

Farklı Boyutluluk Özelliklerindeki Basit ve Karmaşık Yapılı Testlerin Çok Boyutlu Madde Tepki Kuramına Dayalı Parametre Kestirimlerinin İncelenmesi*

Examining the Parameter Estimations of Simple and Complex Structured Tests with Various Dimensionality Properties Based on Multidimensional Item Response Theory

Derya Çakıcı Eser**

Selahattin Gelbal***

Öz

Bu çalışmada basit ve karmaşık yapıdaki iki ve üç boyutlu testlerin farklı test uzunluğu ve örneklem büyüklüğü koşullarında parametre kestirimleri yapılmış, kestirimlere ilişkin olarak elde edilen RMSE, yanlılık ve kestirilen parametreler ile gerçek parametreler arasındaki ilişki incelenmiştir. Çalışmanın verileri iki boyutlu basit, iki boyutlu karmaşık, üç boyutlu karmaşık ve üç boyutlu basit yapıya uygun olacak şekilde simülasyon yoluyla üretilmiştir. Veri setlerinin test uzunluğu 12 ve 48 madde; örneklem büyüklüğü 1000, 2000 ve 4000 olacak şekilde değiştirilmiştir. Bu şekilde ele alınan dört test yapısı ile 2 test uzunluğu ve 3 örneklem büyüklüğü koşulu çaprazlanarak (4x2x3) 24 koşul içeren desen oluşturulmuştur. Oluşturulan her koşula ilişkin olarak, madde ve birey parametreleri sabit tutulmak üzere, 25 tekrar yapılarak toplamda 600 veri seti oluşturulmuş ve analiz edilmiştir. Elde edilen sonuçlara göre madde parametrelerinin parametre kestirim iyiliği hem madde hem birey sayısındaki artışla artmaktadır. Ele alınan her bir koşulda d parametreleri aynı koşuldaki a parametrelerinden daha kararlı kestirilmiştir. Birey parametrelerinin kestirim iyiliği test uzunluğundaki artış ile artmaktadır. Yüksek korelasyon değerleri RMSE'nin düşük olduğu koşullarda elde edilmiştir. Ayrıca çalışmanın sonunda daha kararlı kestirim yapabilmek için gerekli test uzunluğu ve örneklem büyüklüğüne ilişkin öneriler geliştirilmeye çalışılmıştır.

Anahtar Kelimeler: Çok boyutlu madde tepki kuramı, basit yapı, karmaşık yapı, parametre kestirimi

Abstract

Under the scope of this study; parameter estimations of simple and complex structured two and three dimensional tests have been performed according to different test length and sample size conditions; according to the estimations, RMSE, bias and relation between estimated and actual parameters have been investigated. Research data has been generated by means of simulations according to requirements of two dimensional simple structure, two dimensional complex structure, three dimensional complex structure and three dimensional complex structure. The simulation conditions were test length (12 and 48 items) and sample size (1000, 2000 and 4,000). By intercrossing the simulation conditions (4 test structures x 2 test lengths x 3 sample sizes); 24 crossed conditions have been acquired. By fixing item and ability parameters and applying 25 replications; 600 data sets have been generated and analyzed. According to the results, parameter recovery of item parameters increases with the number of items and examinees. For each condition; d parameters have been estimated more accurately than a parameters on the same pattern. Parameter recovery of ability parameters is developed with the increase in test length. High correlation values acquired in the conditions where RMSE values are low. Additionally; suggestions on the required test length and sample size for accurate estimations are provided at the final stage of the study.

* Bu çalışma "Çok Boyutlu Madde Tepki Kuramının Farklı Modellerinden Çeşitli Koşullar Altında Kestirilen Parametrelerin İncelenmesi" başlıklı doktora tezinden oluşturulmuştur.

** Dr., Kırıkkale Üniversitesi, Eğitim Fakültesi, Kırıkkale-Türkiye, deryacacicieser@gmail.com

*** Prof. Dr. Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, sgelbal@gmail.com

Keywords: Multidimensional item response theory , simple structure, complex structure, parameter estimation

GİRİŞ

Madde tepki kuramı (MTK) ilk olarak Lord ve Novick'in (1968) *Statistical Theories Of Mental Test Scores* kitabında Allan Birnbaum tarafından ele alınmıştır. Bu kuramda “yetenek” kavramı ile ilgilenilir ve yetenek θ (teta) simgesi ile gösterilir. Kurama göre bireyin bir testte gösterdiği performansın altında o test ile ölçülmeye çalışılan yeteneği yatar. Kuramın amacı bireyin testle ölçülen yetenek düzeyini tahmin etmek veya kestirmek için bir temel oluşturmaktır. Buna göre bir bireyin yetenek düzeyi, bireyin maddelere verdiği cevaplardan kestirilmeye çalışılır. Kuram ile bireyin bir testte ortaya koyduğu performans ile bu performansın altında yatan gözlenemeyen yeteneği veya özelliği arasında bir ilişki ortaya konur. Gözlenen ve gözlenemeyen bu özellikler arasındaki ilişki matematiksel fonksiyonlar ile açıklanır. Bu sebeple madde tepki modelleri matematiksel modellerdir. MTK modelleri oldukça güçlü varsayımlara dayanır (Hambleton ve Swaminathan, 1985; Embretson ve Reise, 2000). Bu varsayımlar tek boyutluluk ve yerel bağımsızlık başlıkları altında ele alınır. Varsayımlar aşağıda açıklanmıştır:

Tek boyutluluk: Tek boyutluluk test ile tek bir örtük özelliğin ölçülüyor olması şeklinde tanımlanır. Pratikte bunu karşılamak güçtür. Bu sebeple testin baskın bir faktörü ölçtüğünün gösterilmesi tek boyutluluğun sağlanması için yeterlidir (Hambleton ve Swaminathan, 1985; Embretson ve Reise, 2000).

Yerel Bağımsızlık: Yerel bağımsızlık aynı yetenek düzeyindeki bireylerin farklı maddelere verdiği tepkilerin istatistiksel olarak bağımsız olması olarak tanımlanır. Bu varsayım pratikte bireyin bir maddeye verdiği cevabın diğer bir maddeyi olumlu ya da olumsuz etkilememesi anlamına gelir (Hambleton ve Swaminathan, 1985; Embretson ve Reise, 2000).

Yukarıda yer alan varsayımlara ek olarak tüm MTK modellerinin sağlaması gereken koşul kullanılan testin hız testi olmamasıdır. Bir testin hız testi olması testin ilgilenilen yeteneğin yanında performans hızını da ölçtüğü anlamına gelir. Bu durumda testin tek boyutluluk varsayımı ihlal edilmiş olur (Hambleton ve Swaminathan, 1985).

Ancak MTK varsayımlarının karşılanmış olması doğrudan analize geçilebileceği anlamına gelmez. Bunun için öncelikle hangi modele göre analiz yapılacağına karar verilmesi gereklidir. MTK'da modeller cevap kategorisi sayısına göre belirlenir. MTK modelleri iki kategorili modeller ve çok kategorili modeller başlıkları altında toplanır. Bu çalışmanın konusu iki kategorili modeller olduğu için çok kategorili modellere değinilmemiştir.

İki kategorili MTK modelleri lojistik ve ogive modeller şeklinde ele alınır. Modelde yer alan parametrelere bağlı olarak; 1, 2 ve 3 parametrelili modeller şeklinde adlandırılır. 3 parametrelili lojistik model (3PLM)'de bireyin doğru cevap verme olasılığını etkileyen parametreler kişinin yetenek düzeyi ile beraber maddenin; güçlüğü, ayıricılığı ve tahmin (şans) parametresidir. Bu modele göre bir maddenin doğru cevaplanma olasılığı tüm yetenek düzeyleri için her zaman sıfırın üzerindedir. 3 parametrelili lojistik modeldeki bir maddeye ilişkin olasılık fonksiyonu eşitlik 1'de verilmiştir.

$$P(X_{is} = 1 | \theta_s, \beta_i, \alpha_i, \gamma_i) = \gamma_i + (1 - \gamma_i) \frac{\exp[\alpha_i(\theta_s - \beta_i)]}{1 + \exp[\alpha_i(\theta_s - \beta_i)]} \quad (\text{Eşitlik 1})$$

X_{is} = s bireyin i maddesine verdiği cevap (1 veya 0)

θ_s = s bireyinin yetenek düzeyi

β_i = i maddesinin güçlüğü

α_i = i maddesinin ayıricılığı

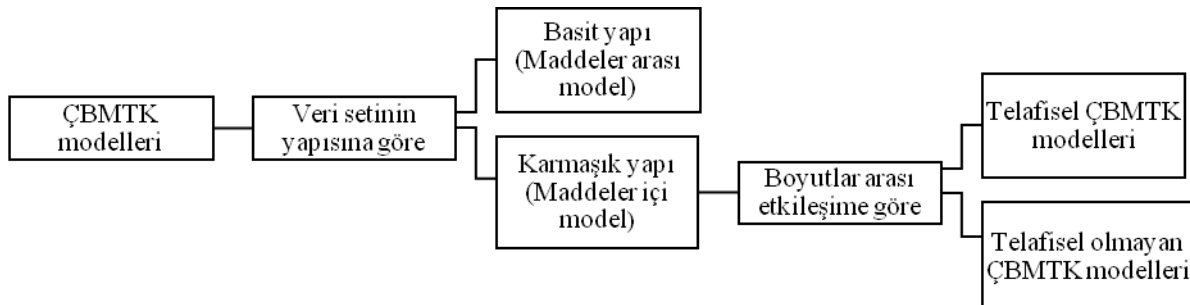
γ_i = i maddesinin en düşük asimptotu (şans parametresi)

Maddelerin şans parametresinin sıfır veya ihmal edilebilir olduğu, ancak maddelerin ayırt edicilik ve güçlük bakımından farklılaştığı durumda 2PLM kullanılır. 2PLM'nin olasılık fonksiyonu eşitlik 1'deki fonksiyondan şans parametresinin çıkartılmış (yi) halidir. Maddelerin sadece güçlük bakımından farklılaştığı, ayırt ediciliklerinin eşit olduğu durumda ise 1PLM kullanılır. 1PLM'nin özel bir durumu olarak Rasch modeli tanımlanır. Rasch modelinde madde güçlükleri farklılaşır, ancak madde ayırtıcılıkları 1'e eşittir. Rasch modelinin lojistik fonksiyonunda madde parametrelerinden sadece güçlük parametresi yer alır.

MTK'da farklı özellikteki durumlara ilişkin birbirinden farklı modeller yer alsa da, varsayımlar karşılanmadığı takdirde bu modellerin kullanılması uygun olmayacaktır. Uygulamalarda en sık karşılaşılan durum, kullanılan testlerin tek boyutlu olmamasıdır. Bu durumda MTK'nın temel varsayımı olan tek boyutluluk ihlal edilmektedir. Bu ve buna benzer durumlarda tek boyutlu MTK modelleri yerine tek boyutlu MTK'nın genişletilmiş hali olan ve çok boyutlu veri setleri için geliştirilmiş olan çok boyutlu madde tepki kuramının (ÇBMTK) kullanılması daha uygundur.

