

## Madde Tepki Kuramına ait Parametrelerin ve Model Uyumlarının Karşılaştırılması: Bir Monte Carlo Çalışması\*

### Comparison of Item Parameters and Model Fit from Item Response Theory Applications: A Monte Carlo Study

Hakan KOĞAR \*\*

#### Öz

Bu araştırmanın amacı, basit ve iki boyutlu yapılarda, çeşitli örneklem büyüklükleri, test uzunlukları ve boyutlar arası korelasyon değerlerinde, madde tepki kuramına ait farklı uygulamalardan elde edilen madde parametreleri, maddelere ve teste ait model veri uyumlarını belirlemek ve sonuçları karşılaştırmaktır. Örneklem büyüklüğü (100, 500, 1000 ve 5000), test uzunluğu (5, 15 ve 25) ve boyutlar arası korelasyon değerlerindeki (0.00, 0.25 ve 0.50) değişim ile elde edilen her bir deneysel desenden, Monte Carlo çalışması kapsamında, 20 tekrar ile veri setleri üretilmiştir. Bu tekrarlar, tek ve çok değişkenli normal dağılım altında üretilmiştir. Madde ve model veri uyumu parametreleri için standart hata ve anlamlılık değerleri hesaplanmıştır. Bulgular incelendiğinde, örneklem büyüklüğü ve test uzunluğundaki artış ile birlikte, teste ait model veri uyumu değerlerinin de arttığı belirlenmiştir. Daha az maddeden oluşan testlerin çok boyutlu madde tepki kuramına daha iyi uyum sağladığı söylenebilmektedir. Tüm simülasyon düzeneklerinde, parametrik olmayan madde tepki kuramında, maddelere ait model veri uyumu oldukça düşük bir hata ile hesaplanmaktadır. Parametrik madde tepki kuramına ait ki-kare, infit ve outfit değerleri incelendiğinde, her üç katsayının da örneklem büyüklüğü ve test uzunluğundaki artış ile birlikte, daha iyi model veri uyumunu gösterdiği belirlenmiştir. Parametrik olmayan madde tepki kuramında  $H_i$  ve  $p$  parametrelerine ait standart hata değerleri, örneklem büyüklüğündeki artış ile azalma eğilimi göstermektedir. Parametrik madde tepki kuramında  $a$  parametresine ait standart hata değerlerinin, örneklem büyüklüğündeki ve test uzunluğundaki artış ile birlikte, sıfıra yaklaştığı sonucuna varılmıştır. Çok boyutlu madde tepki kuramında  $a_1$  ve  $a_2$  parametrelerine ait standart hata değerleri, parametrik madde tepki kuramına benzer sonuçlar vermektedir.

*Anahtar Kelimeler:* madde tepki kuramı, madde parametresi, model veri uyumu, monte carlo

#### Abstract

The purpose of this study is to identify and compare NIRT, PIRT and MIRT across different sample sizes, test length and correlation between dimensions in a two dimensional simple structures. Data sets in various conditions have been simulated. These conditions are sample size (100, 500, 1000 and 5000), test length (5, 15 and 25) and correlation between dimensions (0.00, 0.25 and 0.50). From each experimental design, within the frame of Monte Carlo study, the findings have been obtained through 20 replications. For the item parameters and model data fit for the items, standard errors and significance values have been calculated. Having analyzed the findings of the research, with the increase of sample sizes and test length, it is also found out that the model data fit for the test has increased as well. It can be stated that tests consisting of less items fit better to MIRT models. In all simulation designs, model data fit for the items are calculated with quite low errors in NIRT. When the chi-square, infit and outfit values obtained for PIRT have been analyzed, it has been revealed that along with the increase of sample sizes and test length, all three coefficients exhibit better model fit. In NIRT, the standard errors belonging to  $H_i$  and  $p$  parameters tend to decrease with the increase of sample sizes and test length. In PIRT, a parameters tend to decrease when the sample sizes and test length increase.

*Key Words:* item response theory, item parameter, model data fit, monte carlo

\* Bu çalışma, Madde tepki kuramının farklı uygulamalarından elde edilen parametrelerin ve model uyumlarının örneklem büyüklüğü ve test uzunluğu açısından karşılaştırılması isimli doktora tezinden üretilmiştir.

\*\* Öğr. Gör. Dr., Başkent Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: [hskogar@gmail.com](mailto:hkogar@gmail.com)

## GİRİŞ

Madde tepki kuramı (MTK), katılımcıların test performansı ile performansın altında yatan gizil özellikler arasında bir ilişki tanımlamaya çalışan bir modelleme tekniğidir (Hambleton ve Swaminathan, 1985; Henard, 2000). MTK’da yetenek ölççeği ile bir maddeye doğru cevap verme olasılığı arasındaki ilişki, bir eğri yardımıyla tanımlanmaktadır. S şeklindeki bu eğri, madde karakteristik eğrisidir (MKE). MKE’ye ait standart matematiksel model, bir tür lojistik fonksiyondur. Bu fonksiyona göre  $\theta$ ’ya (yetenek parametresi) ait olasılık monoton olarak artmaktadır. Lojistik fonksiyon değerleri 0 ve 1 aralığında bir değer alabilmekte ve bu nedenle olasılıklarla değerlendirilmektedir. Bu fonksiyon ile ilgili en önemli problem, yalnızca  $\theta$ ’ya bağlı olması ve maddelere ait cevapların olasılıkları ile ilgilenmemesidir. Ancak bazı maddelerin kolay ve bazı maddelerin zor olduğu, madde güçlüğünün de maddeye doğru cevap verme olasılığını etkilediği bilinmektedir. Bu ve benzeri problemlerin önüne geçebilmek amacıyla çeşitli MTK modelleri geliştirilmiştir (Reckase, 2009).

Tek boyutlu parametrik madde tepki kuramı (TBMTK), modern modeller olarak bilinmekte ve olasılığa dayanmaktadır. Buna göre, bir maddenin kolay olması ya da bir bireyin yeteneğinin yüksek olması, doğru cevap verme olasılığının yüksek (ancak 1 değil) olacağı anlamına gelecektir. Maddenin zor olması veya bireyin düşük yetenekte olması ise bu olasılığın düşük olmasına (ancak 0 değil) yol açacaktır. Olasılığın doğru bir şekilde belirlenmesi, kişinin yeteneğine ya da özelliğinin seviyesine bağlıdır (Sijtsma ve Molenaar, 2002, 3). TBMTK modellerinden biri olan iki parametrelili lojistik model, yalnızca güçlük parametresinin kestirildiği Rasch modeline ayırt edicilik parametresinin eklenmesi ile elde edilmiştir (Linden ve Hambleton, 1997). Bu modelde şans parametresinin (c) sıfır olduğu varsayılmaktadır (Lee, 2007; Seungho Yang, 2007). Harris’e (1989) göre a parametresi uygulamada genellikle -3.0 ile +3.0 aralığında bir değer alabilmektedir.

Tek boyutluğun sağlanamadığı durumlarda yalnızca, çok boyutlu madde tepki kuramı (ÇBMTK) modelleri kullanılabilir. Bir testin içindeki maddelerden her biri bir boyut ölçüyor ve hiçbir madde birden fazla yetenek ölçmüyorsa bu türden test verileri basit yapıdadır (Ackerman, Gierl ve Walker, 2003). ÇBMTK, ikili puanlanan maddeler için tamamlayıcı (compensatory) ve tamamlayıcı olmayan (noncompensatory) modeller olmak üzere ikiye ayrılmaktadır. Tamamlayıcı model, çok boyutlu yapılarda bir yetenek düzeyindeki yetkinliğin diğer boyut ya da boyutlardaki eksik olan yetkinliği tamamlaması durumunda kullanılabilecek olan bir modeldir (Ackerman, 1996). Örneğin, kimya dersine ait bir radyoaktivite sorusuna cevap verebilmek için, öğrencinin hem radyoaktivite konu bilgisine hem de matematiksel işlem bilgisine sahip olması gerekir. Ancak, radyoaktivite konusunda yetkin bilgiye sahip bir öğrenci bu maddeyi doğru cevaplayabilir. Tamamlayıcı modelde yer alan çok boyutluluk için genişletilmiş iki parametrelili lojistik model, TBMTK’da yer alan iki parametrelili lojistik modelden türetilmiştir. Tek fark, yetenek parametresinin tek bir yapının düzeyini yorumlaması değil, koordinatların toplamından elde edilen bir değer olmasıdır (Reckase, 2009, 92).

