit ve Karmaşık Yapılı Testlerin Çok Boyutlu Madde Tepki Kuramına Dayalı Parametre Kestirimlerinin İncelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(2), 331-350, DOI: 10.21031/epod.80315.

Feinberg, R. A. & Rubright, J. D. (2016, Haziran). Conducting Simulation Studies in Psychometrics. *Educational Measurement: Issues and Practice,* 35(2), 36-49, DOI: 10.1111/emip.12111

Field, A. (2009). *Discovering Statistics Using SPSS* (3. baskı). London: Sage Publications.

Finch, W. H. & French, B. F. (2019). *Educational and Psychological Measurement.* New York, NY: Routledge.

Gilbert, N. (1999). Simulation: A New Way of Doing Social Science. *American Behavioral Scientist,* 42(10), 1485-1487, DOI: 10.1177/0002764299042010002.

Gulliksen, H. (1950). *Theory of Mental Tests.* New York, NY: John Wiley & Sons.

Haley, D. C. (1952, Ağustos). *Estimation of the DosageMortality Relationship When the Dose is Subject to Error.* (Donanma araştırmaları için hazırlanmış teknik rapor No:15). Stanford, CA: Applied Mathematics and Statistics Laboratory, Stanford University.

Hambleton, R. K. & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications.* New York, NY: Springer.

Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory* (1. baskı). Newbury Park, CA: Sage Publications.

Harwell, M., Stone, C. A., Hsu, Tse-Chi. & Kirişçi, L. (1996, Haziran). Monte Carlo Studies in Item Response Theory. *Applied Psychological Measurement,* 20(2), 101-125, DOI: 10.1177/014662169602000201.

Hulin, C. L., Lissak, R. I. & Drasgow, F. (1982). Recovery of Two- and Three-Parameter Logistic Item Characteristic Curves: A Monte Carlo Study. *Applied Psychological Measurement,* 6(3), 249-260, DOI: 10.1177/014662168200600301.

Karadavut, T. (2017). Estimation of Item Response Theory Models When Ability is Uniformly Distributed. *The Eurasia Proceedings of Educational and Social Science,* 7(), 30-37. Retrived from https://dergipark.org.tr/tr/pub/epess/issue/30770/332649

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, DOI: 10.1177/0146621612446170.

Kilmen, S. (2015). *Eğitim Araştırmacıları İçin SPSS Uygulamalı İstatistik* (2. baskı). Ankara: Edge Akademi.)

Kim, K. Y. & Lee, W. C. (2017). The Impact of Three Factors on the Recovery of Item Parameters for the Three-Parameter Logistic Model. *Applied Measurement in Education,* 30(3), 228-242, DOI: 10.1080/08957347.2017.1316274.

Kirişçi, L., Hsu, Tse-Chi. & Yu, L. (2001, Haziran). Robustness of Item Parameter Estimation Programs to Assumptions of Unidimensionality and Normality. *Applied Psychological Measurement,* 25(2), 146-162, DOI: 10.1177/01466210122031975.

Köklü, N., Büyüköztürk, Ş. & Bökeoğlu, Ö. Ç. (2006). *Sosyal Bilimler İçin İstatistik* (2. baskı). Ankara: Pegem Yayıncılık.

Köse, İ. A. (2010). *Madde Tepki Kuramına Dayalı Tek ve Çok Boyutlu Modellerin Test Uzunluğu ve Örneklem Büyüklüğü Açısından Karşılaştırılması.* (Yayımlanmamış doktora tezi). Ankara Üniversitesi/Eğitim Bilimleri Enstitüsü, Ankara.

Lord, F. M. (1968). An Analysis of the Verbal Scholastic Aptitude Test Using Birnbaum's Three-Parameter Logistic Model. *Educational and Psychological Measurement,* 28(4), 989-1020, DOI: 10.1177/001316446802800401.

Lord, F. M. (1974, Haziran). Estimation of Latent Ability and Item Parameters When There are Omitted Responses. *Psyhcometrika,* 39(2), 247-264, DOI: 10.1007/BF02291471.

Lord, F. M. & Novic, M. R. (2008). *Statistical Theories of Mental Test Scores.* USA: Information Age Publishing.