Tek boyutlu madde tepki kuramı yapı veya boyut olarak adlandırılan gözlenemeyen bir değişken ile bireyin belli bir test maddesine doğru cevap verme olasılığı arasındaki ilişkiyi modellemek için kullanılır. Buna karşılık ÇBMTK iki ya da daha fazla gözlenemeyen değişken ile bireyin belli bir test maddesine doğru cevap verme olasılığı arasındaki ilişkiyi modellemek için kullanılır (Ackerman, Gierl ve Walker, 2003). Tek boyutlu MTK modellerinde doğru cevaplama olasılığı tek bir yeteneğe dayanırken, ÇBMTK'da doğru cevaplama olasılığı ölçülmeye çalışılan çok sayıda yeteneğin bir fonksiyonudur. Buna bağlı olarak da ÇBMTK'da ölçme modeli test sonucu olarak tek bir puan yerine bir yetenek profili ortaya koyar (Hartig ve Höhler, 2009). Bir veri setine ÇBMTK'yı uygulamak için veri setinin çok boyutlu olmasının yanında başka varsayımlara da ihtiyaç duyulur. Bu varsayımlardan biri yerel bağımsızlık diğeri monoton artmadır. Yerel bağımsızlık varsayımı tek boyutlu MTK'daki ile aynı anlamı taşıdığından burada tekrar anlatılmamıştır. Monoton artma varsayımı ise bireyin doğru cevap verme olasılığının, θ vektöründeki herhangi bir elemanın daha yüksek değer alması ile artmasıdır (Reckase, 2009). Buna göre bireyin doğru cevap verme olasılığı ile θ 'sı arasında monoton artan bir fonksiyon ortaya konur. Monoton artma varsayımı tek boyutlu MTK'da karşılanabilmesine rağmen bu varsayıma ihtiyaç duymayan tek boyutlu MTK modelleri de mevcuttur. Bu sebeple tek boyutlu MTK'da bir varsayım olarak ele alınmaz (Roberts ve ark 2000; akt: Reckase, 2009).

ÇBMTK'da kullanılan farklı modeller, yetenek ile doğru cevaplama olasılığı arasında farklı istatistiksel ilişkileri varsayar. Buna ek olarak boyutlar ve maddeler arasındaki ilişki veri setinin yapısına göre tanımlanabilir. ÇBMTK'da modeller veri setinin karmaşıklığına göre ve boyutlar arası etkileşime göre sınıflandırılmıştır (Hartig ve Höhler, 2009). Bu sınıflandırma şekil 1'deki gibi verilebilir.



Şekil 1. Çok Boyutlu Madde Tepki Kuramında Modeller

ÇBMTK'da bir test birden fazla yeteneği içerir. Ancak bu durum iki farklı yapı ile elde edilebilir. Birinci yapı basit yapı adını alır. Basit yapı bir test birden fazla alt testten oluşur. Her bir alt test tek

bir yeteneği veya özelliği ölçer ve kendi içinde homojendir. Her bir madde yer aldığı alt test ile ölçülmeye çalışılan tek bir özellik ile ilişkilidir. Basit yapılu testlerin ortaya koyduğu modele maddeler arası model adı verilir. Bu model genelde geniş ölçekli testlerde karşımıza çıkar. Örneğin TIMSS (Trends in International Mathematics and Science Study) matematik ve fen bilimleri alt testlerinden oluşmaktadır. Matematik testinde yer alan her bir alt test çeşitli konu ve bilişsel alanları içermektedir. Her bir konu alanı ham puanların rapor edildiği tek boyutlu bir alt test olarak ele almır. Bunun yanında testin tamamının matematik yapısını ölçtüğü varsayılır. Bu türdeki çok boyutlu testlerin ortaya koyduğu modele çoklu-tek boyutlu model (multi-unidimensional model) adı verilmektedir (Sheng ve Winkle, 2007; Bulut, 2013).

Çoklu-tek boyutlu model ÇBMTK'nın özel bir durumudur. Modelde madde parametrelerinden güçlük ve şans parametreleri her madde için bir tane kestirilirken; ayırt edicilik parametresi, her madde için testte yer alan boyut sayısı kadar elemana sahip bir vektör şeklinde kestirilir. Buna göre m alt testten oluşan bir testteki j maddesinin ayırt ediciliği $\alpha_j = (\alpha_{1j}, \alpha_{2j}, \dots, \alpha_{mj})$ vektörüne eşittir. Ancak her bir maddenin ayırt ediciliği ilişkili olmadığı alt testlerde sıfırdır. Bu sebeple vektör $\alpha_j = (0, 0, \dots, \alpha_{vj}, 0, \dots, 0)$ şeklinde sadeleştirilir. Üç alt testten oluşan basit yapılu 3PLM'deki bir testin madde parametrelerinin genel görünüşü tablo 1'deki gibidir.

Tablo 1. Basit Yapıda Parametreler

Alt test (i)	Madde (j)	α_{1j}	α_{2j}	α_{3j}	β_{ij}	γ_{ij}
1	1	α_{11}	0	0	β_{11}	γ_{11}
1	2	α_{12}	0	0	β_{12}	γ_{12}
1	3	α_{13}	0	0	β_{13}	γ_{13}
2	4	0	α_{24}	0	β_{24}	γ_{24}
2	5	0	α_{25}	0	β_{25}	γ_{25}
2	6	0	α_{26}	0	β_{26}	γ_{26}
3	7	0	0	α_{37}	β_{37}	γ_{37}
3	8	0	0	α_{38}	β_{38}	γ_{38}
3	9	0	0	α_{39}	β_{39}	γ_{39}

Çoklu-tek boyutlu 3 parametrelili lojistik modelde i bireyinin v alt testinde yer alan j maddesine doğru cevap verme olasılığı eşitlik 2'deki gibidir. Modelden şans parametresi çıkartılırsa çoklu-tek boyutlu 2PL modele ilişkin olasılık fonksiyonu elde edilir.

$$(U_{vij} = 1 | \theta_{vi}, \alpha_{vj}, \beta_{vj}, \gamma_{vj}) = \gamma_{vj} + \frac{(1 - \gamma_{vj})}{1 + e^{[(\sum_{l=1}^v \beta_{jl} \theta_{il}) + \alpha_{vj}]}} \quad (\text{Eşitlik 2})$$

U_{vij} = i bireyin v alt testindeki j maddesine verdiği cevap (1 veya 0)

θ_{vi} = i bireyinin v alt testine ilişkin yetenek düzeyi

β_{vj} = v alt testindeki j maddesinin güçlüğü

α_{vj} = v alt testindeki j maddesinin ayırtıcılığı

γ_{vj} = v alt testindeki j maddesinin en düşük asimptotu (şans parametresi)

ÇBMTK'da veri setinin yapısına dayalı olarak ele alınan ikinci yapı karmaşık yapıdır. Karmaşık yapı, diğer bir ifade ile basit olmayan yapı, maddeler içi model olarak da bilinir. Bu modelde maddeler tek bir yetenek ile ilişkili değildir. Hem testin bütünü hem de testteki maddeler birden çok

yetenek ile ilişkilidir. Bunun bir sonucu olarak maddeler ölçülmek istenen birden fazla yeteneğe faktör yükü verir (Sheng ve Wikle, 2007; Bulut, 2013).

Boyutlar arası etkileşime dayalı yapılacak sınıflandırmada öncelikli koşul, ele alınacak maddelerin maddeler içi modele sahip olmasıdır. Bir madde birden fazla özelliği ölçüyorsa, ölçtüğü özellikler arasındaki etkileşime dayalı olarak telafisel veya telafisel olmayan (kısmi telafisel) model başlığı altında ele alınabilir. Telafisel modelde; bir bireyin bir yetenekteki zayıflığı, maddenin ilişkili olduğu diğer boyutta sahip olduğu yüksek yetenek tarafından telafi edilir ve bu şekilde bireyin maddeyi doğru cevaplama olasılığı artar. Bu modelin matematiksel alt yapısında, θ -koordinatlarının doğrusal kombinasyonuna ogive veya lojistik fonksiyon eklemesi yapılır. Telafisel modele örnek olarak yabancı dil sınavında yer alan astronomi ile ilgili bir paragraf sorusu verilebilir. Böyle bir soruda hem astronomi bilgisi hem de yabancı dil bilgisi yoklanmaktadır. Ancak astronomi bilgisi yüksek olan bir birey yabancı dile ilişkin yeteneği düşük olsa da maddeyi doğru cevaplayabilir. Ya da bireyin astronomi bilgisi yeterli olmasa dahi sahip olduğu yüksek yabancı dil bilgisi ile maddeyi doğru cevaplayabilir. Burada maddenin ilişkili olduğu bu iki yetenek arasında -astronomi ve yabancı dil- telafi edicilik ilişkisi söz konusudur. Telafisel modele bir başka örnek olarak matematikte birden fazla çözüm yolu içeren ve hem cebirsel işlem, hem geometride alan hem de trigonometri ile ilgili bir madde verilebilir. Buna göre öğrenci yoklanan üç boyutta yüksek yeteneğe sahip olmasa dahi sadece bir ya da iki alanda yüksek yeteneğe sahip olarak maddeyi doğru cevaplayabilir.