Parametrik olmayan madde tepki kuramı (TBPoMTK), kişilerin ve maddelerin ölçülmesinde kullanılan ve çok az varsayım gerektiren kullanışlı bir istatistiksel ölçme modelidir. TBPoMTK, gizil değişken ile cevap kategorisine verilen cevaba ait olasılık arasındaki fonksiyonu parametrik olarak tanımlamamaktadır. TBPoMTK modelleri, sıralama düzeyindeki ölçmelere izin verir ve klasik testler ile anketlerdeki her bir katılımcıyı değerlendirme için oldukça uygundur (Stochl, 2007). TBPoMTK modeli, maddelerin sıralanmasına izin vermektedir. Parametrik olmayan bir istatistiksel ölçme modeli olması, çok çeşitli veri setlerinin TBPoMTK’ya uygun olduğunu göstermektedir. TBPoMTK’nın sahip olduğu algoritmalar, TBMTK ve ÇBMTK’ya göre daha basittir (Sijtsma ve Molenaar,

2002). Mokken (1971), ikili puanlanan maddelerin analizi için iki farklı yöntem önermiştir. Bunlar, Monoton Homojenlik Modeli (Monotone Homogeneity Model) ve İkili Monotonluk Modeli'dir (Double Monotonicity Model). Monoton Homojenlik Modeli'nde (MHM) Rasch modelinin bir çeşidi olarak her bir maddede ayırıcılık parametresi olarak  $\alpha_i$  yerine  $\alpha$  kullanılır. Bu modele göre madde parametreleri olan  $a_i$  ve  $\delta_i$ , TBMTK'daki iki parametrelilik lojistik modele karşılık gelmektedir. Diğer bir deyişle bu model, iki parametrelilik lojistik modelin parametrik olmayan karşılığıdır (Molenaar ve Sijtsma, 2000; akt. Chia-Lin vd., 2006).

TBPoMTK modelinde test maddelerinin her biri için ( $H_i$ ) ve tüm test için ( $H$ ) hesaplanabilen ölçeklenebilirlik (scalability) katsayıları geliştirilmiştir. Bu katsayılardan  $H_i$  katsayısı, hem maddeye ait geçerliliği hem de ayırt ediciliği veren bir katsayıdır.  $H_i$  katsayısının 0.00 ve 0.30 aralığında olduğu durumda, bu maddeler düşük ayırt ediciliğe sahiptir. Ayrıca, güvenilir kişi sıralamasına çok az katkı getirmekte olduğu için kullanışlı bir madde değildir.  $H$  katsayısı ise,  $H_i$  katsayılarından elde edilen ve testin geneline ait uyumu ifade eden bir değerdir. Elde edilen  $H_i$  katsayılarının ağırlıklandırılmış ortalamaları alınarak hangi katılımcıların cevapladıkları veri setlerinden yola çıkarak doğru sıralandıkları belirlenir.  $H$  katsayısı, en düşük  $H_i$  katsayısından daha büyük bir değere sahiptir.  $H$  katsayısı aynı zamanda testin güvenilirliği hakkında da bilgi vermektedir. Bu katsayıda ölçülen hata türü, Guttman hatasıdır.  $H$  katsayısının 1'e eşit olması, Guttman hatasının hiç yapılmadığı anlamına gelmektedir. Bu katsayının yorumlanması  $H_i$  katsayısında olduğu gibidir (Sijtsma ve Molenaar, 2002).

Tüm MTK modellerinin uygulanmasında ve yorumlanmasında örneklem büyüklüğü önemli bir etkiye sahiptir. Alanyazında, örneklem büyüklüğü arttıkça parametre tahminlerinin daha doğru olacağına dair ortak bir görüş olarak yer almaktadır. Test uzunluğundaki farklılaşma ise, çeşitli araştırmalarda, model veri uyumu ve parametreler üzerinde etkiye neden olduğunu göstermektedir (Baker, 1991; Stone, 1992; Bolt ve Lall, 2003; Sünbül, 2011). Boyutlar arasındaki korelasyon arttıkça, her bir boyut için elde edilen  $\theta$  değerleri arasındaki ilişki de artmaktadır. Batley ve Boss'un (1993) araştırmasında, boyutlar arasındaki ilişkinin sıfır olduğu durumda, birinci boyuttaki  $\theta$  değerleri en iyi şekilde, ikinci boyuttaki  $\theta$  değerleri ise en kötü şekilde kestirilmiştir.

Bu araştırmanın amacı, basit ve iki boyutlu yapılarda, çeşitli örneklem büyüklükleri, test uzunlukları ve boyutlar arası korelasyon değerlerinde, parametrik olmayan, parametrik ve çok boyutlu madde tepki kuramından elde edilen madde parametreleri, maddelere ait model veri uyumları ve teste ait model veri uyumlarını belirlemek ve elde edilen sonuçları karşılaştırmaktır.

Bu amaca uygun olacak şekilde aşağıdaki alt problemlere cevap aranmıştır:

Farklı test uzunlukları ( $k = 5, 15, 25$ ), farklı örneklem büyüklükleri ( $n = 100, 500, 1000$  ve  $5000$ ) ve farklı boyutlar arası korelasyonlarda ( $0.00, 0.25$  ve  $0.50$ ), TBPoMTK kuramına ait Monoton Homojenlik Modeli, TBMTK kuramına ait 2 Parametrelilik Lojistik Model ve ÇBMTK kuramına ait Tamamlayıcı Modelden elde edilen;

1. teste ait model veri uyumu değerleri nasıldır?
2. maddelere ait model veri uyumu değerleri nasıldır?
3. maddelere ait model veri uyumu için standart hata ve anlamlılık düzeyi değerleri nasıldır?
4. madde parametreleri nasıldır?
5. madde parametrelerine ait standart hata değerleri nasıldır?

Eğitim ve psikoloji alanında çalışılan konular, genellikle tek bir örtük özelliği değil, birden fazla örtük özelliğin ölçüldüğü yapılardan oluşmaktadır. En doğru, kararlı ve kullanışlı sonuçlar elde edebilmek amacıyla, çok boyutlu yapıların analizi üzerinde

çalışılması gerekmektedir. Eğitim ve psikoloji alanında çalışılan konuların bir diğer özelliği de, anket maddeleri gibi birbirinden bağımsız maddelerin analizi ve kişi ve maddeleri sıralamaya dayalı ölçekleme çalışmalarıdır. Örtük özellikler daima belirli bir örüntü dahilinde değil, birbirlerinden bağımsız yapılar olarak da ölçülebilir. Ayrıca, yapılan ölçmenin amacı, yalnızca kişileri ya da maddeleri sıralamak olabilmektedir. Bazı durumlarda, oldukça küçük örneklem büyüklükleri ile çalışmak gerekebilmektedir. Bazı durumlarda ise, kısa testler tercih edilmelidir. Bu araştırma, tanımlanan tüm bu durumlarda kullanılabilecek farklı ölçme modellerini önermeyi amaçladığı için alanyazına katkı getireceği düşünülmektedir. Bu araştırma, özellikle TBPoMTK ve ÇBMTK'nın tanıtılması açısından önemli görülmektedir. Yapılan alanyazın taramasında TBPoMTK ile ilgili Türkiye'de herhangi bir çalışmaya rastlanmamıştır. Araştırma alanyazında sınırlı bulunan simülasyon çalışmalarına Monte Carlo çalışması ile katkı sağlaması; parametrik ve parametrik olmayan modeller ile tek boyutlu ve çok boyutlu modellerin birlikte kullanılmasını sağlaması açısından önemli görülmektedir.

## YÖNTEM

Bu araştırma, iki boyutlu bir veri setinde, TBMTK, TBPoMTK ve ÇBMTK modelleri ile madde parametrelerini ve model uyumlarını belirlemeyi ve karşılaştırmayı amaçladığı için temel bir araştırma niteliğindedir. Bu çalışma, kuramları belirli koşullar altında karşılaştırma ve test etmeye dönüktür.