Luecht, R. & Ackerman, T. A. (2018, Ocak). A Technical Note on IRT Simulation Studies: Dealing With Truth, Estimates, Observed Data, and Residuals. *Educational Measurement: Issues and Practice,* 37(3), 65-76, DOI: 10.1111/emip.12185.

Marso, R. N. & Pigge, F. L. (1988, Nisan). *An Analysis of Teacher-Made Tests: Testing Practices, Cognitive Demands, and Item Construction Errors.* Ulusal Eğitimde Ölçme Kurulu (NCME) Yıllık Toplantısında sunulmuş bildiri, New Orleans, LA.

Mehrens, W. A. & Lehmann, I. J. (1991). *Measurement and Evaluation in Education and Psychology* (4. baskı). Belmont, CA: Wadsworth/Thomson Learning.

Meyers, L. S. & Grossen, N. E. (1974). *Behavioral Science: Theory, Procedure, and Design.* USA: W. H. Freeman and Company.

Miller, M. D., Linn, R. L. & Gronlund, N. E. (2009). *Measurement and Assesment in Teaching* (10. baskı). New Jersey: Pearson Education.

Olmuş, H., Nazman, E. & Erbaş, S. (2016). An Evaluation of the Two Parameter (2-PL) IRT Models Through a Simulation Study. *Gazi University Journal of Science*, 30(1), 235-249. Retrived from https://dergipark.org.tr/tr/pub/gujs/issue/28464/303387

Placko, D. (2007). *Fundamentals of Instrumentation and measurement.* Newport Beach, CA: Iste.

Price, L. R. (2017). *Psychometric Methods: Theory into Practice.* New York, NY: The Guilford Press.

Reckase, M. D. (2009). *Multidimensional Item Response Theory.* New York, NY: Springer.

Reise, S. P. & Yu, J. (1990). Parameter Recovery in the Graded Response Model Using MULTILOG. *Journal of Educational Measurement*, 27(2), 133-144, DOI: 10.1111/j.1745-3984.1990.tb00738.x.

Riley, B. B. & Carle, A. C. (2012). Comparision of Two Bayesian Methods to Detect Mode Effects Between Paper-Based and Computerized Adaptive Assessments: A Preliminary Monte Carlo Study. *BMC Medical Research Methodology, 12:124*, DOI: 10.1186/1471-2288-12-124.

Sass, D. A., Schmitt, T. A. & Walker, C. M. (2008, Nisan). Estimating Non-Normal Latent Trait Distributions within Item Response Theory Using True and Estimated

Item Parameters. *Applied Measurement in Education*, 21(1), 65-88, DOI: 10.1080/08957340701796415.

Seong, Tae-Je. (1990, Eylül). 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, DOI: 10.1177/014662169001400307.

Shultz, K. S. & Whitney, D. J. (2005). *Measurement Theory in Action: Case Study and Exercises.* Thousand Oaks, CA: Sage Publications.

Stone, C. A. (1992, Mart). Recovery of Marginal Maximum Likelihood Estimates in the Two-Parameter Logistic Response Model: An Evaluation of MULTILOG. *Applied Psychological Measurement*, 16(1), 1-16, DOI: 10.1177/014662169201600101.

Svetina, D., Valdiva, A., Underhill, S., Dai, S. & Wang, X. (2017). Parameter Recovery in Multidimensional Item Response Theory Models Under Complexity and Nonnormality. *Applied Psychological Measurement*, 41(7), 530-544, DOI: 10.1177/0146621617707507.

Swaminathan, H. & Gifford, J. A. (Nisan,1979). *Estimation of Parameters in the Three-Parameter Latent Trait Model.* Amerikan Eğitim Araştırmaları Birliği (AERA) ve Ulusal Eğitimde Ölçme Kurulu (NCME)'nun düzenlediği "Pratik Ölçme Sorunlarını Çözme Aracı Olarak Gizil Özellikler Modellerinin Keşfi (Explorations of Latent Trait Models as a Means of Solving Practical Measurement Problems)" başlıklı sempozyumda sunulmuş bildiri, San Francisco, CA.