Telafisel olmayan modelde bireyin maddeyi doğru cevaplayabilmesi için maddenin ilişkili olduğu tüm boyutlarda belli bir yetenek düzeyine sahip olması gereklidir. Buna göre bu modelde, maddeyi doğru cevaplamak için bir boyutta sahip olunması gereken minimum düzeyden yüksek yetenek, diğer boyutta sahip olunması gereken minimum düzeyden düşük yeteneği telafi edemez. Bir bireyin bu tipteki bir maddeyi doğru cevaplama olasılığı maddeyi cevaplamak için gerekli yetenek olasılıklarının çarpımıdır (Ackerman, 1996; Ackerman, Gierl ve Walker, 2003; Reckase, 2009). Telafisel olmayan modele örnek olarak kimya dersindeki bir laboratuvar uygulaması ele alınabilir. Burada öğrencinin hem kimya bilgisine sahip olması hem de deneyi nasıl yapacağını bilmesi gerekir. Öğrencinin kimya alanında bilgisi ne kadar fazla olursa olsun, başarılı olması için ilgili deneyin nasıl yapılacağına ilişkin olarak da belli bir bilgi düzeyine sahip olmalıdır. Yapılan az sayıda çalışmada modellerin kullanışlılığı, gerçek veri setlerini doğru biçimde ele alma ve temsil edebilme bakımından test edilmiştir (Bolt ve Lall, 2003; Babcock, 2009). Sonuçlar telafisel modelin daha iyi sonuçlar ürettiği ve bu bakımdan daha kullanışlı olduğunu göstermektedir (Reckase, 2009). Karmaşık model ortaya koyan telafisel özellikteki 3PLM bir testte j bireyinin i maddesine doğru cevap verme olasılığına ilişkin olasılık fonksiyonu aşağıdaki gibidir.

$$P(U_{ij} = 1 | \theta_j, \alpha_i, \gamma_i, d_i) = \gamma_i + (1 - \gamma_i) \frac{e^{\alpha_i \theta_j' + d_i}}{1 + e^{\alpha_i \theta_j' + d_i}} \quad (\text{Eşitlik 3})$$

U_{ij} = j bireyinin i maddesine verdiği cevap (1 veya 0)

θ_j = j bireyinin yetenek düzeyi

α_i = i maddesinin ayırcılığı

γ_i = i maddesinin en düşük asimptotu (şans parametresi)

d_i = i maddesinin kesim noktası

Denklemden yer alan d parametresi α ve β parametrelerinin etkileşimi ile elde edilen kesim noktasıdır. Buna göre tek boyutlu lojistik modellerde yer alan $(\alpha(\theta - \beta))$ çarpımı yapılırsa $(\alpha\theta - \alpha\beta)$ elde edilir. Burada yer alan $-\alpha\beta$ terimi d ile yer değiştirildiğinde $\alpha\theta + d$ ifadesi elde edilmiş olur. Tek boyutlu model çok boyutlu modele genişletildiğinde; m boyutlu bir test için α parametresi $1 \times m$

elemanlı ayırt edicilik vektörünü; θ parametresi $1 \times m$ elemanlı birey koordinatları vektörünü temsil eder. d parametresi $\alpha\theta + d$ denkleminin kesim noktasını verir. Modelde yer alan “e” teriminin üssünde yer alan ifade α ve θ parametrelerinin etkileştiğini göstermek üzere aşağıdaki şekilde genişletilebilir:

$$\alpha_i \theta_j' + d_i = \alpha_{i1} \theta_{j1} + \alpha_{i2} \theta_{j2} + \alpha_{i3} \theta_{j3} + \dots + \alpha_{im} \theta_{jm} + d_i = \sum_{i=1}^m \alpha_{i1} \theta_{j1} + d_i \quad (\text{Eşitlik 4})$$

(Reckase, 2009).

Yukarıda da bahsedildiği gibi MTK farklı test yapılarına özgü farklı modeller ve çıktılar sunan zengin bir kuramdır. Klasik test kuramı ile aynı alanlarda uygulanabilmesine rağmen çoğu zaman klasik test kuramından daha iyi sonuçlar sunmaktadır (Hambleton ve Jones, 1993; Harvey ve Hammer, 1999; Sinar ve Zickar, 2002; McDonald ve Paunonen, 2002). Temelleri 1980’lerde ortaya konan (Sympson, 1978; Doody-Bogan ve Hattie, 1981; Yen, 1983; akt: Ansley ve Forsyth, 1985) ÇBMTK ile ilgili Türkiye’de doktora düzeyinde tamamlanmış birkaç tez mevcuttur (Köse, 2010; Sünbül, 2011; Özkan, 2012, Koğar, 2014; Yavuz, 2014). Bu çalışmalardan Köse (2010), Sünbül (2011), Özkan (2012) ve Koğar’ın (2014) çalışmaları tek boyutlu MTK ve çok boyutlu MTK’nın telafisel modeldeki iki boyutlu testleri ile sınırlıdır. Yapılan en güncel çalışmalardan Yavuz’un (2014) tezinde boyut sayısı 3 ve 5 olarak değişen çok boyutlu testler ele alınmıştır. Ancak bu çalışmada da testin yapısı basit yapı ve iki-faktör modeli şeklinde sabitlenmiştir. Gerçek uygulamalarda ise boyut sayısı 2’nin üzerinde olan karmaşık yapı testlerinden de yararlanılmaktadır. Yapılan çalışmalarda bu durumu örnekleyecek herhangi bir desen ele alınmamıştır. Ancak bu tür çok boyutlu testler ile çalışırken doğru kestirimler yapabilmek için uygun örneklem büyüklüğü ve test uzunluğu koşullarının belirlenmesi gereklidir. Bu nedenle bu çalışma ile farklı boyutluluk (boyut sayısı 2 ve 3) ve farklı yapı özelliklerindeki (basit yapı ve karmaşık yapı) testlerin ÇBMTK çerçevesinde ele alınmasına, ele alınan yapılara ilişkin farklı örneklem büyüklüğü ve test uzunluğu sınamalarının yapılarak kurama dayalı kestirimlerin değişen koşullardan nasıl etkilendiğinin incelenmesine gerek olduğu görülmüştür. Bu doğrultuda çalışmada “basit ve karmaşık yapı iki ve üç boyutlu testlerden çok boyutlu madde tepki kuramıyla yapılan madde ve birey parametresi kestirimleri örneklem büyüklüğü ve test uzunluğundan nasıl etkilenmektedir?” problem cümlesine cevap aranmıştır.

Bu problem cümlesine dayalı olarak araştırmanın alt problem cümleleri şu şekilde ele alınmıştır:

1. İki boyutlu karmaşık ve basit yapı testler için;
 - 1.1. Örneklem büyüklüğü 1000, 2000 ve 4000; test uzunluğu 12 ve 48 olarak ele alındığında madde parametreleri kestirimlerine ilişkin RMSE, yanlılık, gerçek ve kestirilen parametreler arasındaki korelasyonlar nasıl değerler almaktadır?
 - 1.2. Örneklem büyüklüğü 1000, 2000 ve 4000; test uzunluğu 12 ve 48 olarak ele alındığında birey parametreleri kestirimlerine ilişkin RMSE, yanlılık, gerçek ve kestirilen parametreler arasındaki korelasyonlar nasıl değerler almaktadır?
2. Üç boyutlu karmaşık ve basit yapı testler için;
 - 2.1. Örneklem büyüklüğü 1000, 2000 ve 4000; test uzunluğu 12 ve 48 olarak ele alındığında madde parametreleri kestirimlerine ilişkin RMSE, yanlılık, gerçek ve kestirilen parametreler arasındaki korelasyonlar nasıl değerler almaktadır?
 - 2.2. Örneklem büyüklüğü 1000, 2000 ve 4000; test uzunluğu 12 ve 48 olarak ele alındığında birey parametreleri kestirimlerine ilişkin RMSE, yanlılık, gerçek ve kestirilen parametreler arasındaki korelasyonlar nasıl değerler almaktadır?

Araştırmanın Amacı

Çalışmanın amacı basit ve karmaşık yapıda olan iki ve üç boyutlu testlerin farklı örneklem büyüklüğü ve test uzunluğu koşullarında madde ve birey parametreleri kestirimlerine ilişkin RMSE ve yanlışlık durumlarını belirlemektir. Ayrıca çalışmada elde edilen parametre kestirimleri ile gerçek parametreler arasındaki ilişki de incelemiştir. Elde edilen sonuçlara bağlı olarak çalışılan test yapılarında minimum hataya ve yanlışlığa sahip kestirimler yapabilmek için gerekli örneklem büyüklüğü ve test uzunluğuna ilişkin önerilerin geliştirilmesi amaçlanmaktadır.

Çok boyutlu madde tepki kuramına ilişkin olarak yapılan çalışmalar genel olarak iki boyutlu testler üzerine yoğunlaşmıştır. Yapılan pek çok çalışmada basit ve karmaşık yapı ele alınmıştır. Fakat aynı çalışma içerisinde iki modeli birlikte ele alan çalışmalara pek rastlanmamaktadır. Türkiye’de yapılan ÇBMTK çalışmaları iki boyutlu testleri ve karmaşık yapıyı ele almakta; üç boyutlu yapıları ele alan ÇBMTK çalışması bulunmamaktadır. Bu doğrultuda basit ve karmaşık yapılu üç boyutlu testler ile ilgili bir çalışma da henüz mevcut değildir. Bu çalışma ÇBMTK kapsamında iki ve üç boyutlu testleri ele alması bakımından önem taşımaktadır. Ele alınan çok boyutlu testlerin karmaşık yapının yanında basit yapıyı da kapsamasının araştırmayı önemli kıldığı ve basit yapı olarak adlandırılan çoklu-tek boyutlu modelin tanıtılmasının, farklı boyut sayısı, örneklem büyüklüğü, test uzunluğu koşullarında incelenmesinin alan yazına katkı sağlayacağı düşünülmektedir. Türkiye’de yapılan çalışmalarda farklı kestirim yöntemleri için farklı yazılımlar kullanılmıştır, ancak IRTPRO yazılımı kullanılarak yapılmış bir çalışma bulunmamaktadır. Bu çalışmada IRTPRO 2.1 kullanıldığından, çalışmanın sonucunda yazılım ile ilgili olarak araştırmacılara yönelik öneri geliştirilmesinin, araştırmanın önemini arttırdığı düşünülmektedir.

YÖNTEM

Araştırmanın Türü

Bu çalışma, farklı örneklem büyüklüğü, test uzunluğu ve boyut sayısı koşulları altında çok boyutlu madde tepki kuramının farklı modellerinden yapılan parameter kestirimlerine ilişkin hata, yanlışlık ve yapılan kestirimler ile gerçek parametrelerin korelasyonlarını incelemesi, bu doğrultuda bir kuramın söz konusu modellerini test etme amacını taşımasından dolayı temel araştırma niteliğindedir.

Araştırmanın Verileri

Çalışma için gereken veriler simülasyon yoluyla elde edilmiştir. Yapılan bir çalışmada simülasyon verisi kullanmanın avantajı gerçek madde ve yetenek parametrelerinin bilinmesidir. Böylece araştırmacı gerçek değerler ile kestirilen değerler arasında sağlıklı karşılaştırmalar yapabilir. Ancak dikkat edilmesi gereken iki önemli koşul vardır. Bunlardan ilki araştırmacı veri setini oluşturmak için uygun bir yöntemi seçmelidir. İkinci koşul ise oluşturulan simülasyon veri setlerinin test maddelerine verilecek gerçek cevapları temsil ettiğinden emin olunması gerektiğidir (Way, Ansley ve Forstyh, 1988).

Yukarıda bahsedilen koşulları karşılamak amacıyla verilerin üretimi ÇBMTK’ya uygun veri üretmek üzere hazırlanmış SimuMIRT (2003) yazılımından faydalananarak yapılmış ve parametrelerin gerçekte var olan bir durumu yansıtacak şekilde olmasına özen gösterilmiştir. Bunun için parametre üretilmesinde çeşitli dağılım özellikleri göz önünde bulundurulmuş ve madde ve birey parametreleri ile iki kategorili veri setleri oluşturulmuştur.