### Veri Üretim Çalışması

Araştırmada kullanılacak veri setleri, MIRTGEN 2.0 (Luecht, 2004) programı ile üretilmiştir. Bu program 3 parametrelili tamamlayıcı ÇBMTK modeli altında, belirtilen madde sayısına göre ikili (1-0) cevap örüntüleri üretmektedir. MIRTGEN 2.0, 50 boyuta kadar veri üretimine izin vermektedir. N sayıdaki her boyut için ortalamalar ve standart sapmaları içeren bir vektör, her bir gizil özelliğin birleşmesi ile oluşturulan korelasyon matrisinin köşegenlerindeki en düşük değerler ve çok değişkenli normal dağılım altında oluşturulmaktadır. Bu sayede elde edilen veri seti ile her bir katılımcı için üretilen yeteneklere ait vektör oluşturulur (Goodman, 2008).

3 farklı test uzunluğu (5, 15, 25) \* 4 farklı örneklem büyüklüğü (100, 500, 1000 ve 5000) \* 3 farklı boyutlar arası korelasyon (0.00, 0.25 ve 0.50) olmak üzere 36 farklı deneysel desen elde edilmiştir. Deneysel desen Tablo 1'de özetlenmiştir.

Tablo 1. Araştırma için Tasarlanan Deneysel Desen

Örneklem Büyüküğü	Boyutlar Arası Korelasyon	Birinci	İkinci	İkinci	İkinci
		Boyuttaki Madde Sayısı (25)	Boyuttaki Madde Sayısı (5)	Boyuttaki Madde Sayısı (15)	Boyuttaki Madde Sayısı (25)
100	0,00	-	X	X	X
	0,25	-	X	X	X
	0,50	-	X	X	X
500	0,00	-	X	X	X
	0,25	-	X	X	X
	0,50	-	X	X	X
1000	0,00	-	X	X	X
	0,25	-	X	X	X
	0,50	-	X	X	X
5000	0,00	-	X	X	X
	0,25	-	X	X	X
	0,50	-	X	X	X

Tablo 1’de görüldüğü gibi, birinci boyutta yer alan madde sayısı sabit tutularak, yalnızca ikinci boyuttaki madde sayıları değiştirilerek veri üretimi gerçekleştirilmiştir. İkinci boyut için yer alan veri setlerine ait tek değişkenli ve çok değişkenli normallik değerleri incelenerek, analize hazır hale getirilmiştir. Her bir deneysel desenden, Monte Carlo çalışması kapsamında, 20 tekrar (replications) ile bulgular elde edilmiştir.

### **Verilerin Analizi**

Araştırma bir Monte Carlo çalışması (deneyi) olarak gerçekleştirilmiştir. MTK’da tipik bir Monte Carlo çalışmasının aşamaları maddeler halinde özetlenmiştir (Harwell vd.,1996):

1. Çalışmanın amacına uygun olarak bir ya da daha fazla araştırma problemi belirlenmelidir.
2. Çalışmanın koşulları belirtilmelidir (katılımcı ve madde sayıları, bağımlı ve bağımsız değişkenlerin ne olduğu vb.). Bu araştırmadaki bağımlı değişkenler, madde parametreleri ve maddeler için model veri uyumuna ait standart hata değerleri ile teste ait model veri uyumunu gösteren değerler iken; bağımsız değişkenler boyutlar arası ilişki, test uzunluğu ve örneklem büyüklüğüdür.
3. Uygun bir deneysel desen belirlenmelidir. Bu çalışmada, 3 farklı test uzunluğu, 4 farklı örneklem büyüklüğü ve 3 farklı boyutlar arası korelasyon ile 36 farklı deneysel desen tasarlanmıştır.
4. Öngörülen modellere uygun madde cevap verisi üretilmelidir. Araştırmada kullanılan veri seti, MIRTGEN 2.0 programı ile üretilmiştir. Bu veri iki boyutlu ve ikili (1-0) puanlanan yapıya sahiptir.
5. Üretilmiş madde cevapları ile parametreler belirlenmelidir.
6. Çıktılar tarafından modellenen koşulların etkileri ölçülerek karşılaştırılmalıdır. Bu araştırmada bu amaçla madde parametreleri için standart hata değerleri; model veri uyumları için ise Akaike Bilgi Ölçütü (ABÖ) ve Bayesian Bilgi Ölçütü (BBÖ) değerleri ile diğer model uyumları karşılaştırılmıştır.
7. Bu süreç R defa, desendeki her bir hücre için tekrar edilmelidir. Böylece her bir hücre için R adet çıktı elde edilecektir. Her bir deneysel desenden, 20 tekrar ile bulgular elde edilmiştir.
8. R sayıdaki çıktı betimsel ve çıkarımsal istatistikler ile analiz edilmelidir. Çıkarımsal istatistikler, araştırma problemleri ve deneysel desene göre belirlenmelidir. Bu bulgular araştırma problemlerini destekleyen kanıtlar olacaktır.

Madde parametrelerinin belirlenmesi amacıyla, TBPoMTK kuramına ait Monoton Homojenlik Modeli’nde  $a_i$  (Ölçeklenebilirlik – H katsayısı) ve  $\delta_i$  (p değerleri); TBMTK kuramına ait 2PLM’de a ve b; ÇBMTK kuramına ait Tamamlayıcı Model’de ise  $\alpha_1$ ,  $\alpha_2$  ve d parametreleri tahmin edilmiştir. Ayrıca, bu parametreler için standart hata değerleri de elde edilmiştir. Maddelere ait model uyumlarının belirlenmesi amacıyla, TBPoMTK kuramına ait MHM’de madde için scalability (ölçeklenebilirlik) katsayısı; TBMTK kuramına ait 2PLM’de infit ve outfit değerleri ile ki-kare uyum istatistiği; ÇBMTK kuramına ait Tamamlayıcı Model’de ise infit ve outfit değerleri elde edilmiştir. Ayrıca elde edilen model veri uyumu parametreleri için standart hata ve anlamlılık değerleri hesaplanmıştır. Teste ait model uyumlarının belirlenmesi amacıyla, TBPoMTK kuramına ait MHM’de test için ölçeklenebilirlik (scalability) katsayısı; TBMTK kuramına ait 2PLM ve ÇBMTK kuramına ait Tamamlayıcı Model’de log-benzerlik istatistiği ve bazı bilgi ölçütleri (akaike ve bayesian) elde edilmiştir.



Parametreler ve model veri uyumunun değerlendirilmesinde, R 3.0.2 yazılımı kullanılmıştır. TBPoMTK parametreleri ve model veri uyumu için mokken ve KernSmoothIRT paketleri; TBMTK için eRm ve Itm paketleri; ÇBMTK için ise mirt paketi kullanılmıştır.

## BULGULAR

Birinci alt probleme ait bulgular Tablo 2, 3 ve 4'te yer almaktadır.

Tablo 2. Boyutlar Arası Korelasyonun 0,00 Olduğu Durumda Teste ait Model Veri Uyumunun Değerlendirilmesi

Örneklem Büyüklüğü	Test Uzunluğu	Kuramlar						
		TBPoMT K	TBMTK	ÇBMTK				
		H	-2*Log- Benzerlik	ABÖ*	BBÖ**	-2*Log- Benzerlik	ABÖ	BBÖ
100	5	0,00	-337,1	694,2	720,2	-336,0	700,0	736,4
	15	0,00	-1020,6	2101,3	2179,4	-1010,9	2109,8	2224,5
	25	0,00	-1699,5	3498,9	3629,2	-1681,3	3510,7	3703,5
500	5	0,00	-1726,5	3473,0	3515,1	-1725,0	3478,1	3537,1
	15	0,00	-5179,6	10419,1	10545,6	-5170,0	10428,1	10613,5
	25	0,00	-8630,7	17361,4	17572,2	-8613,3	17374,5	17686,4
1000	5	0,00	-3459,3	6938,6	6987,7	-3458,1	6944,2	7012,9
	15	0,00	-10377,9	20815,8	20963,0	-10368,3	20824,7	21040,6
	25	0,00	-17293,5	34687,1	34932,5	-17275,1	34698,2	35061,4
5000	5	0,00	-17323,0	34666,0	34731,2	-17321,9	34671,9	34763,1
	15	0,00	-51965,6	103991,1	104186,6	-51956,2	104000,5	104287,2
	25	0,00	-86609,0	173318,1	173643,9	-86592,7	173333,5	173815,7

\* ABÖ: Akaike Bilgi Ölçütü, \*\* BBÖ: Bayesian Bilgi Ölçütü

TBPoMTK için teste ait model veri uyumunu değerlendirmek için kullanılan H katsayısı, her örneklem büyüklüğü ve test uzunluğunda, 0,00 değerini almıştır. Bu durum, model veri uyumunun hiçbir koşulda sağlanmadığını gösterebilmekle birlikte, Mokken'in (1971, 120) yaptığı açıklamadan yola çıkılarak, bu araştırmada en az bir maddenin bir ya da daha fazla monoton olarak artan madde yüzey cevap fonksiyonuna sahip olmasından dolayı H = 0 durumu gerçekleşmiş olabileceği söylenebilmektedir. Diğer bir olasılık ise TBPoMTK'nın maddelerin sıralanması ile ölçeklemeye izin vermesidir. Bu araştırma için üretilen veri setlerinde yer alan maddeler herhangi bir örüntü / sıralama olmadan, tamamen bağımsız bir şekilde üretilmiştir. Bu durum, ölçeklemenin gerçekleşmediğine de işaret edebilmektedir. Bu nedenle, H katsayısı için yorum yapılamamıştır.

