



Original article

2 Parametrelili Lojistik Modelde Normal Dağılım İhlalinin Madde Parametre Kestirimine Etkisinin İncelenmesi

Investigation of the Effect of Different Ability Distributions on Item Parameter Estimation Under Two-Parameter Logistics Model

İbrahim Alper Köse ^{a, *} & İsmail Başaran ^b

^aDepartment of Educational Sciences, Faculty of Education, Bolu Abant İzzet Baysal University, Bolu, Turkey

^bMinistry of Education, Turkey

Özet

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 parametrelili 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 xi 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

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

* Corresponding author:

İbrahim Alper Köse, Department of Educational Sciences, Faculty of Education, Bolu Abant İzzet Baysal University, Bolu, Turkey.
Email: i.alper.kose@gmail.com

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) xiii 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.

Received: 07 May 2021 * **Accepted:** 21 June 2021 * **DOI:** <https://doi.org/10.29329/dmer.2021.285.1>

GİRİŞ

Psikolojik yapılar örneğin matematik yeteneği gibi doğrudan gözlemlenip ölçülemediğinden bu yapıları dolaylı yoldan ölçebilecek araçlar 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 kullanılanıdır (Marso ve Piggie, 1988).

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 (Finch ve French 2019). Son yıllarda ise bu ölçme araçlarından elde edilen veri setlerinin MTK modelleri ile analiz edilmesi gittikçe yaygınlaşmaktadır. Ancak ölçme aracının uygulandığı çalışma grubundan elde edilen sonuçların normal dağılmıyor olması normal ogive MTK modellerinde bulunan normallik varsayımını ihlal etmektedir.

Stone (1992), yapmış olduğu çalışmada 2PL model altında MULTILOG yazılımını değerlendirmeye çalışmıştır. Bu amaç doğrultusunda farklı test uzunluğu (10, 20, 40 madde), örneklem büyüklüğü (250, 500, 1000), θ dağılımı (normal dağılım, pozitif çarpık ($\text{ÇK} = 0.75$, $\text{BK} = 0.0$) ve basıklık

(ÇK = 0, BK = -1.0) değişkenlerini manipüle ederek parametre kestirim keskinliğini incelemiştir. Çalışma incelendiğinde bu araştırmadaki gibi normal dağılımın diğer dağılımlardan daha iyi RMSE değerleri verdiği, a ve b parametrelerinin kestiriminden elde edilen RMSE ve Bias değerlerinin bu çalışmayla hemen hemen paralel yönde değerler ürettiği görülmüştür.

Seong (1990), çalışmasında önsel (prior) θ dağılımlarının gerçek θ dağılımlarıyla uyuşmadığında madde ve θ parametrelerinin kestirimindeki hassasiyeti 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. Araştırmacının ulaştığı sonuçlar doğrultusunda bu çalışmayla neredeyse tamamen zıt olduğu görülmüştür.

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. Çalışma incelendiğinde tüm test uzunluklarında örneklem büyüdükçe elde edilen RMSE değerlerinin küçüldüğü bu nedenle bu çalışmayı desteklediği söylenebilir.

Boulet (1996), parametre kestiriminde full-information (FI) ve limited-information (LI) metodlarının kullanıldığı durumlarda θ parametrelerinin çarpık dağılımının parametre kestirim keskinliğine etkisini incelemek istemiştir. Bu amaç doğrultusunda 2PL model üzerinde çalışan araştırmacı 15, 30, 45 ve 60 maddelik testlerin yanında 250, 500, 1,000 ve 10,000 örneklem büyüklükleri kullanılmıştır. Araştırma incelendiğinde bazı durumların bu çalışmayı desteklediği görülürken bazılarının desteklemediği ortaya çıkmıştır.

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 kullanmıştır. Araştırma incelendiğinde bazı durumların bu çalışmayı desteklediği görülürken bazılarının desteklemediği ortaya çıkmıştır.

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 ve R programlama dilinde nasıl performans göstereceğine cevap olacağı için önemlidir.

