



## Tek Boyutlu ve Çok Boyutlu Madde Tepki Kuramına Göre Çok Boyutlu Yapıların İncelenmesi\*

### Examining Multidimensional Structure in View of Unidimensional and Multidimensional Item Response Theory

Emrah GÜL\*\*, Nizamettin KOÇ\*\*\*

• *Geliş Tarihi:* 25.01.2016 • *Kabul Tarihi:* 28.06.2016 • *Yayın Tarihi:* 28.04.2017

**ÖZ:** Bu araştırmanın amacı, ikili puanlanan çok boyutlu yapıların, farklı örneklem büyüklüğü, boyutlararası korelasyon ve boyut sayısı koşulları altında, tek boyutlu ve çok boyutlu Madde Tepki Kuramı'na (MTK) dayalı olarak kestirilen madde ve birey parametrelerini karşılaştırmaktır. Simülasyon verileri ile yürütülen bu araştırma, temel araştırma niteliğindedir. Her iki modele göre kestirilen madde ve birey parametre değerlerine ait standart hatalar, hata kareleri ortalaması ile değerlendirilmiştir. Araştırmanın sonuçlarına göre, iki boyutlu veri yapıları ve boyutlararası korelasyon değerinin yüksek olduğu durumlarda tek boyutlu ve çok boyutlu MTK'dan elde edilen madde parametre değerlerine ait hatalar yüksek oranda farklılaşmamıştır. Bunun yanında, veri yapısının üç ve beş boyutlu olduğu durumda tek boyutlu MTK'dan elde edilen madde parametre değerlerine ait hata miktarları artmaktadır. Madde parametrelerine ait standart hatalar, örneklem büyüklüğünün artması ile daha düşük değerler almıştır. Yetenek parametre kestirimlerinden elde edilen standart hatalar incelendiğinde, tüm koşullar için çok boyutlu MTK'nın daha düşük hatalarla kestirim yaptığı görülmüştür. Araştırma sonucunda, çok boyutlu yapıların analiz edilmesinde çok boyutlu MTK'nın daha iyi sonuçlar verdiği, özellikle birey parametrelerinin kestiriminde ve bireyler hakkında kararlar vermede tek boyutlu MTK'ya göre daha doğru sonuçlar ürettiği belirlenmiştir. Ulusal ve uluslararası düzeyde uygulanan geniş ölçekli sınavlar, birçok alt boyuttan oluştuğundan, bu sınavlarda bireylerin yetenek kestirimleri için çok boyutlu modellerin kullanılması önerilebilir.

**Anahtar sözcükler:** Çok boyutlu MTK, tek boyutlu MTK, çok boyutlu veri, boyutluluk

**ABSTRACT:** The purpose of the study is to compare item-person parameters that are estimated based on unidimensional and multidimensional Item Response Theory (IRT) of dichotomously scored multidimensional constructs, under the following conditions: different sampling sizes, inter-dimensional correlation and the number of dimensions. The study which is carried out with simulative data is an example of basic research. The standard errors of the item-person parameters estimated on the basis of the both models are evaluated with root mean square error. According to the research findings, the two-dimensional data constructs and the errors in the item parameter values obtained from unidimensional and multidimensional Item Response Theory, where the inter-dimensional correlation value is found high, do not significantly vary. Additionally, the number of errors in the item parameter values obtained from unidimensional IRT, in which there is three-dimensional and five-dimensional data construct increases. The standard errors of the item parameters have lower values as the sampling size increases. When the standard errors obtained from the person parameter estimations are considered, it is seen that multidimensional IRT under any conditions estimates with lower errors. As a result of the study, it is concluded that multidimensional IRT is found to give better results in the analysis of multidimensional constructs and to give more accurate results than unidimensional IRT, particularly in person parameter estimations and decision making. It could be advisable to employ multidimensional models since large scale assessments on national and international scales have various sub-dimensions for personal skills estimation.

**Keywords:** Multidimensional item response theory, unidimensional item response theory, multidimensional data, dimensionality

## 1. GİRİŞ

Bilim insanı, her şeyden önce bilimin anlamı ve bilimsel düşünmenin niteliği üzerine sağlam bir anlayış kazanmış bireydir. Modern dünyanın doğru yorumlanması kuşkusuz "bilim" denilen karmaşık bilgi bütününe iyi anlaşılmasına bağlıdır (Yıldırım, 1973). İnsan var olduğu

\* Bu çalışma ilk yazarın doktora tezinden üretilmiştir

\*\* Yrd. Doç. Dr., Hakkâri Üniversitesi, Eğitim Fakültesi, Hakkâri/TÜRKİYE, emrahgul@hakkari.edu.tr