Çalışmada çok boyutlu yapılar sabit olarak ele alınmıştır. Buna göre sabitler: iki boyutlu karmaşık, iki boyutlu basit, üç boyutlu karmaşık ve üç boyutlu basit yapı şeklindedir. Ackerman (1989) çalışmasında simülasyon yoluyla ürettiği verilerin thetaları arasındaki korelasyon arttıkça veri setinin tek boyutlu yapıya yaklaştığını raporlamıştır. Bu çalışmada da desenlerin tek boyutlu yapıya yaklaşmasının önüne geçmek için boyutlar arası korelasyon tüm koşullarda düşük korelasyonu göstermek üzere 0,3 olarak sabitlenmiştir. Test uzunluğu ve örneklem büyüklüğü araştırmanın bağımsız değişkenlerini oluşturmaktadır.

Test uzunluğu: Bağımsız değişkenlerden biri olan test uzunluğunun parametre kestirimi üzerine etkisini test etmek amacıyla kısa ve uzun testleri temsil edecek şekilde iki test uzunluğu koşulu ele alınmıştır. Bu doğrultuda madde sayısının ele alınan boyut sayılarının tam katı olması, basit yapılı testlerin her bir alt testinde en az 3 maddenin yer alması koşullarının sağlanması amacıyla kısa testleri temsil etmek üzere 12, uzun testleri temsil etmek üzere 48 maddeli testlerden yararlanılmıştır. Basit yapılı testlerin her bir alt testinde eşit sayıda madde yer almıştır.

Örneklem Büyüklüğü: Araştırmanın amacına yönelik olarak kestirimler üzerine örneklem büyüklüğünün etkisini test etmek ve araştırmanın sonucunda örneklem büyüklüğü önerisinde bulunmak amacıyla farklı örneklem büyüklükleri ile çalışılmıştır. Literatürde çok boyutlu yapılarda en az 1000 kişilik örneklem ile çalışılması önerilmektedir (Bolt ve Lall, 2003; Yao ve Boughton, 2007; Lee 2012). Bu sebeplerle çalışmada aynı dağılım özelliklerine sahip 1000, 2000 ve 4000 olmak üzere 3 örneklem büyüklüğü koşulu ele alınmıştır.

İşlem

Çalışmada iki boyutlu karmaşık, iki boyutlu basit, üç boyutlu karmaşık ve üç boyutlu basit yapıda, test uzunluğu 12 ve 48; örneklem büyüklüğü 1000, 2000, 4000 şeklinde değişen 4 ÇBMTK modeli * 2 test uzunluğu * 3 örneklem büyüklüğü koşulu olmak üzere 24 koşul içeren desen oluşturulmuştur. Araştırmanın deseni tablo 2’de verilmiştir. Desene ilişkin verilerin oluşturulması için kullanılan SimuMIRT’in (2003) basit ve karmaşık yapılı testler için farklı araçları mevcuttur. Veri üretimi önce madde parametreleri, sonra yanıtlar ve birey parametreleri olmak üzere iki aşamada gerçekleştirilmektedir. Madde parametrelerini üretmek için ele alınan koşuldaki test yapısı ile ilgili yazılım dosyasına madde sayısı, boyut sayısı, basit yapılı testler için her bir boyutta yer alan madde sayısı, madde güçlüğü, madde ayırt ediciliği için ortalama ve standart sapma değerleri girilmiştir. Ayırt edicilik parametrelerinin (a_i) üretilmesinde ortalaması 0,5 ve standart sapması 0,4 olan log-normal dağılım, güçlük parametrelerinin (d_i) üretilmesinde standart normal dağılım kullanılmıştır. Sonuçta ilgili koşula ilişkin madde parametreleri elde edilmiştir ve ikinci aşamaya geçilmiştir. Bu aşamada ilgili dosyaya ilk aşamada elde edilen madde parametreleri kopyalanmış; üzerine madde sayısı, birey sayısı, boyut sayısı, örneklem ortalaması, varyans kovaryans matrisinin alt üçgeni, cevap kategorilerinin sayısı ile birey parametreleri ve yanıtları oluşturmak üzere random seed değerleri girilmiştir. Sonuçta ilk adımda üretilen madde parametrelerine dayalı olarak iki kategorili yanıtlar ile birey parametreleri oluşturulmuştur. Yetenek parametreleri her bir birey için her bir boyuta ilişkin olarak elde edilmiştir ve oluşturulmasında standart normal dağılımdan faydalanılmıştır. Ayrıca boyutlar arası korelasyonun 0,3 olduğunu göstermek üzere varyans kovaryans matrisinin köşegen dışı elemanları 0,3 olarak girilmiştir.

Dağılım özellikleri sabit kalmak üzere boyut sayısı ve test yapısına göre veri üretiminde farklılıklar mevcuttur. Buna göre iki ve üç boyutlu basit yapılı ÇBMTK modellerinde her bir madde sadece bir boyut ile ilişkili olduğu için madde sayısı kadar ayırt edicilik ve güçlük parametresi üretilmiştir (bkz. Tablo 1). Üretilen her bir ayırt edicilik parametresi bir madde ve bir boyut ile ilişkilendirilmiş, ilişkilendirildiği maddenin diğer boyutlardaki ayırt ediciliği sıfır olarak alınmıştır. İki ve üç boyutlu karmaşık yapılı ÇBMTK modellerinde veri üretimi telafisel modele göre yapılmıştır. Bu modelde her bir madde tüm boyutlar ile ilişkili olduğundan her bir madde için boyut sayısı kadar ayırt edicilik parametresi ile bir tane güçlük parametresi üretilmiştir. Böylece karmaşık yapılarda boyut sayısı * madde sayısı kadar ayırt edicilik parametresi ve madde sayısı kadar güçlük parametresi elde edilmiştir.

Her bir koşula ilişkin olarak yapılan bu işlem madde ve birey parametreleri sabit kalmak üzere yanıtlar için girilen random seed değiştirilerek 25 kere tekrar (replikasyon) edilmiştir. Bu tip simülasyon çalışmalarında tekrar örneklem büyüklüğü ile eş anlamlıdır. MTK çalışmalarında tekrar sayısı kestirilen parametrelerin örnekleme hatasının varyansını ve simülasyon sonuçlarına dayalı istatistiklerin yeterli güce sahip olmasını etkilemektedir (Harwell vd. 1996). Bu çalışmada da alan yazın önerileri ve analizlerin uzun sürmesi gibi sınırlılıklar göz önünde bulundurularak bir koşul 25

tekrar ile sınırlandırılmıştır. Sonuçta 24 koşul * 25 tekrar olmak üzere 600 veri seti elde edilmiştir. Verilerin üretilmesinden sonra analize geçilmiştir.

Tablo 1: Araştırmanın Deseni

Sabitler	Değişkenler	
Yapı	Örneklem Büyüklüğü	Test Uzunluğu
2 Boyutlu Karmaşık Yapı	1000	12 madde
		48 madde
	2000	12 madde
		48 madde
	4000	12 madde
		48 madde
2 Boyutlu Basit Yapı	1000	12 madde
		48 madde
	2000	12 madde
		48 madde
	4000	12 madde
		48 madde
3 Boyutlu Karmaşık Yapı	1000	12 madde
		48 madde
	2000	12 madde
		48 madde
	4000	12 madde
		48 madde
3 Boyutlu Basit Yapı	1000	12 madde
		48 madde
	2000	12 madde
		48 madde
	4000	12 madde
		48 madde

Verilerin Analizi

Üretilen veri setlerinin madde ve birey parametreleri test yapısına uygun betikle IRTPRO 2.1 yazılımından faydalanarak kestirilmiştir. IRTPRO 2.1 yazılımında kestirimler Bock-Aitkin Expectation-Maximization (BA-EM), Adaptive Quadrature ve Metropolis-Hastings Robbins-Monro (MH-RM) algoritmaları ile yapılabilmektedir. Ancak MH-RM algoritmasının kullanımı boyut sayısının ikiyi veya üçü aştığı durumlarda önerilmekte, Adaptive Quadrature algoritmasıyla yapılan birey parametresi kestirimlerinde her bir birey için quadrature nod sayısı değişebilmektedir (IRTPRO guide, 2011). Çalışmada ele alınan veri setlerinde boyut sayısının üçü geçmemesi ve quadrature nod sayılarının sabit tutularak birey parametrelerinin kestirilmek istenmesi sebebiyle, analizler BA-EM algoritmasından faydalanarak yapılmıştır. Analizlerin seri biçimde yapılabilmesi için yazılım batch modunda çalıştırılmıştır. Testlerin analizlerinde analiz için gereken süre özellikle karmaşık yapılarda madde ve boyut sayısının artması ile artmış, 3 boyutlu karmaşık yapılarda en yüksek değerini almıştır. Buna göre IRTPRO 2.1, 3 boyutlu karmaşık yapının 48 madde ve 4000 örneklem büyüklüğü koşulundaki bir veri setinin madde ve birey parametrelerini kestirmek için Intel Core i7-4500 CPU 1.80GHz 2.40 GHZ 8GB (RAM) 64 bit işletim sistemi olan bir bilgisayarda ortalama 22 saat çalıştırılmıştır.

Analizlerin tamamlanmasından sonra madde ve birey parametrelerinin kestirilen ve gerçek değerlerinden faydalanarak GOR istatistikleri (kestirim iyiliği - goodness of recovery) hesaplanmıştır. GOR istatistikleri farklı koşulların etkisini araştırmak ve kestirim kararlılığını belirlemek amacıyla parametreler için hesaplanan istatistiklerdir (Maris, 1999; Turhan, 2006). Bu

çalışmada her bir parametre için RMSE ve yanlılık olmak üzere iki GOR istatistiği hesaplanmıştır. RMSE gerçek ve kestirilen parametre değerleri arasındaki farkın kareleri ortalamasının kareköküdür. Bu istatistik aşağıda yer alan formülden faydalanarak hesaplanmıştır.

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (\bar{\tau}_{nj} - \tau_{nj})^2}{N}} \quad (\text{Eşitlik 5})$$

RMSE formülündeki terimlerden N terimi testte yer alan parametre sayısını temsil etmektedir. Parametre sayısı, RMSE'nin madde parametreleri için hesaplandığı durumda testteki madde sayısı; birey parametreleri için hesaplandığı durumda testteki birey sayısı olarak alınmıştır. τ_{nj} n. maddeye/bireye ait j parametresinin gerçek değeri iken, $\bar{\tau}_{nj}$ n. maddenin/bireyin j parametresinin kestirilen değeridir. Her bir koşulda her bir tekrar için parametrelerin RMSE değeri hesaplanmıştır. Daha sonra koşulda yer alan 25 tekrardan elde edilen RMSE değerlerinin ortalaması, ilgili parametrenin RMSE değeri şeklinde raporlanmıştır. RMSE minimum sıfır olmak üzere her zaman pozitif değer almaktadır. İlgili koşulda elde edilen RMSE değerlerinin sıfıra yakın olması kestirim kararlılığının yüksek, sıfırdan uzak olması kestirim kararlılığının düşük olduğu şeklinde yorumlanmaktadır.