Tablo 3. Boyutlar Arası Korelasyonun 0,25 Olduğu Durumda Teste ait Model Veri Uyumunun Değerlendirilmesi

Örneklem Büyüklüğü	Test Uzunluğu	Kuramlar						
		TBPoMTK	TBMTK	ÇBMTK				
		H	-2*Log-Benzerlik	ABÖ*	BBÖ**	-2*Log-Benzerlik	ABÖ	BBÖ
100	5	0,00	-339,9	699,9	725,9	-338,4	704,7	741,2
	15	0,00	-1019,5	2099,0	2177,2	-1008,8	2105,6	2220,2
	25	0,00	-1700,4	3500,8	3631,1	-1683,2	3514,3	3707,1
500	5	0,00	-1726,7	3473,4	3515,5	-1725,3	3478,5	3537,5
	15	0,00	-5179,5	10418,9	10545,4	-5170,0	10428,1	10613,5
	25	0,00	-8628,6	17357,2	17568,0	-8609,8	17367,6	17679,5
1000	5	0,00	-3460,1	6940,1	6989,2	-3458,3	6944,5	7013,2
	15	0,00	-10375,4	20810,8	20958,0	-10365,8	20819,6	21035,5
	25	0,00	-17295,6	34691,1	34936,5	-17277,7	34703,5	35066,6
5000	5	0,00	-17322,6	34665,3	34730,4	-17321,2	34670,4	34761,7
	15	0,00	-51964,4	103988,7	104184,2	-51960,2	104008,4	104295,2
	25	0,00	-86609,5	173318,9	173644,8	-86595,2	173338,5	173820,7

\* ABÖ: Akaike Bilgi Ölçütü, \*\* BBÖ: Bayesian Bilgi Ölçütü

Tablo 4. Boyutlar Arası Korelasyonun 0,50 Olduğu Durumda Teste ait Model Veri Uyumunun Değerlendirilmesi

Örneklem Büyüklüğü	Test Uzunluğu	Kuramlar						
		TBPoMTK	TBMTK	ÇBMTK				
		H	-2*Log-Benzerlik	ABÖ*	BBÖ**	-2*Log-Benzerlik	ABÖ	BBÖ
100	5	0,00	-340,6	701,2	727,3	-339,2	706,5	742,9
	15	0,00	-1020,6	2101,2	2179,3	-1010,5	2109,0	2223,7
	25	0,00	-1699,7	3499,4	3629,7	-1681,3	3510,5	3703,3
500	5	0,00	-1727,4	3474,8	3516,9	-1726,2	3480,3	3539,3
	15	0,00	-5177,0	10414,0	10540,5	-5168,5	10425,0	10610,5
	25	0,00	-8630,7	17361,4	17572,2	-8616,8	17381,5	17693,4
1000	5	0,00	-3459,1	6938,1	6987,2	-3457,7	6943,5	7012,2
	15	0,00	-10375,6	20811,2	20958,4	-10365,7	20819,3	21035,2
	25	0,00	-17295,4	34690,8	34936,2	-17274,0	34695,9	35059,1
5000	5	0,00	-17322,3	34664,6	34729,8	-17320,7	34669,3	34760,6
	15	0,00	-51966,7	103993,4	104189,0	-51954,5	103996,9	104283,7
	25	0,00	-86609,4	173318,9	173644,7	-86591,4	173330,9	173813,1

\* ABÖ: Akaike Bilgi Ölçütü, \*\* BBÖ: Bayesian Bilgi Ölçütü

Tablo 2, 3 ve 4 incelendiğinde, boyutlar arası korelasyondaki değişimden bağımsız bir şekilde, örneklem büyüklüğü ve test uzunluğu arttıkça TBMTK ve ÇBMTK için üretilen -2\*log-benzerlik, ABÖ ve BBÖ değerlerinin de arttığı görülmektedir. Daha küçük model veri uyumu değerleri, daha iyi model veri uyumuna işaret edeceği için daha az maddeden oluşan testlerin TBMTK ve ÇBMTK'ya daha iyi uyum sağladığı söylenebilmektedir. TBMTK'dan elde edilen -2\*log-benzerlik, ABÖ ve BBÖ değerlerinin, her örneklem büyüklüğü ve test uzunluğunda, ÇBMTK'dan daha küçük değerler ürettiği ortaya koyulmuştur. Bu durum, örneklem büyüklüğü ve test uzunluğundan bağımsız olarak her koşulda, TBMTK ile daha iyi model veri uyumu değerleri sağlandığını göstermektedir.

İkinci alt probleme ait bulgular Tablo 5'te yer almaktadır.

Tablo 5. Farklı Simülasyon Düzeneklerinde Elde Edilen Maddelerin Model Veri Uyumu Değerlerinin Ortalamaları

Örneklem Büyükliği	Test Uzunluğu	Boyutlar arası korelasyon	TBPoMTK	TBMTK			ÇBMTK	
			$H_i$	Outfit	Infit	Ki-kare	Outfit	Infit
100	5	,00	0,008	0,998	0,994	33,573	0,916	0,948
		,25	-0,002	0,994	0,996	29,764	0,934	0,950
		,50	-0,012	0,996	0,996	31,262	0,952	0,966
	15	,00	0,000	0,998	0,998	11,457	0,998	0,999
		,25	-0,003	0,999	0,999	11,391	0,999	0,999
		,50	0,001	0,999	0,998	11,508	0,997	1,000
	25	,00	0,000	0,999	0,999	9,638	1,000	0,998
		,25	0,000	0,999	1,000	9,649	0,999	0,999
		,50	-0,001	0,999	0,999	9,480	0,998	1,000
500	5	,00	0,002	1,000	0,998	160,654	0,948	0,960
		,25	-0,002	1,002	0,998	161,566	0,944	0,964
		,50	0,002	1,000	0,998	169,776	0,946	0,962
	15	,00	0,000	0,999	0,999	33,829	1,000	0,999
		,25	0,001	1,001	0,999	31,753	0,999	0,999
		,50	0,000	1,000	0,999	34,189	1,001	0,999
	25	,00	0,000	1,000	1,000	21,097	0,998	0,999
		,25	0,000	1,000	1,000	21,294	0,996	0,997
		,50	0,000	1,000	1,000	21,382	0,999	0,999
1000	5	,00	-0,002	1,000	1,000	349,162	0,948	0,964
		,25	-0,002	0,998	0,998	329,700	0,948	0,964
		,50	0,004	0,998	0,998	332,736	0,950	0,964
	15	,00	0,000	0,999	0,999	62,800	1,000	1,000
		,25	0,000	0,999	1,000	65,631	1,000	0,999
		,50	0,000	0,999	0,999	65,152	1,000	0,999
	25	,00	0,000	1,000	1,000	38,738	1,000	1,000
		,25	0,000	1,000	1,000	37,389	1,000	1,000
		,50	0,000	1,000	1,000	38,819	1,000	1,000
5000	5	,00	0,000	1,000	1,000	1934,454	0,952	0,962
		,25	0,000	1,002	1,000	1861,510	0,950	0,966
		,50	0,000	1,000	1,000	1814,638	0,950	0,964
	15	,00	0,000	0,999	1,000	334,139	1,001	1,001
		,25	0,000	1,000	1,000	314,307	1,000	1,000
		,50	0,001	1,000	1,000	344,761	0,999	0,999
	25	,00	0,000	1,000	1,000	178,720	1,000	1,000
		,25	0,000	1,000	1,000	188,994	1,000	1,000
		,50	0,000	1,000	1,000	202,866	0,999	1,000

Tablo 5'te yer alan bilgiler doğrultusunda  $H_i$  değerleri incelendiğinde, örneklem büyüklüğü ve test uzunluğu arttıkça,  $H_i$  değerlerinin de sifıra yaklaşma eğilimi gösterdiği belirlenmiştir. TBMTK için elde edilen infit ve outfit değerleri incelendiğinde, her iki katsayının da örneklem büyüklüğü ve test uzunluğundaki artış ile birlikte, optimum değer olan bire yaklaştığı belirlenmiştir. ÇBMTK için elde edilen infit ve outfit değerleri incelendiğinde, her iki katsayının da örneklem büyüklüğü ve test uzunluğundaki artış ile birlikte, optimum değer olan bire yaklaştığı belirlenmiştir.