\*\*\* Prof. Dr., Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ankara/TÜRKİYE, nkoc@ankara.edu.tr

günden bu yana, bir taraftan evrende yaşananları anlama, tanımlama ve evreni çözümlenme, diğer yandan doğa ile güvenli bir hayat kurma isteği içinde olmuştur. Bu başatma çabaları, bilim denilen karmaşık bilgi bütününü oluşturmuştur. Bilimin kuşkusuz en önemli kavramlarından biri ölçmedir. Bilimin işlevleri dikkate alındığında, ölçmesiz bilimin varlığından söz etmenin mümkün olmadığı görülmektedir. Bu durum eğitim bilimi için de geçerlidir. Bilim ve teknolojiye gelişmeler ışığında çağın gereksinimleri ile uyumlu, yenilikçi ve entelektüel bireyleri yetiştirmede eğitimin rolü şüphesiz çok açıktır. Eğitim sisteminde bireylerin gereksinimlerinin hangi hedeflere ne düzeyde ulaşıldığının saptanması ve eğitim politikalarının bu doğrultuda geliştirilmesi, ölçme ve değerlendirme uygulamalarının gerekliliğini ve önemini ortaya koymaktadır (Baykul, 2000). Eğitimde ve psikolojide bireyler hakkında verilen kararların doğru ve güvenilir olması için araştırmacılar, geçerli ve güvenilir ölçmeler sağlayacak yeni yaklaşımlar geliştirmişlerdir. Bu yaklaşımların gelişimi de bilimsel bilginin gelişimine paraleldir. Eğitimde ve psikolojide ölçme ve değerlendirme bilim alanının temelini oluşturan kuramlar, Klasik Test Kuramı (KTK), tek boyutlu MTK ve çok boyutlu MTK'dır.

Eğitimde ve psikolojide ölçme ve değerlendirme bilim alanının temel kuramı olan KTK, 1905 yılında Charles Edward Spearman tarafından temellendirilmiştir (Akt: Crocker ve Algina, 1986). KTK'da bireyin maddeye verdiği yanıtlardan, bireyin gerçek puanı belirlenmeye çalışılmaktadır. Gerçek puan, ölçmeye karışan farklı hata türlerinden dolayı, ölçme aracılığıyla doğrudan elde edilemez (Baykul, 2000; Turgut, 1995). KTK'nın sınırlılıklarına karşın "MTK (Item Response Theory-IRT)", 1930'lu yılların sonlarında ortaya çıkmıştır. MTK aynı zamanda "Örtük Özellik Teorisi (Latent Trait Theory)" olarak da adlandırılmaktadır (Crocker ve Algina, 1986). Lord'un 1950'li yıllarda yaptığı çalışmalar MTK'nın matematiksel modellerinin ortaya konması bakımından kuramın başlangıcı olarak kabul edilebilir. Lord'un "normal ogive" modeli, 1950'li yıllarda Birnbaum'un "lojistik" modeli, 1960'lı yıllarda Danimarkalı matematikçi Rasch'ın geliştirdiği ve kendi adıyla anılan Rasch model, MTK'nın gelişiminde önemli roller oynamıştır (Akt: Baker, 2001; Hambleton ve Swaminathan, 1989). MTK belirli varsayımlar altında madde ve birey parametre değerleri kestirmektedir. MTK'nın, KTK ile ortak olan varsayımlarından biri tek boyutluluktur. Tek boyutluluk, bir ölçme işleminde, ölçülmek istenen özelliğin/yetenegin tek bir yapı göstermesidir. Bir diğer ifade ile maddelerin tek bir boyutu, tek bir özelliği ölçmesi demektir (Hambleton, Swaminathan ve Rogers 1991). Gerçek test uygulamalarında ve psikolojik yapılarda tek boyutluluk varsayımının karşılanması, bu kurama dayalı yapılan analizlerde ciddi problemlere yol açmaktadır.