Yanlılık, parametrenin kestirilen ortalama değeri ile gerçek değeri arasındaki farktır. Çalışmada yanlılık değerleri her bir madde ve her bir birey parametresi için hesaplanmıştır. Bunun için aşağıda yer alan eşitlikten yararlanılmıştır.

$$Yanlılık(\tau_j) = \left(\frac{\sum_{r=1}^R \bar{\tau}_{jr}}{R} \right) - \tau_j \quad (\text{Eşitlik 6})$$

Eşitlik 6'da yer alan R tekrar sayısını temsil etmektedir. $\bar{\tau}_{jr}$ j parametresinin r. tekrardaki kestirilen değeri, τ_j j parametresinin gerçek değeridir. Formüle göre her bir parametrenin her tekrarda kestirilen değerlerinin ortalamasının gerçek parametreden farkı bulunmuştur. Böylece madde parametreleri için madde sayısı kadar, birey parametreleri için birey parametresi kadar yanlılık değeri elde edilmiştir. Elde edilen yanlılıkların madde/birey sayısına bölünerek hesaplanan değeri ilgili koşuldaki parametrenin yanlılığı olarak raporlanmıştır. Yanlılık hem pozitif hem de negatif değerler alabilmektedir. Sıfır ve sıfıra yakın yanlılık değerleri parametre kestiriminin yansız yapıldığı şeklinde yorumlanmaktadır.

Araştırmada gerçek parametreler ile kestirilen parametreler arasındaki ilişkiyi görmek için korelasyonlar incelenmiştir. Bunun için her bir tekrardan kestirilen parametreler ile gerçek parametreler arasındaki Pearson Momentler Çarpımı Korelasyon katsayısı hesaplanmıştır. Desende yer alan 25 tekrar için bu işlem yapılmış ve ortalaması hesaplanarak rapor edilmiştir. Korelasyon katsayılarının hesaplanmasına dair izlenen bu süreç hem madde hem birey parametreleri için aynı şekildedir.

BULGULAR

Bu başlıkta araştırmanın bulguları alt problemler halinde verilmiştir:

1. Birinci alt problem için bulgular

1.1. İki boyutlu karmaşık ve basit yapıları testlerden kestirilen madde parametrelerine ilişkin bulgular:

İki boyutlu karmaşık ve basit yapıları testlerden kestirilen madde parametrelerine ilişkin bulgular tablo 3'teki gibidir:

İki boyutlu karmaşık yapıları testlerden yapılan madde parametresi kestirim sonuçlarına göre RMSE'ler 0,048 ile 0,323 arasında değişen değerlere sahiptir. Basit yapıları testlerde madde parametrelerine ait RMSE'ler 0,067 ile 0,174 arasında değişen ve genel itibarı ile aynı koşullarda karmaşık yapıda ortaya çıkan değerlerden daha düşük değerler almıştır. Özellikle karmaşık yapıları testlerin kısa test koşulunda madde parametrelerinden elde edilen RMSE değerleri, uzun test koşulundaki RMSE değerlerinden belirgin biçimde yüksektir. Her iki yapıda da tüm parametreler örneklem büyüklüğünün artması ile daha düşük hatalı kestirilmiştir. Ayrıca d parametresi, her iki test yapısında da test uzunluğu ve örneklem büyüklüğüne ilişkin koşulların çoğunda a parametrelerinden daha düşük RMSE değerleri ile kestirilmiş; d parametresinin a1 ve a2 parametresine göre daha kararlı davrandığı gözlenmiştir.

Tablo 3: İki Boyutlu Testler İçin Madde Parametrelerine Ait Bulgular

Karmaşık Yapı										
Madde sayısı	Örneklem	RMSE			Yanlılık			Korelasyon		
		a1	a2	d	a1	a2	d	a1	a2	D
12 madde	1000	0,317	0,323	0,106	0,124	-0,151	-0,008	0,805	0,905	0,995
	2000	0,226	0,275	0,065	0,124	-0,169	0,006	0,891	0,945	0,998
	4000	0,212	0,262	0,048	0,140	-0,197	-0,016	0,917	0,966	0,999
48 madde	1000	0,135	0,139	0,090	0,000	-0,026	-0,020	0,946	0,946	0,997
	2000	0,093	0,096	0,061	-0,010	-0,020	-0,006	0,972	0,983	0,999
	4000	0,068	0,071	0,049	-0,013	-0,020	-0,019	0,985	0,991	0,999
Basit Yapı										
Madde sayısı	Örneklem	RMSE			Yanlılık			Korelasyon		
		a1	a2	d	a1	a2	d	a1	a2	d
12 madde	1000	0,171	0,174	0,120	-0,012	-0,008	0,013	0,959	0,967	0,983
	2000	0,132	0,131	0,094	-0,013	-0,003	0,017	0,969	0,976	0,985
	4000	0,097	0,102	0,076	-0,009	-0,021	0,013	0,978	0,981	0,987
48 madde	1000	0,122	0,114	0,132	-0,015	0,003	-0,003	0,974	0,979	0,956
	2000	0,092	0,084	0,111	-0,023	-0,003	0,003	0,981	0,985	0,957
	4000	0,073	0,067	0,096	-0,012	-0,001	-0,007	0,982	0,988	0,958

İki boyutlu karmaşık yapıları testlerde madde parametrelerine ait yanlılık değerleri -0,197 ile 0,140 arasında değişmektedir. İki boyutlu basit yapıları testlerde yanlılığın -0,023 ile 0,017 arasında değişen değerler aldığı görülmektedir. Aynı koşullar altında basit ve karmaşık yapıdan hesaplanan yanlılık değerleri karşılaştırıldığında basit yapının daha yansız kestirim sonuçlarına sahip olduğu görülmektedir. Yanlılık sonuçlarına göre örneklem büyüklüğü ve madde sayısındaki artış ile hem a parametrelerinin hem de d parametresinin yanlılığında düzenli bir azalış ya da artışın olmadığı ifade edilebilir.

Kestirilen madde parametreleri ile gerçek parametrelerin korelasyonlarının ortalaması iki boyutlu karmaşık yapıları testlerin tüm koşullarında 0,805 ve üzerinde korelasyon değerlerine sahiptir. İki boyutlu basit yapıları testlerde korelasyonlar 0,956 ile 0,988 arasında değişen mükemmel değerlerdir. RMSE ve yanlılık bakımından en düşük değerlere sahip olan d parametresi her iki test yapısında da tüm koşullarda gerçek parametrelerle 0,956 ve üzerinde pozitif yönde, mükemmel düzeyde ilişki göstermiştir. a parametrelerine ait korelasyonlar her iki test yapısında da örneklem büyüklüğünün artması ile daha yüksek değere ulaşmıştır.

1.2. İki boyutlu karmaşık ve basit yapıli testlerden kestirilen birey parametrelerine ilişkin bulgular:

İki boyutlu karmaşık ve basit yapıli testlerden kestirilen birey parametrelerine ilişkin bulgular tablo 4'teki gibidir:

İki boyutlu karmaşık yapıli testlerden kestirilen yetenek parametrelerine ilişkin RMSE değeri incelenğinde değeri kısa testte 0,715 ile 0,621; uzun testte 0,540 ile 0,489 arasında değıştiği görülmektedir. İki boyutlu basit yapıli testlerin birey parametrelerine ilişkin RMSE değeri madde sayısının artması ile hem teta1 hem teta2 parametresi için 0,7 değeri etrafında iken 0,5 değeri etrafına gerilemiştir. Tablonun geneli incelendiğinde yetenek parametrelerine ilişkin RMSE değeriörlerinin örneklem büyüklüğünün artması ile ciddi bir değışime uğramadığı görülmektedir. Ancak her iki test yapısında da kısa testten uzun teste geçilmesi ile yetenek parametrelerinin RMSE değeriörlerinde düşme gözlenmiştir. Bu durum, birey parametrelerinin test uzunluğundan etkilendiği; örneklem büyüklüğünden etkileneceği şeklinde yorumlanabilir.

Tablo 2: İki Boyutlu Testler İçin Birey Parametrelerine Ait Bulgular

Karmaşık Yapı							
Madde Sayısı	Örneklem	RMSE		Yanlılık		Korelasyon	
		01	02	01	02	01	02
12 madde	1000	0,715	0,634	-0,016	0,040	0,686	0,774
	2000	0,704	0,621	-0,210	0,022	0,693	0,786
	4000	0,708	0,623	-0,008	0,030	0,696	0,788
48 madde	1000	0,540	0,496	-0,005	0,039	0,835	0,876
	2000	0,531	0,489	-0,012	0,022	0,838	0,880
	4000	0,533	0,491	0,005	0,030	0,840	0,880
Basit Yapı							
Madde Sayısı	Örneklem	RMSE		Yanlılık		Korelasyon	
		01	02	01	02	01	02
12 madde	1000	0,761	0,749	0,027	0,072	0,631	0,668
	2000	0,756	0,741	0,022	0,054	0,630	0,676
	4000	0,761	0,744	0,040	0,061	0,635	0,677
48 madde	1000	0,499	0,526	0,022	0,043	0,865	0,865
	2000	0,492	0,523	0,015	0,025	0,863	0,862
	4000	0,495	0,524	0,033	0,033	0,865	0,864

Karmaşık yapıli iki boyutlu testlerden hesaplanan yanlılık sonuçlarına göre birey parametreleri, -0,210 ile 0,040 değeri arasında değışen bir yanlılıkla kestirilmiştir. Basit yapıli iki boyutlu testlerden hesaplanan yanlılık değeri ele alınan tüm koşullarda pozitif ve 0,015 ile 0,072 arasında değışen değeri almaktadır. Yine basit yapıda tüm koşullarda birey sayısı arttıkça yanlılık değeri önce azalmış sonra artmıştır. Karmaşık yapıli testte teta2 için aynı durum söz konusu iken teta1 için tam tersi gerçekleşmiştir. Yanlılık değeri genel olarak incelendiğinde değeriörlerinin örneklem büyüklüğü ve test uzunluğundaki artış ile aynı yönde olmayan bir örüntüye sahip olduğu söylenebilir.

Her bir tekrarda kestirilen birey parametreleri ile gerçek birey parametrelerinin korelasyonlarının ortalamalarını veren tablo 4'te görüldüğü gibi yetenek parametrelerine ait korelasyon değeri karmaşık yapıda 0,686 ile 0,880 arasında; basit yapıda 0,630 ile 0,865 arasında değışmektedir. Her iki yapıda da genel olarak testteki madde sayısı arttıkça teta parametrelerine ait korelasyon değeri artmıştır. Hem basit hem karmaşık yapıda RMSE'si daha düşük değeri alan teta parametrelerinin aynı koşullarda gerçek parametrelerle gösterdiği ilişki daha yüksek değeri almaktadır. Madde sayısı sabit kalmak üzere örneklem büyüklüğü arttırıldığında korelasyon değeri ciddi bir artış meydana

gelmemiştir. Bu bulgu RMSE değerleri ile birlikte ele alındığında birey parametrelerinin kestiriminin örneklem büyüklüğünden etkilenmediği bulgusunu destekler niteliktedir.