Ölçme Keskinliği

Eğitim ve psikoloji gibi dolaylı ölçüm işlemlerinin yapıldığı bilim dallarında uygulamalı ya da deneysel bilim dallarında olduğu gibi ölçme işlemlerine karışan hataların kestiriminde kesin bir yöntem

ya da katsayı belirlenemez. Eğitimde ve psikolojide ölçme işlemine karışan hatalar ölçme işleminin çalışma grubuna tekrar tekrar uygulanması ile tespit edilebilir (Turgut ve Baykul, 2015). 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 anlaşılabilir. 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, \hat{X}_i 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.

YÖNTEM

Araştırmanın Modeli

Bu araştırma, farklı örneklem büyüklüklerinin ve farklı yetenek düzeyi (θ) dağılımlarının kullanılarak 2 parametrelili lojistik (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 betimleme modelidir.

Simülasyon Faktörleri ve Koşulları

Bu çalışmada, 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.

Araştırmada manipüle edilen değişkenlerden ilki θ dağılımlarıdır. 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, 1000 ve 2000 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). 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 diğer modellerden daha tutarlı kestirimler vermesi ve bununla birlikte gerçek durumlara olabildiğince yakın olmasıdır (de Ayala, 2009).

Sonuç olarak 5 adet θ dağılımı ve 4 adet örneklem büyüklüğünün kullanıldığı bu çalışmada toplam $5 \times 4 = 20$ koşul incelenmiştir. Araştırma 100 replikasyon sayısı ile gerçekleştirilmiştir. Yani bu 20 koşulun her biri 100 kez tekrarlanmıştır ve bu işlemlerin sonucunda toplam $20 \times 100 = 2000$ veri seti elde edilmiştir. Tablo 1'de simülasyon koşulları gösterilmiştir..

Tablo 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= 1000	-2.00
	-1.00
	0.00
	1.00
	2.00

Verilerin Üretilmesi

Bu çalışmada veri setleri 2PL model altında yapay olarak üretilmiştir. 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.

Bu çalışmada kullanılan Monte Carlo (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; Sünbül ve Bulut, 2017; Feinberg ve Rubright, 2016).

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 $a \sim \ln N(0,0, 0,5)$, b parametresi standart normal dağılımdan rastgele $b \sim N(0,0, 1,0)$ üretilmiştir (Seong, 1990; Feinberg ve Rubright, 2016; Bulut ve Sünbül, 2017). Madde parametrelerini üretmek için RStudio yazılımında R programlama dili kullanılmıştır.

Nihayetinde madde parametrelerinin üretimi için kullanılan parametreler Tablo 2’de görülebilir.

Tablo 2. Madde parametrelerini (a, b) üretmek için kullanılan parametre değerleri

Madde Parametreleri	Kullanılan Parametre Değerleri	
	Ortalama	Standart Sapma
Madde Ayırt Edicilik (a)	0	0.5
Madde Güçlük (b)	Ortalama	Standart Sapma
	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 (-2,0, -1,0) pozitif çarpıklık katsayısına (1,0, 2,0) doğru sırasıyla $\theta \sim \text{BETA}(21,0, 0,8; 9,0, 1,45; 1,45, 9,0; 0,8, 21,0)$ katsayıları kullanılmıştır. Ardından istenilen θ değerlerine ulaşmak amacıyla üretilen her bir değerden $\text{ÇK} = -2,00$ durumunda 0,965 çıkarılmış ve elde edilen sonuç 25,7 ile çarpılmıştır bu işlem diğer çarpıklık katsayılarının üretildiği durumlarda $\text{ÇK} = -1,00$ için 0,86 değerinin çıkarılıp 10,7 ile çarpılarak, $\text{ÇK} = 1,00$ için 0,14 değeri çıkarılıp 10,7 ile çarpılarak, $\text{ÇK} = 2,00$ 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. Standart normal dağılıma $N(0,1)$ sahip yetenek (θ) parametreleri de dahil olmak üzere θ parametreleri RStudio yazılımında R programlama dili kullanılarak üretilmiştir.

Sonuç olarak kullanılan çarpık dağılımlar, bu dağılımları üretmek için kullanılan katsayılar ve istenilen özelliklere sahip olması için gerekli olan çıkarma ve çarpma işlemlerinde kullanılan değerler Tablo 3'te görülebilir.