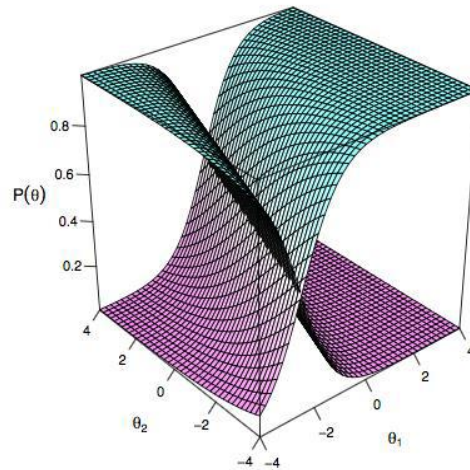
Tek boyutluluk varsayımının karşılanması veya veri yapısının tek boyutluluğu ne düzeyde gerektiğini ortaya koymak çok zordur (Reckease, 2007). Bunun bir sonucu olarak veri yapısının doğru bir şekilde temsil edildiği ve elde edilen bilgi miktarını artıracak farklı modellere ihtiyaç duyulmaktadır. Tek boyutlu MTK'nın bu sınırlılığına karşı geliştirilmiş çok boyutlu MTK modelleri bulunmaktadır. Çok boyutlu MTK tek boyutlu MTK'nın bir uzantısı olarak ortaya çıkmıştır. Çok boyutlu MTK, veri yapısı iki veya daha fazla boyuttan oluştuğunda, madde ile birey arasındaki etkileşimi inceleyen bir modeldir. Alanyazında birçok çok boyutlu MTK modeli olduğu görülmektedir. Bu modellerden biri de, tüm parametre değerlerini barındıran 3 parametrelili lojistik modeldir.

### 1.1. Çok Boyutlu Tamamlayıcı 3 Parametrelili Lojistik Model

Çok boyutlu MTK modelleri tamamlayıcı ve kısmen tamamlayıcı modeller olarak kategorize edilmiştir. Tamamlayıcı modeller bir boyuttaki yüksek yetenek ile ikinci boyuttaki düşük yetenek arasındaki etkileşime izin verir ve bu modeller çok boyutlu veri yapılarında bir yetenek düzeyindeki üstünlüğün, diğer boyut veya boyutlarda eksik olan yeteneği tamamlaması durumunda kullanılırlar. Örneğin bir fen testinde okuma becerisi, bireyin fen maddesine doğru yanıt verme olasılığını etkileyecektir. Böylelikle fen başarısı ve okuma başarısı arasında bir etkileşimden söz etmek mümkündür. Kısmen tamamlayıcı modellerde ise yetenek düzeyleri

arasında etkileşim sınırlıdır. Örneğin psikomotor davranışların ölçüldüğü bir testte, duyma ve görme yeteneklerinin bireyin işe alınmasında iki boyut olduğu düşünüldüğünde, işe alınmak için her iki yeteneğin belirli bir düzeyin üzerinde yüksek olmasını beklenir. Bir başka deyişle, bireyin duyma yeteneği, görme yeteneğini telafi edemez. Bu modeller için “tamamlayıcı olmayan” kavramı kullanılmamalıdır; çünkü bir boyutta yeteneği yüksek olan bireyin, bir maddeyi doğru yanıtlama olasılığı, o boyutta yeteneği düşük olan bireyden daha yüksek değildir. Kısmen tamamlayıcı modellerde her iki boyutta da bireyin yeteneğinin belirli bir seviyede olması gerekmektedir (Yoo, 2011).

Çok boyutlu MTK, iki veya daha fazla boyutlu bir testi alan birey ile bireyin testteki bir maddeye doğru cevap verme olasılığı arasındaki ilişkiyi matematiksel modeller ile açıklamaktadır. Bir başka deyişle, tek boyutlu MTK modellerinin çok boyutluluğa uyarlanmasıdır (Ackerman, Gierl ve Walker, 2003). Tek boyutlu MTK'nın uzantısı olarak çok boyutlu modellerden, ogive modeli Bock, Gibbons ve Muraki'nin, lojistik modelleri ise, Reckase ve McKinley'in temellendirdikleri görülmektedir. (Seungho Yang, 2007). Bir parametrelili lojistik tamamlayıcı modelde tek bir parametre kestirilir ve bu parametre, (d) tek boyutlu modellerde kullanılan güçlük parametresidir. Bu parametrede maddelerin şans olasılığından bağımsız, eşit ayırt edicilikte oldukları varsayılır. İki parametrelili modelde güçlüğü ek olarak madde ayırt ediciliği de modele girer ve her bir boyut için ayrı ayrı hesaplanır. Örneğin üç boyutlu modeller için  $a_1$ ,  $a_2$  ve  $a_3$  değerleri hesaplanır. Üç parametrelili tamamlayıcı modelde ise üçüncü parametre değeri olarak şans parametresi eklenmektedir. Tek boyutlu MTK'dan farklı olarak  $d$  sembolü ile gösterilen güçlük parametresi, karşılaştırmalar yapılırken karışıklıklara neden olmaması için bu çalışmada her iki modele dayalı olarak yapılan kestirimlerde  $b$  sembolü ile gösterilmiştir. Çok boyutlu MTK'da bir ya da daha fazla boyut için maddeye doğru cevap verme olasılığı madde karakteristik yüzeyleri (MKY) ile gösterilmektedir. Şekil 1'de MKY'leri görülmektedir.