2. İkinci alt problem için bulgular

2.1. Üç boyutlu karmaşık ve basit yapıları testlerden kestirilen madde parametrelerine ilişkin bulgular:

Üç boyutlu karmaşık ve basit yapıları testlerden kestirilen madde parametrelerine ilişkin bulgular tablo 5'te verilmiştir. Tablo 5'e göre üç boyutlu karmaşık yapıları testlerden kestirilen madde parametrelerine ilişkin RMSE değerleri, iki boyutlu desenlerde elde edilen RMSE değerlerinden farklılık göstermekte ve karmaşık yapıda hesaplanan RMSE'ler oldukça yüksek değerler almaktadır. RMSE'ler 0,054 ile 1,508 arasında değişen geniş bir aralığa sahiptir. En yüksek hata ile kestirilen parametre 12 maddeli test ve 1000 örneklem büyüklüğü koşulunda kestirilen a3 parametresi olmuştur. Üç boyutlu karmaşık yapılarında elde edilen yüksek RMSE değerlerine karşılık üç boyutlu basit yapılarında RMSE değerleri 0,241 ile 0,042 arasında değerler almaktadır. Üç boyutlu basit ve karmaşık yapılarında kısa testten uzun teste geçildiğinde madde parametrelerine ait RMSE değerlerinde düşme gözlenmiştir. Hem basit hem de karmaşık yapının uzun test koşulunda a1, a2 ve a3 parametrelerinin RMSE değerleri birbirine çok yakındır. İki test uzunluğu koşulunda da her iki yapıda en düşük RMSE değerlerine sahip parametre d parametresi olmuştur. Bu bulgu iki boyutlu desenlerde elde edilen bulgulara paralellik göstermektedir. Bu sonuca dayalı olarak üç boyutlu basit yapılarında da en kararlı kestirilen madde parametresinin d parametresi olduğu yorumuna gidilebilir.

Tablo 3: Üç Boyutlu Testler İçin Madde Parametrelerine Ait Bulgular

Karmaşık Yapı													
Madde sayısı	Örneklem	RMSE				Yanlılık				Korelasyon			
		a1	a2	a3	d	a1	a2	a3	d	a1	a2	a3	d
12 madde	1000	0,446	1,289	1,508	0,602	0,002	0,210	0,247	-0,174	0,836	0,358	0,435	0,899
	2000	0,474	1,350	1,290	0,530	0,172	0,193	0,077	-0,152	0,888	0,336	0,601	0,912
	4000	0,371	0,624	0,640	0,181	0,173	0,051	-0,163	-0,042	0,903	0,562	0,615	0,977
48 madde	1000	0,298	0,325	0,279	0,118	-0,011	-0,150	0,092	-0,042	0,857	0,884	0,889	0,996
	2000	0,273	0,294	0,254	0,078	-0,039	-0,159	0,110	-0,015	0,880	0,911	0,915	0,998
	4000	0,274	0,273	0,245	0,054	-0,008	-0,149	-0,113	-0,014	0,872	0,928	0,913	0,999
Basit Yapı													
Madde sayısı	Örneklem	RMSE				Yanlılık				Korelasyon			
		a1	a2	a3	d	a1	a2	a3	d	a1	a2	a3	d
12 madde	1000	0,241	0,206	0,186	0,103	0,008	-0,037	0,025	-0,016	0,955	0,942	0,980	0,993
	2000	0,137	0,134	0,132	0,072	-0,014	-0,043	0,012	-0,007	0,984	0,972	0,990	0,996
	4000	0,119	0,092	0,089	0,049	0,004	-0,024	-0,019	0,002	0,987	0,986	0,995	0,998
48 madde	1000	0,114	0,110	0,109	0,082	0,004	-0,038	-0,008	-0,015	0,990	0,987	0,987	0,996
	2000	0,075	0,078	0,086	0,060	-0,001	-0,031	-0,007	-0,004	0,996	0,994	0,992	0,998
	4000	0,058	0,053	0,061	0,042	0,005	-0,011	-0,019	-0,005	0,997	0,997	0,996	0,999

Üç boyutlu karmaşık yapıları veri setlerinde yapılan en yanlı kestirim testin kısa ve örneklem büyüklüğünün 1000 olduğu koşulda a3 parametresine (0,247); en yansız kestirim ise testin kısa ve örneklem büyüklüğünün 1000 olduğu koşulda a1 parametresine (0,002) aittir. Diğer taraftan üç boyutlu basit yapıları veri setlerinden kestirilen madde parametrelerinin yanlılıkları 0,043 ile 0,025 arasında değişen, birbirine ve sıfıra yakın değerler almıştır. Basit yapının tüm koşulları göz önünde bulundurulduğunda en yanlı kestirilen parametre, kısa test ve 2000 örneklem büyüklüğü koşulunda a2 parametresi iken, en yansız kestirilen parametre kısa test ve 4000 örneklem koşulunda kestirilen d parametresi olmuştur. Üç boyutlu veri setlerinden kestirilen madde parametrelerinin yanlılıkları

daha önce incelenen iki boyutlu yapılara benzer biçimde madde sayısının ve örneklem büyüklüğünün artması ile düzenli bir artma ya da azalma göstermemiştir.

Tablo 5'te yer alan kestirilen madde parametreleri ile gerçek madde parametreleri arasındaki korelasyon değerlerine göre üç boyutlu karmaşık yapı testler için korelasyon ortalamaları 0,336-0,999 arasında; basit yapı testlerde ise korelasyonlar 0,942 ile 0,999 arasında değişen mükemmel değerlere sahiptir. Bu sonuçlar RMSE sonuçları ile paralellik göstermektedir. RMSE'nin 1'den büyük olarak hesaplandığı koşullarda korelasyonlar orta düzeyde çıkmıştır. Örneklem büyüklüğü ve test uzunluğu arttıkça her bir parametre için elde edilen korelasyon değerleri de artmıştır. d parametresi hem basit hem de karmaşık yapıdaki her iki test uzunluğunda ve tüm örneklem büyüklüklerinde 0,899 ve üzerindeki değerlerle gerçek parametrelerle mükemmel düzeyde ilişkiye sahiptir. Bu bulgu d parametresinin daha kararlı kestirildiği bulgusuna paraleldir.

2.2. Üç boyutlu karmaşık ve basit yapı testlerden kestirilen birey parametrelerine ilişkin bulgular:

Üç boyutlu karmaşık ve basit yapı testlerden kestirilen birey parametrelerine ilişkin bulgular tablo 6'daki gibidir. Üç boyutlu testlerden kestirilen birey parametrelerine ilişkin RMSE değerleri karmaşık yapıda 0,542 ile 0,848 arasında; basit yapıda 0,553 ile 0,838 arasında değişmektedir. Her iki yapıdan aynı koşullarda kestirilen RMSE değerleri birbirine yakın değerler almakla beraber; uzun testte boyut başına düşen madde sayısının artması ile basit yapıda genel olarak karmaşık yapıdan daha düşük RMSE değerleri görülmektedir. Örneklem büyüklüğünün artması parametrelerin RMSE değerlerini ciddi bir biçimde etkilememiş, test uzunluğu sabit tutulmak üzere örneklem büyüklüğü arttırıldığında bazı yetenek parametrelerinin RMSE'lerinde düşüş, bazılarında artış gerçekleşmiştir. Testte yer alan madde sayısı arttığında her iki test yapısında da elde edilen RMSE'lerde fark edilir bir düşüş olmuştur.

Tablo 4 : Üç Boyutlu Testler İçin Madde Parametrelerine Ait Bulgular

Karmaşık Yapı										
Madde Sayısı	Örneklem	RMSE			Yanlılık			Korelasyon		
		01	02	03	01	02	03	01	02	03
12 madde	1000	0,721	0,730	0,848	0,043	0,038	0,067	0,695	0,626	0,526
	2000	0,753	0,753	0,838	0,007	0,034	0,070	0,718	0,603	0,531
	4000	0,686	0,753	0,819	0,010	0,036	0,046	0,731	0,649	0,556
48 madde	1000	0,619	0,542	0,583	0,009	0,006	0,046	0,785	0,821	0,823
	2000	0,608	0,546	0,575	-0,029	-0,001	0,044	0,793	0,825	0,830
	4000	0,605	0,558	0,576	-0,021	0,003	0,029	0,797	0,833	0,823
Basit Yapı										
Madde Sayısı	Örneklem	RMSE			Yanlılık			Korelasyon		
		01	02	03	01	02	03	01	02	03
12 madde	1000	0,827	0,832	0,768	0,044	0,045	0,085	0,542	0,480	0,650
	2000	0,817	0,838	0,763	0,015	0,037	0,084	0,574	0,493	0,655
	4000	0,819	0,819	0,752	0,024	0,042	0,068	0,578	0,515	0,655
48 madde	1000	0,567	0,580	0,612	0,054	0,018	0,064	0,828	0,803	0,803
	2000	0,553	0,587	0,611	0,016	0,010	0,062	0,833	0,804	0,805
	4000	0,554	0,592	0,605	0,024	0,014	0,047	0,835	0,815	0,802

Üç boyutlu testlerden elde edilen birey parametrelerine ilişkin yanlılık değerleri karmaşık yapıda -0,029 ile 0,070 arasında; basit yapıda 0,085 ile 0,010 arasında değişen değerler almıştır. Basit yapıda ele alınan tüm koşullarda; karmaşık yapıda ise neredeyse tüm koşullarda birey parametresi kestirimleri pozitif yanlılık ile yapılmıştır. Her iki test yapısında da birey parametrelerine ait yanlılık, örneklem büyüklüğü ve madde sayısından ciddi biçimde etkilenmemiştir. Ele alınan koşullarda

hesaplanan yanlılık değerlerinin yakın olduğu görülmektedir. Buna göre yetenek parametrelerinin tüm koşullarda benzer yanlılıkla kestirildiği sonucuna ulaşılabilir.

Üç boyutlu yapılardan birey parametreleri için elde edilen korelasyon ortalamaları genel itibari ile RMSE değerleri ile paralel bulgulara sahiptir. Buna göre RMSE'nin yüksek olduğu kısa test koşulunda madde parametreleri için korelasyonlar düşük; RMSE'nin daha düşük değerler aldığı uzun test koşulunda korelasyonlar daha yüksek değerler almıştır. Korelasyon ortalamaları karmaşık yapıda 0,526 ile 0,833; basit yapıda 0,480 ile 0,835 arasında değişmektedir. Testlerde yer alan madde sayısının artması korelasyon değerlerinin artmasını sağlamıştır. Genel olarak 12 maddeli testlerde örneklem büyüklüğünün artışı, 48 maddeli testlerde örneklem büyüklüğü artışından daha etkili olmuştur.