Tablo 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	Standart Sapma		
ÇK= 0,00	0	1		
	α	β		
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

Verilerin Analizi

Araştırmaya dair tüm işlemler RStudio v1.2.1335 programında R v3.6.1 yazılımı kullanılarak gerçekleştirilmiştir. Madde ve yetenek (θ) parametrelerinin üretimi R yazılımına ait çekirdek paket programlarla gerçekleştirilmiştir. Madde parametrelerinin tahmini için "mirt" paketi kullanılmıştır. Bu paket açıklayıcı 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 Expectation-Maximization (EM) algoritmaları da bulunmaktadır. Bu çalışma için MML kestirim yöntemi ve EM algoritması kullanılmıştır.

Veri üretiminden analizine geçişteki ilk adımda üretilen parametreler doğrultusunda her bir birey için 0 ve 1'lerden oluşan bir cevap matrisi oluşturmaktadır. Daha sonra oluşan cevap matrisleri kullanılarak parametre tahmini gerçekleştirilmiştir. Yapılan kestirim işlemi sonucunda parametre değerleri elde edilmiş ve bir fonksiyona toplanmıştır. Sonuç olarak fonksiyonlardan gerçek parametre

değerleri ile kestirilen parametre değerleri arasındaki farklar Bias ve RMSE istatistikleri yardımıyla değerlendirilmiştir.

BULGULAR ve YORUMLAR

Lord (1968) ve Drasgow (1989)'un önerileri doğrultusunda örneklem büyüklüğünün $N=1,000$ 'e sabitlendiği ve θ parametrelerinin dağılım durumlarının parametre kestirimlerine etkisinin incelendiği çalışma deseninden elde edilen sonuçlar Tablo 4 ve Tablo 5'te incelenebilir.

Tablo 4. Farklı Dağılımlar İçin a Parametresinin Ortalama RMSE ve Bias Değerleri

Örneklem Büyüklüğü (N)	Test Uzunluğu	Çarpıklık Katsayısı (ÇK)	Ortalama RMSE	Ortalama Bias
N=1000	30	ÇK= 2,00	0,243	0,140
	30	ÇK= -2,00	0,254	-0,147
	30	ÇK= 1,00	0,154	0,030
	30	ÇK= -1,00	0,147	0,028
	30	ÇK= 0,00	0,121	0,005

Tablo 5. Farklı dağılımlar için b parametrelerinin ortalama RMSE ve Bias değerleri

Örneklem Büyüklüğü (N)	Test Uzunluğu	Çarpıklık Katsayısı (ÇK)	Ortalama RMSE	Ortalama Bias
N=1000	30	ÇK= 2,00	0,136	-0,066
	30	ÇK= -2,00	0,143	0,071
	30	ÇK= 1,00	0,153	-0,091
	30	ÇK= -1,00	0,151	0,094
	30	ÇK= 0,00	0,093	-0,001

Tablo 4 incelendiğinde en düşük RMSE ve Bias değeri (0,121, 0,005) ÇK= 0,00 yani θ dağılımı standart normal dağılımken en yüksek RMSE ve Bias değeri (0,254, -0,147) 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 çarpıklık katsayısının (ÇK) 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 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 görülebilir.

Tablo 5 incelendiğinde en düşük RMSE ve Bias değerinin (0,093, -0,001) θ parametreleri standart normal dağılıma sahipken en yüksek RMSE değerinin (0,153) θ parametreleri $\text{ÇK}= 1,00$ olan dağılıma ait olduğu ve mutlak değeri en büyük Bias değerinin (0,094) θ parametreleri $\text{ÇK}= -1,00$ olan dağılımdan 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 $\text{ÇK}= 1,00$ ve $\text{ÇK}= -1,00$ iken çok yüksek değerler almaktayken dağılımların ÇK 'si arttıkça bu değerlerde azalma görülmektedir. Bias sonuçları değerlendirilecek olunursa 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.

Örneklem büyüklüklerinin (250, 500, 1000, 2000) ve θ parametrelerinin çeşitli dağılım özelliklerine sahip olduğu durumların (çarpıklık katsayısı (ÇK) = -2,00, -1,00, 0,00, 1,00, 2,00) parametre kestirimlerini nasıl etkilediğine dair elde edilen sonuçlar Tablo 6 ve Tablo 7'de incelenebilir.