Üçüncü alt probleme ait bulgular Tablo 6’da yer almaktadır.

Tablo 6. Farklı Simülasyon Düzeneklerinde Elde Edilen Maddelerin Model Veri Uyumlarına ait Standart Hata,  $p$  ve  $t$  Değerlerinin Ortalamaları

Örneklem Büyüküğü	Test Uzunluęu	Boyutlar Arası Korelasyon	TBPoMTK			ÇBMTK		
			$H_i$ (SH)	Outfit (t)	Infit (t)	Ki-kare (p)	Outfit (t)	Infit (t)
100	5	,00	-0,058	-0,038	-0,034	0,072	-0,602	-0,550
		,25	-0,060	-0,050	-0,052	0,132	-0,656	-0,608
		,50	-0,052	-0,040	-0,044	0,102	-0,498	-0,454
	15	,00	-0,030	-0,025	-0,018	0,317	-0,013	-0,017
		,25	-0,030	-0,017	-0,018	0,333	-0,018	-0,015
		,50	-0,030	-0,024	-0,018	0,329	-0,018	-0,006
	25	,00	-0,020	-0,012	-0,013	0,391	-0,008	-0,011
		,25	-0,020	-0,017	-0,014	0,388	-0,024	-0,024
		,50	-0,020	-0,018	-0,015	0,398	-0,008	-0,006
500	5	,00	-0,020	-0,024	-0,016	0,000	-1,222	-1,080
		,25	-0,020	-0,022	-0,024	0,000	-1,242	-1,104
		,50	-0,020	-0,020	-0,014	0,000	-1,224	-1,098
	15	,00	-0,010	-0,012	-0,008	0,147	-0,015	-0,015
		,25	-0,010	-0,014	-0,009	0,168	-0,011	-0,011
		,50	-0,010	-0,013	-0,013	0,160	-0,010	-0,011
	25	,00	-0,010	-0,010	-0,006	0,225	-0,012	-0,006
		,25	-0,010	-0,009	-0,009	0,206	-0,020	-0,017
		,50	-0,010	-0,008	-0,008	0,202	-0,005	-0,002
1000	5	,00	-0,020	-0,018	-0,016	0,000	-1,616	-1,434
		,25	-0,020	-0,018	-0,010	0,000	-1,624	-1,440
		,50	-0,020	-0,018	-0,022	0,000	-1,664	-1,488
	15	,00	-0,010	-0,006	-0,009	0,100	-0,009	-0,005
		,25	-0,010	-0,010	-0,010	0,115	-0,035	-0,028
		,50	-0,010	-0,007	-0,007	0,106	-0,041	-0,037
	25	,00	-0,010	-0,005	-0,006	0,152	-0,010	-0,010
		,25	-0,010	-0,003	-0,006	0,141	-0,005	-0,004
		,50	-0,010	-0,006	-0,006	0,146	-0,006	-0,006
5000	5	,00	-0,010	-0,008	-0,008	0,000	-3,540	-3,140
		,25	-0,010	-0,008	-0,010	0,000	-3,530	-3,132
		,50	-0,010	-0,008	-0,004	0,000	-3,626	-3,216
	15	,00	0,000	-0,003	-0,003	0,073	-0,013	-0,005
		,25	0,000	-0,005	-0,002	0,051	-0,008	-0,003
		,50	0,000	-0,005	-0,003	0,033	-0,031	-0,011
	25	,00	0,000	-0,004	-0,001	0,057	-0,001	-0,003
		,25	0,000	-0,002	-0,003	0,065	-0,007	-0,002
		,50	0,000	-0,003	-0,002	0,060	-0,001	-0,006

Tablo 6 incelendięinde,  $H_i$  katsayısına ait standart hatanın örneklem büyüklüęü ve test uzunluęu arttıkça, bu katsayıda azalma eğilimi gösterdięi belirlenmiştir. TBMTK için elde edilen infit ve outfite ait standart hata deęerleri incelendięinde, her iki durumda da örneklem büyüklüęü ve test uzunluęundaki artış ile birlikte,  $t$  deęerlerinin de sıfıra yaklaştıęı görülmektedir. ÇBMTK için elde edilen infit ve outfite ait standart hata deęerlerine göre, her iki durum için de örneklem büyüklüęü ve test uzunluęundaki artış ile

birlikte, t değerlerinin de sıfıra yaklaştığı görülmektedir. TBMTK’da elde edilen değerler, ÇBMTK’da elde edilen değerlerden daha küçüktür.

Dördüncü alt probleme ait bulgular Tablo 7’de yer almaktadır.

Tablo 7. Farklı Simülasyon Düzeneklerinde Elde Edilen Madde Parametresi Değerlerinin Ortalamaları

Örneklem Büyükliği	Test Uzunluğu	Boyutlar Arası Korelasyon	TBPoMTK		TBMTK		ÇBMTK		
			H <sub>i</sub>	p	a	b	a1	a2	d
100	5	,00	0,008	0,510	0,462	0,696	0,106	-0,328	0,212
		,25	-0,002	0,502	0,030	-0,562	0,116	0,042	0,002
		,50	-0,012	0,492	0,758	0,414	-0,434	0,174	-0,106
	15	,00	0,000	0,495	0,215	-1,217	0,383	-0,009	-0,078
		,25	-0,003	0,501	0,001	-9,328	0,351	-0,604	0,027
		,50	0,001	0,501	0,109	0,148	0,007	0,315	-0,024
	25	,00	0,000	0,498	0,078	0,038	-0,112	0,084	0,008
		,25	0,000	0,498	-0,076	0,186	0,030	0,141	-0,005
		,50	-0,001	0,498	-0,062	-0,892	0,005	0,114	0,008
500	5	,00	0,002	0,500	0,424	-0,270	0,006	0,026	-0,014
		,25	-0,002	0,504	0,314	-1,922	0,078	0,002	0,020
		,50	0,002	0,500	0,386	0,176	0,222	0,100	0,000
	15	,00	0,000	0,499	-0,143	0,215	-0,055	0,079	-0,003
		,25	0,001	0,499	0,003	0,145	0,058	0,007	-0,004
		,50	0,000	0,500	-0,085	1,029	-0,001	-0,001	0,001
	25	,00	0,000	0,499	0,024	0,009	0,073	-0,001	0,008
		,25	0,000	0,505	0,026	-0,604	0,001	0,014	0,003
		,50	0,000	0,501	0,020	0,286	-0,040	0,020	0,004
1000	5	,00	-0,002	0,498	0,106	0,192	0,042	-0,002	-0,010
		,25	-0,002	0,500	0,090	0,422	0,066	-0,034	-0,002
		,50	0,004	0,500	0,334	0,418	0,012	0,010	-0,002
	15	,00	0,000	0,501	-0,058	3,081	-0,261	0,053	0,028
		,25	0,000	0,500	0,085	0,700	0,043	0,048	0,000
		,50	0,000	0,499	-0,030	-1,979	0,057	0,031	-0,001
	25	,00	0,000	0,499	0,000	0,218	-0,049	0,024	0,001
		,25	0,000	0,499	-0,012	-0,094	0,040	0,034	0,002
		,50	0,000	0,500	-0,012	0,076	0,000	0,007	0,001
5000	5	,00	0,000	0,500	0,026	0,136	-0,010	0,006	-0,002
		,25	0,000	0,500	0,012	1,180	-0,118	0,012	0,002
		,50	0,000	0,500	0,204	-0,184	0,188	-0,064	0,004
	15	,00	0,000	0,500	0,015	-0,139	0,017	-0,001	-0,001
		,25	0,000	0,499	0,001	1,623	0,003	0,015	0,000
		,50	0,001	0,499	0,015	0,241	0,020	0,015	-0,001
	25	,00	0,000	0,499	0,000	-1,052	-0,012	0,012	0,001
		,25	0,000	0,499	0,018	-0,516	-0,010	-0,002	0,003
		,50	0,000	0,501	0,006	0,018	-0,003	0,006	0,001