SONUÇLAR ve TARTIŞMA

Bu çalışmada iki ve üç boyutlu yapıların madde ve birey parametresi kestirimlerinin farklı örneklem büyüklüğü ve test uzunluklarından nasıl etkilendiği RMSE, yanlılık ve korelasyonlar ölçüt alınarak incelenmiştir. Örneklem büyüklüğünün ve test uzunluğunun parametre kestirimi üzerinde çok etkili olduğu bilinmektedir. Bu etkilerin niteliği ve niceliği pek çok çalışmanın konusu olmuştur. (Stone, 1992; Sinar ve Zickar, 2002; Bolt ve Lall, 2003; De Mars, 2003; de la Torre ve Patz, 2005; Sheng ve Wikle, 2007; Babcock, 2009; Finch, 2010; Finch, 2011; Sünbül, 2011; Lee, 2012; Kieftenbeld and Natesan, 2012; Zhang, 2012). Bu çalışmanın bulgularına ait sonuç ve öneriler madde ve birey parametrelerine ilişkin olmak üzere iki başlık altında verilmiştir.

1. Madde Parametrelerine İlişkin Tartışma ve Yorum

Çalışmadan elde edilen bulgulara göre örneklem büyüklüğünün artması hem iki hem de üç boyutlu testlerde madde parametrelerine ilişkin RMSE değerlerinin düşmesine sebep olmuştur. Bu bulgu Bolt ve Lall (2003), Finch (2010) ve Zhang'ın (2012) elde ettiği sonuçlar ile paralellik göstermektedir. Zhang (2012) yaptığı çalışma ile boyutlar arası korelasyon ve madde sayısı sabit iken birey sayısı arttıkça parametrelerin RMSE değerlerinin azaldığını ifade etmiş ve birey sayısının artmasının çok boyutlu yapıların kestirimleri için iyi sonuçlar oluşturduğunu ifade etmiştir. Finch (2010) ise madde ve örneklem büyüklüğünün yüksek olduğu durumda düşük hatalı ve kararlı kestirimler yapılabileceğini belirtmiş ve çalışmasında madde ayıricılığına ait RMSE değerlerinin ve güçlük parametresine ilişkin standart hataların birey ve madde sayısı arttıkça düştüğünü, ancak yanlılığın büyük örneklemelerde daha yüksek değerlere sahip olduğunu rapor etmiştir. Bu sebeple araştırmacı, çalışmasında, çok sayıda birey ve fazla sayıda madde kullanmanın her zaman doğru bir ÇBMTK parametre kestirimini sağlamayacağını ifade etmiştir. Bu çalışmada elde edilen bulgular Finch'in (2010) çalışmasından elde edilen bulgulara paralellik göstermektedir. Çalışmadan elde edilen sonuçlara göre madde parametrelerine ilişkin yanlılık değerleri örneklem büyüklüğünün ve/veya madde sayısının artmasına bağlı olarak düzenli bir artış veya azalış sergilememiştir. Örneğin 3 boyutlu basit yapıya kısa testten kestirilen a1 parametresinin yanlılık değeri örneklem büyüklüğü 1000 iken 0,008; 2000 iken -0,014 ve 4000 iken 0,004 olarak hesaplanmıştır. Buna göre bu parametrenin kestirim yanlılığı örneklem büyüklüğünün ilk kez artışı (1000'den 2000'e) ile artmış, ikinci kez artışı (2000'den 4000'e) ile azalmıştır. Benzer durum farklı koşullarda diğer parametreler için de geçerlidir. Bu çalışmada ele alınan yapılara göre madde parametrelerinin hesaplanan mutlak yanlılık değerlerinin alt ve üst sınırları incelenirse, boyut sayısı arttıkça ve yapı basitten karmaşığa geçtikçe yanlılık değerlerinin arttığı görülmektedir. Bu değerlere göre yanlılık değerlerinin artışı boyut sayısı ve yapının karmaşık olup olmaması ile ilişkilidir. Bu bulgu Lee'nin (2012) ifadeleriyle paralellik göstermektedir. Lee (2012) çalışmasında ÇBMTK modelinin parametre kestirimini etkileyen faktörlerin artmasının kestirim yanlılığının artışına sebep olduğunu belirtmiştir.

Çalışmadan elde edilen bir diğer bulgu da elde edilen yanlılık ve RMSE değerlerinin aynı boyut sayısı, test uzunluğu ve örneklem büyüklüğü koşullarındaki basit yapı ve karmaşık yapı testlerde farklılık göstermesidir. Buna göre iki boyutlu basit yapılarda a1 ve a2 parametreleri için elde edilen

yanlılık ve RMSE değerleri tüm test uzunluğu ve örneklem büyüklüğü koşullarında karmaşık yapıdan daha düşük değerlerdedir.

İki boyutlu karmaşık yapıli testlerde d parametresi daha kararlı kestirilmiştir ve bu parametreye ilişkin RMSE değerleri tüm koşullarda, yanlılık değerleri ise iki koşul (uzun testteki 1000 ve 4000 örneklem koşulları) dışındaki tüm koşullarda basit yapıdan daha düşük değerlere sahiptir. Bu bulguları doğrular biçimde madde parametrelerine ilişkin elde edilen korelasyon ortalamaları da a1 ve a2 parametreleri için basit yapıli testte, d parametresi için karmaşık yapıli testte daha yüksek değerlerde olma eğilimindedir. Üç boyutlu yapılarda tüm örneklem büyüklüğü ve test uzunluğu koşullarında basit yapıdan elde edilen RMSE ve yanlılık değerleri, karmaşık yapıdan elde edilen RMSE ve yanlılık değerlerinden daha düşüktür. Bu durum sadece kısa test ve 1000 örneklem büyüklüğü koşulunda a1 parametresinin yanlılık değeri için istisna teşkil etmektedir. Üç boyutlu yapılardan elde edilen madde parametrelerinin korelasyonlarının ortalaması ise yukarıdaki bulguyu destekler biçimde basit yapılarda daha yüksek değerlere sahiptir. Uzun test koşulunda kestirilen d parametrelerinin gerçek d parametreleri ile korelasyonlarının ortalaması hem basit hem karmaşık yapıda bire bir aynı değerlere sahiptir. Basit ve karmaşık yapıların parametre kestirimlerinin RMSE, yanlılık ve korelasyonların bu şekilde farklılık göstermesi Finch (2011) ile tutarlık göstermektedir. Finch (2011) yaptığı çalışmada basit yapıda olmayan maddelerin basit yapıya göre daha yüksek yanlılık ve standart hata ile kestirildiğini, basit yapıdaki maddelerin daha kararlı kestirildiğini ifade etmiştir.

Bu çalışmadan madde parametrelerine ilişkin olarak elde edilen diğer bir bulgu ele alınan koşulların büyük bir kısmında d parametresine ait RMSE değerlerinin aynı koşullarda kestirilen a parametrelerinden daha düşük, d parametresine ait korelasyon ortalamalarının a parametrelerinininkinden daha yüksek olmasıdır. Bu bulgu Zhang (2012) ile Sheng ve Wikle (2007) tarafından yapılan çalışmalar ile örtüşmektedir. Zhang (2012) d parametrelerinin daha kararlı kestirildiğini; Sheng ve Wikle (2007) ise ele aldıkları koşullarda d parametrelerine ait RMSE'lerin a parametrelerinininkinden daha düşük olduğunu, dolayısıyla d parametrelerinin a parametrelerinden daha kararlı kestirildiğini ifade etmişlerdir.

Elde edilen sonuçlara dayanarak yapılacak parametre kestirimlerinde boyut sayısı arttıkça madde parametresi kestirimlerinin iyileştirilmesi için madde ve birey sayısı artırılmalıdır. BA-EM ile yapılacak kestirimlerde madde parametrelerinin RMSE değerlerinin düşük olması için iki boyutlu basit yapılarda en az 12 maddelik testler ve en az 2000 kişilik örneklemin kullanılması önerilmektedir. Ancak iki boyutlu karmaşık yapılar için 12 maddelik testlerde tatmin edici sonuçlara ulaşılamamaktadır. Bu yapılarda 48 maddenin yer aldığı testler ve 1000 kişilik örneklem büyüklüğü kullanıldığında kabul edilebilir düzeyde RMSE sonuçlarına ulaşılabilir. Üç boyutlu basit yapılarda 4000 örneklem ve üstünde 12 maddelik kısa testlerden faydalanılması önerilmektedir. Üç boyutlu karmaşık yapılarda ise 48 madde ve 4000 örneklemin üstüne çıkılmasına ihtiyaç vardır.

2. Birey Parametrelerine İlişkin Tartışma ve Yorum

Bu çalışmada ele alınan koşullarda madde parametrelerinin yanı sıra birey parametreleri de kestirilmiştir. Buna göre; ele alınan tüm boyutluluk koşullarında madde sayısı sabit tutulduğunda birey sayısının artması birey parametrelerinin (tetaların) kestirim iyiliğini önemli biçimde etkilememiş, teta parametrelerinin RMSE değerlerinde ciddi bir fark meydana getirmemiştir. Buna paralel olarak gerçek teta parametreleri ile kestirilen teta parametreleri arasındaki korelasyonlarının ortalaması da örneklem büyüklüğü artışından etkilenmemiştir. Ancak birey parametrelerinin kestirim iyiliği test uzunluğunun artması ile değişim göstermiştir. Buna göre 12 maddeli testten 48 maddeli teste geçildiğinde birey parametresi kestirimine ilişkin RMSE değerlerinde ciddi bir azalma meydana gelmiştir. Kirisci, Hsu ve Yu (2001), de la Torre ve Patz (2005), Köse (2010) ve Kieftenbeld ve Natesan (2012) çalışmalarında örneklemin büyümesinin birey parametrelerinin daha iyi kestirilmesine bir etkisinin olmadığını, ancak test uzunluğunun artmasıyla teta kestirim iyiliğinin geliştirilebileceğini belirtmişlerdir. Bu bakımdan bu çalışmadan elde edilen sonuçlar literatür ile paralellik göstermektedir.

Bu çalışmaya göre madde sayısının artmasıyla gerçek birey parametreleri ile kestirilen birey parametreleri arasındaki korelasyonlar yükselmiştir. Bu bulgu de la Torre ve Patz'ın (2005) bulguları ile tutarlıdır. de la Torre ve Patz (2005) yaptıkları çalışmada test uzunluğu ve boyut sayısı koşulları tek başına arttığında elde edilen korelasyonların da arttığını göstermişlerdir. Sheng ve Wikle (2007) ise yaptıkları çalışmada boyutlar arası korelasyonu değişen iki boyutlu basit yapıların teta parametrelerini incelemişlerdir. Buna göre bu çalışmaya en yakın koşulda teta1 ve teta2 parametrelerinin gerçek ve kestirilen değerleri arasındaki korelasyonu 0,80 ve 0,86 olarak bulmuşlardır. Eldeki çalışmada da Sheng ve Wikle'ın (2009) koşullarına benzeyen iki boyutlu basit yapının uzun test koşulunda elde edilen korelasyonlar 0,86 değerindedir. Bu açıdan elde edilen korelasyon değerleri ilgili alan yazın ile tutarlık göstermektedir. Madde parametrelerine benzer biçimde birey parametrelerinde de boyut sayısı arttıkça ve yapı basitten karmaşığa geçtikçe aynı koşullar için elde edilen RMSE değerleri artmış ve korelasyon değerleri azalmıştır.