Tablo 6. Farklı dağılımlar ve örneklem büyüklükleri için a parametrelerinin ortalama RMSE ve Bias değerleri

Çarpıklık Katsayısı (ÇK)	Örneklem Büyüklüğü (N)	Ortalama RMSE	Ortalama Bias
$\text{ÇK}= 2,00$	N= 250	0,310	-0,126
	N= 500	0,265	-0,138
	N= 1000	0,243	0,140
	N= 2000	0,230	-0,140
$\text{ÇK}= -2,00$	N= 250	0,306	-0,128
	N= 500	0,264	-0,141
	N= 1000	0,254	-0,147
	N= 2000	0,231	-0,143
$\text{ÇK}= 1,00$	N= 250	0,254	0,021
	N= 500	0,189	0,019
	N= 1000	0,154	0,030
	N= 2000	0,117	0,025
$\text{ÇK}= -1,00$	N= 250	0,262	0,024
	N= 500	0,191	0,019
	N= 1000	0,147	0,028
	N= 2000	0,122	0,024
$\text{ÇK}= 0,00$	N= 250	0,254	0,036
	N= 500	0,175	0,006

N= 1000	0,121	0,005
N= 2000	0,088	0,003

Not: Tüm çalışma desenlerinde test uzunluğu 30 maddeden oluşmaktadır.

Tablo 7. Farklı dağılımlar ve örneklem büyüklükleri için b parametrelerinin ortalama RMSE ve Bias değerleri

Çarpıklık Katsayısı (ÇK)	Örneklem Büyüklüğü (N)	Ortalama RMSE	Ortalama Bias
ÇK= 2,00	N= 250	0,221	-0,088
	N= 500	0,167	-0,081
	N= 1000	0,136	-0,066
	N= 2000	0,120	-0,064
ÇK= -2,00	N= 250	0,214	0,085
	N= 500	0,168	0,077
	N= 1000	0,143	0,071
	N= 2000	0,119	0,064
ÇK= 1,00	N= 250	0,200	-0,071
	N= 500	0,174	-0,091
	N= 1000	0,153	-0,091
	N= 2000	0,137	-0,095
ÇK= -1,00	N= 250	0,205	0,077
	N= 500	0,170	0,089
	N= 1000	0,151	0,094
	N= 2000	0,143	0,099
ÇK= 0,00	N= 250	0,195	0,007
	N= 500	0,131	-0,0001
	N= 1000	0,093	-0,001
	N= 2000	0,065	-0,00002

Not: Tüm çalışma desenlerinde test uzunluğu 30 maddeden oluşmaktadır.

Tablo 6 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 görülebilir. Aynı tablo 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) $\zeta K = -2,00$ olan dağılımda N=1000 örneklem büyüklüğünde elde edildiği görülebilir. Ayrıca a parametreleri için ortalama Bias değerlerinin θ parametrelerinin çarpıklık durumuna daha hassas oldukları ve ζK standart normal dağılıma yaklaştıkça elde edilen ortalama Bias değerlerinin küçüldüğü görülebilir.

Tablo 7 incelendiğinde bütün dağılım türlerinde örneklem büyüklüğü arttıkça b parametrelerine ait ortalama RMSE değerlerinin küçüldüğü görülmüştür. Son derece ilginç bir şekilde örneklem büyüklüğünün N=250 olduğu durum dışında $\zeta K = -1,00$, $\zeta K = 1,00$ ile $\zeta K = -2,00$, $\zeta K = 2,00$ karşılaştırıldığında beklenenin aksine örneklem büyüdükçe $\zeta K = -1,00$ ile $\zeta K = 1,00$, $\zeta K = -2,00$ ile $\zeta K = 2,00$ '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 $\zeta K = 2,00$ 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 ζK '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 ζK 'nin hem pozitif hem de negatif 1 durumu ile standart normal dağılım durumu karşılaştırıldığında çok daha net şekilde gözlemlenebilir. Aynı tablo incelendiğinde örneklem büyüklüğündeki artışın $\zeta K = 2,00$, $\zeta K = -2,00$ ve standart normal dağılıma sahip durumlarda b parametrelerinin ortalama Bias değerlerini genelde 0'a yaklaştırdığı görülmüştür. $\zeta K = 1,00$ ve $\zeta K = -1,00$ olan dağılımlarda ise örneklem büyüklüğünde yaşanan artış ortalama Bias değerlerini 0'dan uzaklaştırmıştır. Mutlak değerleri eşit olan ζK '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 $\zeta K = -1,00$ olan dağılımda ve N=2000 olan örneklem büyüklüğünden (0,099) elde edildiği görülebilir. Ayrıca en kötü sonuçların ζK 'nin negatif değerlerinde elde edildiği de gözlemlenebilir.