Tablo 5 ve 7’de hesaplanan H<sub>i</sub> değerlerinin aynı olduğu görülmektedir. TBPoMTK’da güçlük indeksi olarak kullanılan klasik güçlük parametresi (p), tüm simülasyon düzeneklerinde orta düzeyde güçlüğü ifade etmektedir. TBMTK ve ÇBMTK’da madde ayırt ediciliğinin, örneklem büyüklüğündeki artış ile birlikte, sıfıra yaklaşma eğilimi gösterdiği belirlenmiştir. TBMTK’da örneklem büyüklüğünün küçük olduğu durumlarda, b

parametresi daha düşük değerler almıştır. ÇBMTK’da ise madde güçlüğünü belirlemek için üretilen d parametresine ait bulgular incelendiğinde, örneklem büyüklüğündeki artış ile birlikte, bu parametreye ait değerlerde de artış gözlenmiştir.

Beşinci alt probleme ait bulgular Tablo 8’de yer almaktadır.

Tablo 8. Farklı Simülasyon Düzeneklerinde Elde Edilen Madde Parametrelerine ait Standart Hata Değerlerinin Ortalamaları

Örneklem Büyükliği	Test Uzunluğu	Boyutlar Arası Korelasyon	TBPoMTK		TBMTK		ÇBMTK		
			H <sub>i</sub> (SH)	p (SH)	a (SH)	b (SH)	a1 (SH)	a2 (SH)	d (SH)
100	5	,00	-0,058	0,050	9,530	50,102	1,830	1,434	1,276
		,25	-0,060	0,050	12,256	23,398	1,994	2,016	0,584
		,50	-0,052	0,050	7,924	29,992	1,450	1,560	0,500
	15	,00	-0,030	0,050	10,734	408,527	0,867	0,816	0,375
		,25	-0,030	0,050	8,454	32,580	1,326	1,562	0,439
		,50	-0,030	0,050	5,578	29,103	1,238	1,282	0,397
	25	,00	-0,020	0,050	1,543	0,078	0,439	0,538	0,274
		,25	-0,020	0,050	1,523	31,302	0,544	0,694	0,272
		,50	-0,020	0,050	0,982	22,449	0,466	0,513	0,248
500	5	,00	-0,020	0,020	2,508	7,732	0,426	0,414	0,116
		,25	-0,020	0,020	3,320	334,138	0,310	0,302	0,112
		,50	-0,020	0,020	2,836	14,636	0,224	0,234	0,104
	15	,00	-0,010	0,020	4,062	26,612	0,220	0,218	0,104
		,25	-0,010	0,020	1,783	18,064	0,221	0,217	0,101
		,50	-0,010	0,020	2,847	225,313	0,207	0,219	0,105
	25	,00	-0,010	0,020	0,220	387,654	0,122	0,166	0,094
		,25	-0,010	0,020	0,212	140,422	0,238	0,238	0,106
		,50	-0,010	0,020	0,281	56,620	0,256	0,266	0,118
1000	5	,00	-0,020	0,020	1,862	2,722	0,206	0,202	0,076
		,25	-0,020	0,020	2,606	41,868	0,174	0,182	0,074
		,50	-0,020	0,020	1,526	17,994	0,308	0,258	0,078
	15	,00	-0,010	0,020	1,696	3565,255	0,107	0,123	0,064
		,25	-0,010	0,020	0,643	95,335	0,127	0,125	0,070
		,50	-0,010	0,020	1,201	2401,455	0,133	0,136	0,070
	25	,00	-0,010	0,020	0,190	58,553	0,111	0,110	0,063
		,25	-0,010	0,020	0,305	250,394	0,125	0,128	0,065
		,50	-0,010	0,020	0,180	31,674	0,118	0,120	0,064
5000	5	,00	-0,010	0,010	0,484	8,080	0,066	0,062	0,030
		,25	-0,010	0,010	0,784	169,836	0,090	0,094	0,030
		,50	-0,010	0,010	0,840	2,542	0,084	0,086	0,030
	15	,00	0,000	0,010	0,121	22,133	0,073	0,073	0,030
		,25	0,000	0,010	0,124	677,529	0,077	0,077	0,030
		,50	0,000	0,010	0,129	13,505	0,071	0,068	0,030
	25	,00	0,000	0,010	0,103	839,074	0,051	0,049	0,030
		,25	0,000	0,010	0,106	237,863	0,047	0,049	0,030
		,50	0,000	0,010	0,101	32,481	0,062	0,059	0,030

Tablo 8 incelendiğinde, H<sub>i</sub> katsayısına ait standart hatanın örneklem büyüklüğü ve test uzunluğu arttıkça, azalma eğilimi gösterdiği belirlenmiştir. TBPoMTK’da güçlük düzeyini belirlemek için kullanılan p parametresine ait standart hata değerleri incelendiğinde, örneklem büyüklüğü 100 iken 0,05; örneklem büyüklüğü 500 ve 1000 iken

0,02 ve örneklem büyüklüğü 5000 iken 0,01 değerlerinin elde edildiği belirlenmiştir. TBMTK için madde ayırt ediciliğini belirlemek amacıyla kullanılan a parametresine ait standart hata değerleri incelendiğinde, örneklem büyüklüğündeki artış ile birlikte, standart hata değerlerinin sıfıra yaklaşma eğilimi gösterdiği belirlenmiştir. TBMTK için madde güçlüğünü belirlemek için kullanılan b parametresine ait standart hata değerleri incelendiğinde, örneklem büyüklüğü, test uzunluğu ve boyutlar arası korelasyon koşullarındaki değişim ile birlikte, bu parametreye ait standart hata değerlerindeki değişim bir örüntü ortaya koymamaktadır. ÇBMTK’da madde ayırt ediciliği için kullanılan  $a_1$  ve  $a_2$  parametrelerine ait bulgular incelendiğinde, örneklem büyüklüğündeki artış ile birlikte, standart hata değerlerinin sıfıra yaklaşma eğilimi gösterdiği belirlenmiştir. ÇBMTK için madde güçlüğünü belirlemek amacıyla kullanılan d parametresine ait standart hata değerleri incelendiğinde, örneklem büyüklüğündeki artış ile birlikte, standart hata değerlerinin sıfıra yaklaşma eğilimi gösterdiği belirlenmiştir. Test uzunluğundaki artış ile birlikte, d parametresine ait standart hata değerlerinin azalma eğilimi gösterdiği, ancak birbirine oldukça yakın değerler verdiği belirlenmiştir.

### SONUÇLAR ve TARTIŞMA

Meijer, Sijstma & Smid (1990) yaptıkları araştırmada, MHM’nin veriye uygulandığı her durumda kişilere ait sıralı ölçmeler yapılabilmesi açısından TBMTK’dan daha üstün olduğunu öne sürmüşlerdir. Bu araştırma ile bu bulgu desteklenememiştir. Teste ait model veri uyumu örneklem büyüklüğü ve test uzunluğu açısından genel olarak incelendiğinde, örneklem büyüklüğü ve test uzunluğu arttıkça, boyutlar arası korelasyon ve test uzunluğuna göre değişen simülasyon düzeneklerinde elde edilen model veri uyumu değerlerinin de arttığı belirlenmiştir. Teste ait model veri uyumu boyutlar arası korelasyon açısından incelendiğinde, boyutlar arası korelasyon ile farklılaşan model veri uyumu değerleri arasında bir örüntü bulunamamıştır. Teste ait model veri uyumunu en fazla etkileyen bağımsız değişkenlerin örneklem büyüklüğü ve test uzunluğu olduğu tespit edilmiştir.

Zhou (2011), TBMTK ve TBPoMTK’da maddelere ait uyum indekslerinin birbirine yakın değerler ürettiğini belirtmiştir. Bu bulgu, bu araştırma bulguları ile farklılık göstermektedir. Köse’nin (2010) araştırma bulgularından farklı olarak, ÇBMTK’da maddelere ait model veri uyumu değerleri, TBMTK’ya göre daha düşüktür. Maddeler, TBMTK’ya daha iyi uyum göstermişlerdir. Maddelere ait model veri uyumunu en fazla etkileyen bağımsız değişkenler örneklem büyüklüğü ve test uzunluğudur.