Şahin, A. & Anıl, D. (2017). The Effects of Test Length and Sample Size on Item Parameters in Item Response Theory. *Educational Science: Theory & Practice,* 17(1), 321-335, DOI: 10.12738/estp.2017.1.0270.

Şahin, M. G. & Yıldırım, Y. (2018). The Examination of Item Difficulty Distribution, Test Lenght and Sample Size in Different Ability Distribution. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi,* 9(3), 277-294, DOI: 10.21031/epod.385000.

Turgut, M. F. & Baykul, Y. (2015). *Eğitimde Ölçme ve Değerlendirme* (7.baskı). Ankara: Pegem Akademi.

Uysal, İ., Ertuna, L., Ertaş, F. G. & Kelecioğlu, H. (2019). Performance Based on Ability Estimation of the Methods of Detecting Differential Item Functioning: A Simulation Study. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi,* 10(2), 133-148, DOI: 10.21031/epod.534312.

van der Linden, W. J. (2016). *Handbook of Item Response Theory: Three Volume Set.* Boca Raton, FL: CRC Press.

van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory.* New York, NY: Springer.

Wichmann, B. A. & Hill, I. D. (1982). An Efficient and Portable Pseudo-random Number Generator. *Journal of the Royal Statistical Society,* 31(2), 188-190, DOI: 10.2307/2347988.

Wright, B. D. & Stone, M. H. (1979). *Best Test Design.* Chicago, IL: Mesa Press.

Xu, X. & Jia, Y. (2011). *The Sensitivity of parameter estimates to the latent ability distribution* (Research Report 11-40). Princeton, NJ: Educational Test Service.

Yen, W. M. (1987, Haziran). A Comparison of the Efficiency and Accuracy of BILOG and LOGIST. *Psychometrika,* 52(2), 275-291, DOI: 10.1007/BF02294241.

Yıldırım, Y. (2015). *Derecelendirilmiş Tepki Modeli Temelli Parametre Kestiriminde Normalliğin İhlalinin Ölçme Kesinliğine Etkisi.* Yayımlanmamış Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Yoes, M. E. (1995). *An Update Comparison of Microcomputer-Based Item Parameter Estimation Procedures Used with the 3-Parameter IRT Model* (ASC Technical Report 95-1). St. Paul, MN: Assessment Systems Corporation.

Zwinderman, A. H. & van den Wollenberg, A. L. (1990). Robustness of Marginal Maximum Likelihood Estimation in the Rasch Model. *Applied Psychological Measurement*, 14(1), 73-81, DOI: 10.1177/014662169001400107.