SONUÇLAR ve TARTIŞMA

Örneklem büyüklüğünün sabit tutulup yalnızca ζK 'nin değişim gösterdiği durumlarda a parametrelerinin ortalama RMSE değerlerinin olması gerekenden büyük olduğu görülmüştür (Browne ve Cudeck, 1993). Buna rağmen θ dağılımlarının ζK 'si standart normal dağılıma yaklaştıkça RMSE

değerlerinde de küçülme görülmüştür. Ortalama RMSE değerlerinin bu kadar yüksek olmasının test uzunluğunun veya örneklem büyüklüğünün yetersiz olmasından kaynaklandığı düşünülmektedir.

Örneklem büyüklüğünün sabit tutulduğu ve yalnızca ÇK 'nin değişim gösterdiği durumlarda a parametrelerinin ortalama Bias değerlerinin ÇK standart normal dağılıma yaklaştıkça sifıra yaklaştığı görülmüştür. Ayrıca ortalama Bias değerlerinin ÇK 'nin hem pozitif hem de negatif 2,00 değeri dışında kabul edilebilir düzeylerde olduğu tespit edilmiştir. Elde edilen sonuçlar beklendiği gibidir.

Örneklem büyüklüğünün sabit olduğu ve yalnızca ÇK 'nin değişim gösterdiği durumlarda b parametrelerinin ortalama RMSE değerlerinin $\text{ÇK}=-1,00$ ve $\text{ÇK}=1,00$ olduğu durumlarda ilginç bir şekilde en büyük değerleri ürettiği 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. Yine bu durumun yukarıda açıklanan sebepten kaynaklandığı düşünülmektedir.

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.

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. 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 büyüklüğünde 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 sifıra yaklaşmaktadır. Ancak dağılımlar kendi içlerinde değerlendirildiğinde örneklem büyüklüğü arttıkça ortalama Bias değerlerinin de sifıra 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. Tüm bu bilgiler değerlendirildiğinde $\text{ÇK}=-2,00$ ve $\text{ÇK}=2,00$ 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 $\text{ÇK}=-2,00$ ve $\text{ÇK}=2,00$ olan dağılımların $\text{ÇK}=-1,00$ ve $\text{ÇK}=1,00$ 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. . 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 $\text{ÇK}=1,00$ ve $\text{ÇK}=-1,00$ olan dağılımlar hariç örneklem büyüklüğü arttıkça sıfıra yaklaşma eğiliminde olmuşlardır. $\text{ÇK}=-1,00$ ve $\text{ÇK}=1,00$ olan dağılımlarda ise örneklem büyüklüğü arttıkça ortalama Bias değerleri sıfırdan uzaklaşma eğilimindedirler. $\text{ÇK}=-1,00$ ve $\text{ÇK}=1,00$ dışında sonuçlar beklenildiği gibidir. $\text{ÇK}=-1,00$ ve $\text{ÇK}=1,00$ 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 $\text{ÇK}=-1,00$ için ortalama Bias değerlerinin maksimum farkının 0,022, $\text{ÇK}=1,00$ için ortalama Bias değerlerinin maksimum farkının 0,024 olduğu görülmektedir. Bu farklar doğrultusunda aslında $\text{ÇK}=-1,00$ ve $\text{ÇK}=1,00$ 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. 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.