Cavanagh & Waugh’a (2011, 251) göre, outfit ve infit t değerlerinin -2 ile +2 aralığında yer alması gerekmektedir. Sıfıra yaklaşan değerler, ilgili maddelere ait model veri uyumunun yüksek olduğunu ifade etmektedir. Bu değerlerin dışında kalan değerlere sahip maddeler ise testten çıkarılmalıdır. Outfit ve infit için, tüm simülasyon düzeneklerinde elde edilen t değerleri ( $-0,001 \leq t \leq -0,052$ ), Cavanagh & Waugh’un (2011, 251) belirttiği sınır değerlerin içindedir. TBMTK için maddelere ait model veri uyumu, tüm simülasyon düzeneklerinde sağlanmıştır. ÇBMTK için ise, outfit ve infit için, örneklem büyüklüğü 5000 ve test uzunluğu 5 iken, tüm boyutlar arası korelasyon koşullarında elde edilen t değerleri ( $-3,132 \leq t \leq -3,626$ ), Cavanagh & Waugh’un (2011, 251) belirttiği sınır değerlerin dışındadır. Stone’un (1992) araştırması ile bu araştırma bulguları benzerlik göstermektedir. Her iki araştırmada da küçük örneklem büyüklükleri ve kısa testlerde, daha büyük ayırt edicilik parametresi değerleri elde edilmiştir. Zeng (1989), bazı durumlarda ayırt edicilik parametresinin boyutluluktan etkilendiğini ortaya koymuştur. Ancak bu araştırmada, benzer bir bulgu elde edilmemiştir. Ansley & Forsyth (1985), Way, Ansley & Forsyth (1988) ve Zeng (1989), bu araştırmada olduğu gibi, basit yapıda iki boyutlu bir veri seti üzerinde çalışmışlardır. TBMTK ve ÇBMTK’dan elde edilen a parametresine ait bulgular, Ansley &

Forsyth (1985), Way, Ansley & Forsyth (1988) ve Zeng (1989) araştırmalarına benzerlik göstermemektedir. Ansley & Forsyth (1985) ve Way, Ansley & Forsyth (1988), bu iki kurama ait a parametresi değerlerini birbirine yakın bulurken, bu araştırmada oldukça farklı bulgular elde edilmiştir. Güçlük parametresine ait bulgular ise, Ansley & Forsyth (1985) ve Way, Ansley & Forsyth'un (1988) araştırmalarına benzerlik göstermektedir. Her iki araştırmada da TBMTK'ya ait bulgular, ÇBMTK'ya ait bulgulardan daha yüksektir. Ancak Zeng (1989), her iki kuramdan elde edilen b parametresi değerlerini de %80'e varan oranlarda benzer bulmuştur. Köse'nin (2010) araştırmasına benzer olarak, madde parametresini etkileyen en önemli bağımsız değişken örneklem büyüklüğüdür. Bununla birlikte, test uzunluğunun da madde parametresini etkilediği belirlenmiştir.

Stochl, Jones & Croudace'in (2012) araştırmasında da desteklediği gibi, az madde sayısı ve az katılımcı ile gerçekleştirilen araştırmalarda TBPoMTK kullanılabilir. Boyutlar arası korelasyondaki değişim ise standart hatayı etkilememiştir. Smits, Timmerman & Meijer (2012) ise, özellikle boyutlar arası korelasyonun düşük olduğu durumlarda Mokken ile ölçeklemenin iyi bir alternatif olduğunu ortaya koymuşlardır. Bu iki araştırma bulgusu farklılık göstermektedir. Test uzunluğu ve boyutlar arası korelasyondan bağımsız olarak, örneklem büyüklüğündeki artış, p parametresine ait standart hata değerlerinde azalmaya neden olmaktadır. Bu bulgu, Smits, Timmerman & Meijer'in (2012) bulgularında benzerlik göstermektedir. TBMTK için madde ayırt ediciliğini belirlemek amacıyla kullanılan a parametresine ait standart hata değerleri incelendiğinde, örneklem büyüklüğündeki artış ile birlikte, standart hata değerlerinin sıfıra yaklaşma eğilimi gösterdiği belirlenmiştir. Elde edilen bu bulgu Baker'ın (1991) ve Stone'un (1992) araştırmalarıyla benzerlik göstermektedir. Ancak Baker (1991), yalnızca 30 ile 240 arasında değişen bir örneklem büyüklüğü tercih ettiği için bu iki araştırma bulgusunun tam anlamıyla paralellik gösterdiği söylenemez. Test uzunluğundaki artış ile birlikte, a parametresine ait standart hata değerlerinin azalma eğilimi gösterdiği belirlenmiştir. Baker'ın (1991) ve Stone'un (1992) araştırmalarıyla bu bulgu benzerlik göstermektedir. TBMTK için madde güçlüğünü belirlemek için kullanılan b parametresine ait standart hata değerleri incelendiğinde, örneklem büyüklüğü, test uzunluğu ve boyutlar arası korelasyon koşullarındaki değişim ile birlikte, bu parametreye ait standart hata değerlerindeki değişim bir örüntü ortaya koymamaktadır. Bu bulgular Baker'ın (1991) araştırmasıyla benzerlik göstermemektedir. Bu durumun sebebi olarak, Baker'ın (1991) araştırmasına göre çok daha büyük örneklem büyüklüklerinin kullanılması gösterebilir. Ancak, Stone'un (1992) araştırmasıyla benzerlik göstermektedir. ÇBMTK'da madde ayırt ediciliği için kullanılan  $a_1$  ve  $a_2$  parametrelerine ait bulgular incelendiğinde, örneklem büyüklüğündeki artış ile birlikte, standart hata değerlerinin sıfıra yaklaşma eğilimi gösterdiği belirlenmiştir. Bu bulgu, Bolt & Lall'in (2003) araştırmasındaki bulgulara benzerlik göstermektedir. Test uzunluğundaki artış ile birlikte, d parametresine ait standart hata değerlerinin azalma eğilimi gösterdiği, ancak birbirine oldukça yakın değerler verdiği belirlenmiştir. Bolt & Lall'in (2003) araştırma bulguları ile bu araştırma bulguları benzerlik göstermektedir. Köse'nin (2010) araştırma bulgularına benzer olarak, ÇBMTK'da madde parametreleri için elde edilen hata değerleri, TBMTK'ya göre daha düşüktür.

Teste ait model uyumunun daha iyi düzeyde sağlanabilmesi için kısa testlerin ve küçük örneklem büyüklüklerinin kullanılması önerilmektedir. Ancak, testteki madde sayısı ve örneklem büyüklüğü arttıkça, teste ait model veri uyumu daha kararlı sonuçlar vermektedir. Her simülasyon koşulunda, maddeler için model veri uyumunu değerlendirmek amacıyla TBPoMTK kullanılabilir. TBMTK ve ÇBMTK'da, daha az hatalı madde parametresi değerleri elde edebilmek için, daha büyük örneklem büyüklüğü ve test uzunluğu ile çalışılmalıdır. Boyutlar arası korelasyon değerlerindeki değişim, madde ve test için



model uyumları ile madde parametreleri ve hata değerleri üzerinde çok az etkili olmuştur. Bunun nedeni olarak araştırma için kullanılan boyutlar arası korelasyon değerlerinin 0,00 - 0,50 arasında değişmesi gösterilebilmektedir. Ansley & Forsyth (1985) ve Ackerman'ın (1987) araştırmalarında 0,90 ve üzeri boyutlar arası korelasyon değerlerinde ayırıcılık ve güçlük düzeylerinin değiştiği belirlenmiştir. Bu nedenle, araştırmacıların boyutlar arası korelasyondaki değişime bağlı kalmamaları önerilmektedir.

Daha küçük örneklemelerde (100, 250, 500 ve 1000), daha büyük test uzunluklarında (>25) ve daha yüksek boyutlar arası korelasyon değerlerinde ( $\geq 0,75$ ) benzer bir çalışma gerçekleştirilebilir. Benzer simülasyon koşullarında yetenek parametreleri de tahmin edilerek, benzer bir araştırma yapılabilir. Standart hata dışında farklı hata değerleri incelenerek çalışma tekrarlanabilir. Farklı parametreler sabit tutularak, diğer parametrelerdeki değişim, farklı bir araştırma ile incelenebilir.