Birey parametresi için basit yapılı ÇBMTK kestirimleri sonunda pozitif yanlışlıklar elde edilmiş diğer bir ifade ile birey parametreleri için üst kestirimde bulunulmuştur. Boyut sayısı arttıkça hesaplanan yanlışlık değerlerinin ranjı genişlemiştir. Bu sebeple birey parametresi için kestirimi etkileyen faktör sayısının artmasının kestirim iyiliğini düşürdüğü söylenebilir. Bu bulgu Lee'nin (2012) ÇBMTK'da madde parametrelerinin kestirim yanlışlığı ile ilgili bulgularıyla paralellik göstermektedir.

Elde edilen sonuçlara göre boyut sayısı arttıkça birey parametresi kestirimlerinin iyileştirilmesi için madde sayısı artırılmalıdır. Karmaşık yapıli testlerde kararlı kestirimler yapmak için, aynı boyut sayısına sahip basit yapıdaki testlerde kararlı kestirim yapılmasını sağlayan madde ve birey sayısından daha fazla madde ve bireye gerek olduğu göz önünde bulundurulmalıdır. Birey parametresi kestirimlerini iyileştirmek için, hesaplanan korelasyon değerleri göz önünde bulundurulduğunda, iki boyutlu (basit ve karmaşık) yapılar ve üç boyutlu basit yapılar için en az 12 maddelik, üç boyutlu karmaşık yapılarda ise en az 48 maddelik testlerin kullanılması önerilmektedir.

KAYNAKÇA

- Ackerman, T.A. (1996). Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement*, 20(4), 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(3), 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(1), 37-48.
- Babcock, B. G. E. (2009). *Estimating a noncompensatory IRT model using a modified Metropolis Algorithm*. Unpublished Doctoral Dissertation. University Of Minnesota Faculty Of The Graduate School.
- 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.
- Bulut, O. (2013). *Between-person and within-person subscore reliability: Comparison of unidimensional and multidimensional IRT models*. Unpublished doctoral dissertation, University Of Minnesota Faculty Of The Graduate School.
- de la Torre, J., & Patz, R. L. (2005). Making the most of what we have: A practical application of multidimensional item response theory in test scoring. *Journal of Educational and Behavioral Statistics*, 30(3), 295-311.
- DeMars, C.E. (2003). Sample size and the recovery of nominal response model item parameters. *Applied Psychological Measurement*, 27(4), 275-288.
- Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associate, Inc.
- Finch, H. (2010) Item parameter estimation for MIRT model: Bias and precision of confirmatory factor analysis based models. *Applied Psychological Measurement*, 34(1), 10-26.
- Finch, H. (2011) Multidimensional item response theory parameter estimation with nonsimple structure items. *Applied Psychological Measurement*, 35(1), 67-82.
- Hambleton, R. K., & Jones, R. W. (1993). An NCME module on comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.

- Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory principles and applications*. Kluwer-Nijhoff Publishing. Boston-USA.
- Hartig, J., & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*, 35(2009), 57–63
- Harvey, R.J., & Hammer, A. L. (1999). Item response theory. *The Counselling Psychologist*, 27(3), 353-383.
- Harwell, M., Stone, C.A., Hsu, T.C., & Kirisci L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement*, 20(2), 101-125.
- IRTPRO 2.1 for Windows (Item Response Theory for Patient-Reported Outcomes). *Scientific Software International*.
- IRTPRO: Users Guide (2011). *Scientific Software International*.
- 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.
- Kirisci, L., Hsu, T., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement*, 25(2), 146-162.
- Koğar, H. (2014). *Madde tepki kuramının farklı uygulamalarından elde edilen parametrelerin ve model uyumlarının örneklem büyüklüğü ve test uzunluğu açısından karşılaştırılması*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü.
- 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.
- Lee, J. (2012). *Multidimensional item response theory: An investigation of interaction effects between factors on item parameter recovery using Markov Chain Monte Carlo*. Unpublished doctoral dissertation, Michigan State University Measurement and Quantitative Methods.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64(2), 187-212.
- McDonald, P., & Paunone, S. V. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement*, 62(6), 921-943.
- Özkan, Y. Ö. (2012). *Öğrenci başarılarının belirlenmesi sınavından (ÖBBS) klasik test kuramı, tek boyutlu ve çok boyutlu madde tepki kuramı modelleri ile kestirilen başarı puanlarının karşılaştırılması*. Yayınlanmamış doktora tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü.
- Reckase, M. D. (2009). *Multidimensional item response theory (Statistics for social and behavioral sciences)*. New York: Springer.
- Sheng Y. and Wikle C. K. (2007). Comparing multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, 68(3), 413-430.
- SimuMIRT (2003). Software. *Lihua Yao*.
- Sinar, E.F., & Zickar, M. J. (2002). Evaluating the robustness of graded response model and classical test theory parameter estimates to deviant items. *Applied Psychological Measurement*, 26(2), 181-191.
- Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of MULTILOG. *Applied Psychological Measurement*, 16(1), 1-16.
- Sünbül, Ö. (2011). *Çeşitli boyutluluk özelliklerine 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.
- Turhan, A. (2006) *Multilevel 2PL item response model vertical equating with the presence of differential item functioning*. Unpublished doctoral dissertation. Florida State University.
- Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). Unidimensional IRT estimates the comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimates. *Applied Psychological Measurement*, 12(3), 239-252.
- Yao, L., & Boughton, K. A. (2007). A multidimensional item response modeling approach for improving subscale proficiency estimation and classification. *Applied Psychological Measurement*, 31(2), 83-105.
- Yavuz, G. (2014). *Çok boyutlu madde tepki kuramı modelleri ve paket programları için karşılaştırmalı analizler*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü.
- Zhang, J. (2012). Calibration of response data using MIRT models with simple and mixed structures. *Applied Psychological Measurement*, 36(5), 375-398.

EXTENDED ABSTRACT

Introduction

Through item response theory (IRT), a relationship between the performance of an examinee in the test and the observed ability or trait of that examinee which underlie this performance is presented. The relation between these observable and unobservable traits is explained by means of mathematical functions. Therefore, item response models are mathematical models. These models are based on rather significant assumptions (Hambleton and Swamington, 1985; Embretson and Reise, 2000). Assumptions of IRT are assessed under unidimensionality and local independence titles. Even though there are different models for different conditions in IRT, these models are not suitable for usage if the assumptions are not addressed. In terms of application, the situation is influenced due to the fact that tests that have been used are not unidimensional. Therefore, unidimensionality, one of the most fundamental assumptions of IRT, is violated. Therefore, in these and other similar cases; multidimensional item response theory (MIRT) shall be used instead of IRT; as MIRT is the expanded version of IRT which has been developed for multidimensional data sets.

Various models that are used in MIRT; assumes various statistical relationships between ability and probability for correct response. Additionally, the relation between dimensions and items can be defined according to the structure of the data set. The first structure is referred as "simple structure". A simple structured test includes multiple sub-tests. Each of these sub-tests measure a single ability or trait and has a homogenous dispersion. Each item is related to a single trait which is tried to be measured with the sub-test. Complex structure is the second type of structure than can be considered for a MIRT data set. Complex structure, which is also referred as a non-simple structure, is also known as within item model. The test itself, as well as the items of the test, is related to multiple abilities (Sheng and Wikle, 2007; Bulut, 2013).

IRT often provides better results than the classical test theory despite its applicability in the same fields (Hambleton and Jones, 1993; Harvey and Hammer, 1999; Sinar and Zickar, 2002; Macdonald and Paunonen, 2002). MIRT; fundamentals of which have been set forth in 1980s (Sympson, 1978; Doody-Bogan and Hattie, 1981; Yen, 1983; akt:1978 Ansley and Forsyth, 1985); have been addressed by various studies that have been performed outside Turkey. On the other hand, number of studies which have been performed in Turkey on MIRT is rather low. There are a number of doctorate dissertation studies which have been performed on the subject (Köse, 2010; Sünbül, 2011; Özkan, 2012, Koğar, 2014, Yavuz, 2014). These studies include either simple or complex structures. None of them include both simple and complex test structures with 2- and 3-dimensional tests. Accordingly; the study is structured to seek the answer for the problem of "What is the influence of sample size and test length on item and ability parameter estimations of simple and complex structured two and three dimensional tests that have been performed by means of multidimensional item response theory?"

Method

The data which is required for the study is acquired via simulation. SimuMIRT (2003) software has been used for generating data. Multidimensional structures have been considered as constants under the scope of the study. Therefore constants are as: two dimensional complex, two dimensional simple, three dimensional complex and three dimensional simple structures. Test length (12 and 48) and sample size (1000, 2000, 4000) are independent variables of the study. At the end a test design consists of 24 conditions (4 MIRT models * 2 test length * 3 sample size) is obtained. For each condition 25 replication is used. In total 600 data sets (24 conditons* 25 reps) are analyzed by IRTPRO 2.1. Following the data analysis; RMSE and bias statistics have been calculated by using estimated and actual values of item and ability parameters. In additon Pearson Product-Moment Correlation have been assessed to evaluate the relation between actual parameters and estimated parameters.

Results and Discussion

According to the results of the study; increase of the sample size has caused decrease in RMSE values of item parameters in both two and three dimensional tests. This finding is in compliance with results of the studies performed by Bolt and Lall (2003), Finch (2010) and Zhang (2012). A regular increase or decrease of bias, which is directly related to increase of item number or sample size, has

not been observed. But bias values showed that; the increase in bias is directly related to number of dimensions and complexity of the structure. This finding is in compliance with the assumptions of Lee (2012). Mean correlation values that have been acquired for item parameters have been measured to be higher for a1 and a2 parameters in simple structured tests; and for d parameter in complex structured tests. Increase in number of examinees haven't had a significant influence on estimation quality of ability parameters (thetas) when number of items is kept the same; and haven't caused a significant difference in terms of RMSE values under all dimensional conditions. Accordingly; mean correlation between actual tetra parameters and estimated tetra parameters hasn't been influenced by the increase of population. On the other hand; estimation accuracy of ability parameters has also shown a difference when test length is changed. In this context; RMSE values have decreased significantly in terms of estimation of ability parameters when 12-item tests have been expanded to 48 item tests. Kirisci, Hsu and Yu (2001), de la Torre and Patz (2005), Köse (2010) and Kieftenbeld and Natesan (2012) have indicated that; expansion of the sample size does not have any influence on estimation accuracy of ability parameters, yet, theta estimation success may be developed if the test length is increased. Under this perspective, it can be said that findings of this study are in compliance with the current literature.