# EKLER

**EK 1.** Çalışmada Kullanılan R Kodları

```
# kodların çalışması için mirt ve moments paketlerinin yüklü olması gerekmektedir.
itemrecovery<- function(nitem, samplesize, seed){
  set.seed(seed)

  item_disc<- as.matrix(round(rlnorm(nitem, meanlog = 0.0, sdlog = 0.5), 3), ncol=1)
#madde ayırt edicilik parametrelerini üretir.

  item_diff<- as.matrix(round(rnorm(nitem, mean = 0,sd=1 ),3), ncol=1) # madde güçlük
parametrelerini üretir.
  #Aşağıda beta dağılımında istenen çarpıklık katsayısını elde etmek için gereken
katsayılar bulunmaktadır.
  #ÇK= 1 için rbeta(1.45,9) (ability-0.14)*10.7 2k, 1k, 500 ve 250 örneklem
  #ÇK= 2 için rbeta(0.8,21) (ability-0.035)*25.7 2k, 1k, 500 ve 250 örneklem
  #ÇK= -1 için rbeta(9,1.45) (ability-0.86)*10.7 2k, 1k, 500 ve 250 örneklem
  #ÇK= -2 için rbeta(21,0.8) (ability-0.965)*25.7 2k, 1k, 500 ve 250 örneklem
  repeat{
    ability<- as.matrix(round(rbeta(samplesize,0.8,21),3),ncol=1) #istenen çarpık dağılımı
üretir ancak beta dağılımı bunu 0 ve 1 sayıları arasında yaptığından ranjı genişletmek
gerekmektedir.

    sability<- matrix(ncol = 1)
```

```
  for(i in 1:samplesize){ #ranjı genişletilmiş dağılım

    sability[i]<- as.matrix(round(((ability[i]-0.035)*25.7),3),ncol=1) #ranjı genişletilmiş dağılım burada üretilmektedir.

    sability<- as.matrix(sability,ncol=1) #verileri analiz için matrix içinde tek sütunda toplar

  }

if(round(skewness(sability),1)==2&&round(sd(sability),1)==1&&round(mean(sability),1)== 0){ #istenilen özellikleri sağladığında kod döngüyü durdurur ve bir sonraki aşamaya geçer aksi takdirde başa döner.

    break

  }else{

    next

  }

 }

 dat<- as.matrix(simdata(a=item_disc,d=item_diff,N=samplesize,itemtype = "dich",Theta = sability),ncol=nitem) #mirt paketi üzerinden üretilmiş olan parametreleri kullanarak 0-1 cevap matrisine çevirir.
```

```
model2pl<- mirt(data = dat, model=1, itemtype = "2PL", SE=TRUE, verbose = FALSE) # üretilen cevap matrixinden parametre kestirimi yapar.

parameters <- as.data.frame(coef(model2pl, simplify=TRUE)$items)# kestirilen parametre değerleri buraya aktarılır.

bias.item_disc<- round(mean(parameters[,1]-item_disc), 3) #her replikasyonda madde ayırt edicilik (a) parametrelerinin üretilen değerleri ile kestirilen değerleri arasında Bias değeleri hesaplanır.

bias.item_diff<- round(mean(parameters[,2]-item_diff), 3) #her replikasyonda madde güçlük (b) parametrelerinin üretilen değerleri ile kestirilen değerleri arasında Bias değerleri hesaplanır.

rmse.item_disc<- round(sqrt(mean((parameters[,1]-item_disc)^2)), 3) #her replikasyonda madde ayırt edicilik (a) parametrelerinin üretilen değerleri ile kestirilen değerleri arasında RMSE değerleri hesaplanır.

rmse.item_diff<- round(sqrt(mean((parameters[,2]-item_diff)^2)), 3) #her replikasyonda madde güçlük (b) parametrelerinin üretilen değerleri ile kestirilen değerleri arasında RMSE hesaplanır.

results<-data.frame(sample.size=samplesize, nitem=nitem, bias.item_disc=bias.item_disc, bias.item_diff=bias.item_diff, rmse.item_disc=rmse.item_disc, rmse.item_diff=rmse.item_diff) #sonuçları buraya aktarmak üzere tanımlanmıştır.

return(results)
```

```
}

myseed <- sample.int(n = 1000000, size = 100) # seed ve iterasyonun belirlendiği kısım
buradan değiştirilebilir.

write.csv(myseed, "2000~+2 simulation seeds.txt", row.names = FALSE)# seedi dışarı
aktarır.

results <- data.frame(samplesize=0, nitem=0, bias.item_disc=0, bias.item_diff=0,
rmse.item_disc=0, rmse.item_diff=0) #sonuçları aktarmak için tanımlanmıştır.

for (i in 1:length(myseed)) {

  results[i,] <- itemrecovery(nitem = 30, samplesize = 2000, seed = myseed[i]) #analiz ve
sonuçların aktarılması işi burada gerçekleştirilir madde sayısı ve örneklem büyüklüğü
buradan gerçekleştirilmektedir.

}

write.csv(results,"2000~+2 sim results.txt", row.names = FALSE) #sonuçları dışarıya
kaydeder.
```

# ÖZGEÇMİŞ

**İsmail Başaran:** 1994 Bolu doğumludur. İlköğretimini Bolu'nun Gerede ilçesinde ortaöğretimini Düzce'de tamamlamıştır. 2016 yılında Kastamonu Üniversitesi Türkçe Öğretmenliği Lisans bölümünü, 2020 yılında Bolu Abant İzzet Baysal Üniversitesi Eğitimde Ölçme ve Değerlendirme Yüksek Lisans programını tamamlamıştır. 2017 yılından itibaren çeşitli özel kurumlarda çalışmaktadır.

**İletişim Adresleri**

e-posta: basaranismail1414@gmail.com

Telefon: +905387829388