## KAYNAKLAR

- Ackerman, T. A. (1996). Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement*, 20, 311-329.
- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*, 22, 37-51.
- Ansley, T. N., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement*, 9, 37-48.
- Baker, F. B. (1991). Comparison of minimum logit chi-square and bayesian item parameter estimation. *British Journal of Mathematical and Statistical Psychology*, 44, 299-313.
- Batley, R. M. & Boss, M. W. (1993). The effects on parameter estimation of correlated dimensions and a distribution-restricted trait in a multidimensional item response model. *Applied Psychological Measurement*, 17, 131 – 141.
- Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using markov chain monte carlo. *Applied Psychological Measurement*, 27(6), 395-414.
- Cavanagh, R. F., & Waugh, R. F. (2011). *Applications of rasch measurement in learning environments research* (Eds. Vol. 2). Sense Publishers.
- Chia-Lin, K., I-Ping, H., Wen-Chung, W., Ching-Fan, S., Tzu-Ying, Y., Chun-Hou, W., & Ching-Lin, H. (2006). Validation of the action research arm test using item response theory in patients after stroke. *J Rehabil Med*, 38, 375-380.
- Goodman, J. T. (2008). *An examination of the residual covariance structures of complex performance exercises under various scaling and scoring methods*. Unpublished Doctoral Dissertation. The University of North Carolina.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory principles and applications*. Boston: Kluwer.
- Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice*, 8(1), 35-41.
- Harwell, M. R., Stone, C. A., Hsu, T.-C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement*, 20(2), 101-125.
- Henard, D. H. (2000). Item response theory, in *Reading and Understanding MORE Multivariate Statistics*, Vol. II, Larry Grimm and Paul Yarnold (Eds). Washington, DC: American Psychological Association, 67-97.
- Köse, İ. A. (2010). *Madde tepki kuramına dayalı tek boyutlu ve çok boyutlu modellerin test uzunluğu ve örneklem büyüklüğü açısından karşılaştırılması*, Yayınlanmamış Doktora Tezi. Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü.
- Linden, W., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.
- Lee, S. H. (2007). *Multidimensional item response theory: A SAS MDIRT MACRO and empirical study of PIAT MATH test*. Unpublished Doctoral Dissertation. The University of Oklahoma.
- Luecht, R. M. (2004). *Mirtgen (Version 2)*. Greensboro, NC.



- Meijer, R. R., Sijtsma, K., & Smid, N. G. (1990). Theoretical and empirical comparison of the mokken and the rasch approach to IRT. *Applied Psychological Measurement*, 14(3), 283-298.
- Mokken, R. J. (1971). *A theory and procedure of scale analysis*. De Gruyter, Berlin, Germany.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.
- Seungho Yang, M. A. (2007). *A comparison of unidimensional and multidimensional rasch models using parameter estimates and fit indices when assumption of unidimensionality is violated*. Unpublished Doctoral Dissertation. The Ohio State University.
- Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.
- Smits, I. A. M., Timmerman, M. E., & Meijer, R. R. (2012). Exploratory mokken scale analysis as a dimensionality assessment tool: why scalability does not imply unidimensionality. *Applied Psychological Measurement*, 36, 516-539.
- Stochl, J. (2007). Nonparametric extension of item response theory models and its usefulness for assessment of dimensionality of motor yests. *Acta Universitatis Carolinae*, 42(1), 75-94.
- Stochl, J., Jones, P. B., & Croudace, T. J. (2012). Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric irt method in empirical research for applied health researchers. *BMC Med Res Methodol*, 12:74.
- Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: an evaluation of MULTILOG. *Applied Psychological Measurement*, 16, 1-16.
- Sünbül, Ö. (2011). *Çeşitli boyutluluk özelliğine sahip yapılarda, madde parametrelerinin değişmezliğinin klasik test teorisi, tek boyutlu madde tepki kuramı ve çok boyutlu madde tepki kuramı çerçevesinde incelenmesi*, Yayınlanmamış Doktora Tezi. Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü.
- Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). The comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimates. *Applied Psychological Measurement*, 12(3), 239-259.
- Zeng, L. (1989). *Robustness of unidimensional latent trait models when applied to multidimensional data*. Unpublished Doctoral Dissertation, Georgia University, Athens.
- Zhou, Y. (2011). *Comparing parametric item response theory and nonparametric item response theory: application in psychological research using polytomous items* Unpublished Doctoral Dissertation. Fordham University, New York.

## EXTENDED ABSTRACT

### **Introduction**

The worked topics about educational sciences and psychology are structures with generally more than one latent trait, not just one latent trait. To achieve most valid, stable and practical conclusions, multidimensional structure analysis must be practicing. Another feature of the worked topics about educational sciences and psychology are analyzing the survey questions (independent questions) and scaling studies like ordering the person and items. Latent traits not only form a pattern, but also independent structures can be analyzed. Also, the purpose of the measure can be just ordering the person and items. Some situations, analysis must be done with quite a little sample sizes. In case of other situations, small tests preferred. To achieve all this purposes, item response theory applications developed. The purpose of this study is to identify NIRT, PIRT and MIRT across different sample sizes, test length and correlation between dimensions in a two dimensional simple structures. The aim is also to identify and compare and contrast the results of item parameters, model data fit for the items and model data fit for the test.

### **Method**

In order to achieve the purpose, the data sets with two dimensional structures and binary scoring properties, have been generated. These generated data sets display a simple structured feature. In doing so, NIRT, PIRT and MIRT model applications have been applied in the same data sets. Data in various sample sizes have been simulated. These sample sizes have been identified as 100, 500, 1000 and 5000. Moreover, to investigate the effect of test length, while limiting the first factors with the number of 25 items, the numbers

of second factor items are 5, 15 and 25. As for the correlation between dimensions, the correlation between simulated data sets and estimated ability parameters have been determined to vary from 0.00, 0.25 to 0.50. From each experimental design, within the frame of Monte Carlo study, the findings have been obtained through 20 replications. These replications have been simulated under univariate and multivariate normal distributions.

With the purpose of specifying the item parameters, for NIRT theory  $\alpha_i$  (scalability) and  $\delta_i$  (p values); for PIRT theory a and b values; for MIRT theory  $\alpha_1$ ,  $\alpha_2$  and d values have been estimated. In order to specify the model data fit for the items, for NIRT theory scalability coefficients for items; for PIRT theory infit, outfit and chi-square fit statistics; for MIRT theory infit and outfit statistics have been obtained. For specifying the model data fit for the test, however, for NIRT theory scalability coefficients for test; for PIRT and MIRT theories log-likelihood statistics and some information criteria (akaike and bayesian) have been revealed. In addition, for the item parameters and model data fit for the items, standard errors and significance (p) values have been calculated.

### ***Results and Discussion***

Having analyzed the findings of the research, with the increase of sample sizes and test length, it is also found out that the model data fit for the test has increased as well. Different correlation between dimensions setup, model data fit for the test values has not formed a pattern. It can be stated that tests consisting of less items fit better to MIRT models.

In all simulation designs, model data fit for the items are calculated with quite low errors in NIRT. When the chi-square, infit and outfit values obtained for PIRT have been analyzed, it has been revealed that along with the increase of sample sizes and test length, all three coefficients exhibit better model fit. When MIRT infit and outfit values have been analyzed, similar to PIRT, the MIRT infit and outfit values have been found to be closer to optimum value one with the increase of sample sizes and test length. The correlation between dimensions showed little effect on chi-square, infit and outfit values. When standard errors belonging to infit and outfit values by PIRT and MIRT, have been analyzed, it has been observed that t values get closer to one along with the increase of sample sizes and test length.

In NIRT, the standard errors belonging to  $H_i$  and p parameters tend to decrease with the increase of sample sizes and test length. In PIRT, a parameters tend to decrease when the sample sizes and test length increase. The change in any simulation designs hasn't showed an impact on b parameters. In MIRT,  $a_1$  and  $a_2$  parameters tend to decrease along with increase of sample sizes. There are not any patterns between the change of test length and correlation between dimensions with  $a_1$  and  $a_2$  parameters. In PIRT, the standard errors belonging to a parameters are found to get closer to zero with the increase in sample sizes and test length. In MIRT, the standard errors belonging to  $a_1$  and  $a_2$  parameters displayed similar results.