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 parametrelili normal ogive model altında 45 maddelik bir test ve $N=1,000$

örneklem büyüklüğünü kullanan Seong (1990)'un çalışmasıyla uyuşmamaktadır. Stone (1992)'un çalışması incelendiğinde 2 PL model altında test uzunluğunu, örneklem büyüklüğünü ve θ parametrelerinin dağılımını manipüle eden araştırmacının elde ettiği bulgular bu çalışmayı desteklemektedir. Kieftenbeld ve Natesan (2012)'in 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. Boulet (1996)'in çalışması incelendiğinde 2 PL 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 2 PL model altında b parametrelerinin ve θ parametrelerinin dağılımını manipüle eden araştırmacıların RMSE istatistiğinden elde ettikleri sonuçlar bu 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 parametrelilik normal ogive model altında 45 maddelik bir test ve N=1,000 örneklem büyüklüğünü kullanan Seong (1990)'un çalışmasıyla uyuşmamaktadır. Stone (1992)'un çalışması incelendiğinde 2 PL 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. Abdel-fattah (1994)'in 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=1,000 olduğu veri setinden elde ettiği sonuçlar bu çalışmayla tamamen çelişmektedir.

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 parametrelilik normal ogive model altında 45 maddelik bir test ve N=1,000 örneklem büyüklüğünü kullanan Seong (1990)'un çalışmasıyla uyuşmamaktadır. 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 ışığında şu anki çalışmadan elde edilen sonuçları desteklediği söylenemez. 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 çalışmada da b parametrelerinin kestirimi için θ parametreleri standart normal dağılıma sahipken en küçük RMSE değerlerini verdiğinden Sass vd. (2008)'in çalışmasının bu çalışmayı desteklediği söylenebilir. Abdel-fattah (1994)'in 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=1,000 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 (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 durumlardan elde edilen sonuçlar değerlendirildiğinde Bias değerlerinin 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 çalışmayı 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 dağılım θ parametrelerinden elde edilen ortalama Bias değerleri diğer dağılım türlerine göre sıfıra daha yakındır. Bu nedenle en yakın ortalama Bias değerlerinin standart normal dağılımdan elde edilmesi şu anki çalışmayı desteklemektedir. Abdel-fattah (1994)'in çalışması incelendiğinde MML kestirim yönteminde test uzunluğunun 20 madde ve örneklem büyüklüğünün N=1,000 olduğu veri setinde standart normal dağılımın diğer dağılımlardan 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

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 $\text{ÇK}=0,00$ '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. Stone (1992)'un çalışması incelendiğinde tüm test uzunlukları (10, 20, 40 madde) ve θ parametrelerinin dağılımında (normal, çarpık ve basık) örneklem büyüklüğündeki artışın ($N=250, 500, 1,000$) 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)'in çalışmaları incelendiğinde MML kestirim yöntemi için tüm dağılım türlerinde (normal, çarpık ve uniform) örneklem büyüklüğü arttıkça ($N=75, 150, 300, 500, 1,000$) 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, kestirimin uygulandığı yazılım türlerinde ve θ parametrelerinin dağılım türlerinde örneklem büyüklüğü arttıkça 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)'in çalışması incelendiğinde elde edilen sonuçların MML kestirim yöntemi için tüm test uzunluklarında ve θ parametrelerinin dağılım türlerinde örneklem büyüklüğü arttıkça elde edilen RMSE değerlerinde küçülme gözlemlenmiştir. Bu nedenle araştırmacının çalışmasında ulaştığı sonuçların bu çalışmadan elde edilen sonuçları desteklediği söylenebilir. Akour ve AL-Omari (2013)'nin çalışmaları incelendiğinde 3 PL model altında test uzunluklarını ve örneklem büyüklüklerini 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.

Madde ayırt edicilik (a) parametreleri için hem örneklem büyüklüğü hem de çarpıklık katsayıları manipüle edildiğinde ortalama Bias değerleri $\text{ÇK}=-2,00$ ve $\text{ÇK}=2,00$ için örneklem büyüklüğü arttıkça sıfırdan uzaklaşma, $\text{ÇK}=0,00$ için sıfıra yaklaşma eğilimindedir. $\text{ÇK}=-1,00$ ve $\text{ÇK}=1,00$ için örneklem büyüklüğündeki artış ortalama Bias değerlerinde belirli şekilde sıfıra yaklaşma 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 $\text{ÇK}=0,00$ için sonuçlar desteklenirken diğer ÇK 'ler için sonuçlar desteklenmemektedir. Stone (1992)'un çalışmasından elde edilen bulgular değerlendirildiğinde şu anki çalışmayla benzer sonuçlara ulaşılabilmektedir. Bu nedenle Stone (1992)'un çalışması bu araştırmayı desteklemektedir. Abdel-fattah (1994)'ın çalışmasından elde edilen bulgular incelendiğinde 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 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 $\text{ÇK}=-2,00$ ve $\text{ÇK}=2,00$ olan dağılımlar $\text{ÇK}=-1,00$ ve $\text{ÇK}=1,00$ 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 desteklemediği 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 maddeden oluştuğu ve θ parametrelerinin truncated 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)'ın çalışması incelendiğinde 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. 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)'in çalışmaları incelendiğinde 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 sifıra yakınlaşma eğilimindedir. ÇK=-1,00 ve ÇK=1,00 olan dağılımlarda ise örneklem büyüklüğü arttıkça ortalama Bias değerleri sifirdan uzaklaşmaktadır. Seong (1990)'un çalışması incelendiğinde prior ve posterior dağılımların eşleştiği durumlarda θ parametrelerinin tüm dağılım türlerinde örneklem büyüklüğü arttıkça ortalama Bias değerlerinin de sifıra yaklaştığı görülmektedir. Bu bulgular değerlendirildiğinde elde edilen sonuçların şu anki araştırmayı desteklemediği söylenebilir. Stone (1992)'un çalışması incelendiğinde test uzunluğunun 10 ve 40 maddeden oluşan veri setlerinden elde edilen sonuçların şu anki araştırmının sonuçlarını desteklediği söylenebilir. Abdel-fattah (1994)'ın çalışması incelendiğinde elde edilen sonuçların şu anki çalışmayı desteklemediği söylenebilir.

Yukarıdaki tüm çalışmalar ve RMSE ile Bias sonuçları değerlendirildiğinde yalnızca θ parametreleri standart normal dağılıma sahipken ve N=2,000 örneklem büyüklüğünde madde parametre kestirim işlemi gerçekleştirmek uygun görünmektedir.

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.
- 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 Ottawa/Faculty of Education, Ottawa.
- Browne, M.W. & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. Bollen, K.A. & Long, J.S.. *Testing Structural Equation Models*. Newbury Park, CA:Sage kitabından
- 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.
- Chalmers, R. P. (2012). mirt:A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6), 1-29.
- Crocker, L. & Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. Mason, Ohio: Cengage Learning.
- 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 Psychological Measurement*, 27(4), 275-288.
- Dragow, F. (1989, Mart). An Evaluation of Marginal Maximum Likelihood Estimation for the Two-Parameter Logistic Model. *Applied Psychological Measurement*, 13(1), 77-90.

- Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of Educational Measurement* (5. baskı). New Delhi: Prentice-Hall.
- 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.
- Feinberg, R. A. & Rubright, J. D. (2016, Haziran). Conducting Simulation Studies in Psychometrics. *Educational Measurement: Issues and Practice*, 35(2), 36-49.
- Finch, W. H. & French, B. F. (2019). *Educational and Psychological Measurement*. New York, NY: Routledge.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Meyers, L. S. & Grossen, N. E. (1974). *Behavioral Science: Theory, Procedure, and Design*. USA: W. H. Freeman and Company.
- 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.
- Reise, S. P. & Yu, J. (1990). Parameter Recovery in the Graded Response Model Using MULTILOG. *Journal of Educational Measurement*, 27(2), 133-144.
- 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.
- 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.

- 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.
- Sünbül, Ö. & Bulut, O. (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.
- 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)'nin 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.
- 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.
- Yen, W. M. (1987, Haziran). A Comparison of the Efficiency and Accuracy of BILOG and LOGIST. *Psychometrika*, 52(2), 275-291.
- Yıldırım, Y. (2015). *Derecelendirilmiş Tepki Modeli Temelli Parametre Kestiriminde Normalliğin İhlalinin Ölçme Keskinliğine Etkisi*. Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- 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.