**Şekil 1. İki boyutlu ikili puanlanan veri yapısı için madde karakteristik yüzeyleri**

Çok boyutlu tamamlayıcı 3 parametrelili lojistik model (3PLM), Reckase (2009) tarafından tek bir skalar değerden çok, parametre vektörü olarak tanımlanan 3PLM'nin genişletilmiş hali olarak tanımlanabilir. Model uyumu iyi olduğunda, elde edilen model,  $k$ -boyutlu uzay için verilen  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$  koordinatlarında doğru yanıtların koşullu olasılığını kestirir. Burada  $k$ , modelde ele alınan boyut sayısıdır. Çok boyutlu tamamlayıcı 3PLM Eşitlik 1 ile açıklanabilir:

$$P(u_{ij} = 1 | \theta_j, \mathbf{a}_i, c_i, d_i) = c_i + (1 - c_i) \frac{e^{a_i \theta_j + d_i}}{1 + e^{a_i \theta_j + d_i}} \dots \text{Eşitlik 1}$$

Burada;

$P(u_{ij} = 1 | \theta_j, \mathbf{a}_i, c_i, d_i)$ ,  $\theta$  düzeyindeki bir bireyin  $i$  maddesini doğru cevaplama olasılığıdır.

$\mathbf{a}_i$ ,  $i$  maddesinin ayırıcılık gücü ile ilgili parametrelerin  $1 \times k$  vektörüdür.

$c_i$ , şans parametresidir.

$d_i$ , tek boyutlu MTK'da güçlük parametresine karşılık gelmektedir.

$\theta$ , bireyin tüm boyutlarındaki yetenek puanlarının  $1 \times k$  vektörüdür.

Eğitimde ve psikolojide kullanılan testlerde, birey ve madde arasındaki gerçek etkileşim tek boyutlu modellerde açıklandığı gibi basit değildir. Gibbons, Immekus ve Bock (2007) tek boyutlu olmayan veri yapılarında, tek boyutlu MTK'ya göre kestirilen madde parametreleri ile ilgili bir alanyazın taraması yapmışlardır. İncelenen çalışmalarda iki önemli sonuca dikkat çekilmiştir. Birincisi, veri yapısı genel bir boyut altında toplanıyorsa ve boyutluluk yapısı çoğunlukla bu faktör tarafından açıklanıyorsa, tek boyutlu modellerin parametre kestiriminin kesinliği artmaktadır. İkinci olarak ise, veri yapısı tek bir güçlü boyutun dışında farklı boyutları da barındırıyorsa, tek boyutlu analizler veri yapısındaki yalnızca en güçlü boyut ile ilgili kestirim yapmaktadır. Bu özelliklere sahip veri yapılarına uygulanan tek boyutlu modeller ölçme kesinliğini azaltmaktadır.

Klasik Test Kuramı ve Tek Boyutlu MTK'ya göre ölçülen özelliğin tek boyutlu bir yapıya sahip olması gerekmektedir (Hambleton ve Swaminathan, 1989). Fakat psikolojik özelliklerin tek boyutla ifade edilmesinin sınırlı olduğu ve özellikle başarı testlerinde maddeye tepki verilirken birden çok bilişsel sürecin etkili olduğu, bu süreçlerin işlevlerinin bireyden bireye farklılık gösterdiği fikirleri ortaya çıkmaya başlamıştır. Bu eleştiriler özellikle KTK ve MTK'nın tek boyutluluk varsayımına yönelik olmuştur (Kırisci, Hsu ve Yu, 2001). Ackerman (1996) eğitimde ve psikolojide kullanılan birçok testin farklı düzeylerde çok boyutlu özellikleri ve bu özelliklerin bir bileşimini ölçtüğünü belirtmiştir. Tek boyutlu MTK modellerinin tek boyutluluk varsayımı ile başarı veya tutum gibi karmaşık psikolojik yapıları basitleştirdiği ve bireyin gerçek performansı hakkında daha az bilgi verdiği öne sürülmektedir (Reckase, 2007).

Bir özelliği ölçmek için geliştirilen bir testin yalnızca o özelliği ölçmesi veya tek boyutlu olması görüşünün eğitimde ve psikolojideki yapılar için geçerli olması çok zordur. Farklı boyutluluk düzeylerinde geliştirilen testler, test ile ölçülen özelliğin dışında kalan özellikleri de ölçebilmektedir. Örneğin bir birey bir fen testi ile ilgili bir maddeyi yanıtlarken okuma becerisinin de maddeye doğru yanıt verme olasılığını etkilediği görülmektedir. Bu durumda fen başarısı ile okuma becerisi farklı boyutlar altında toplanabilecektir. Tek boyutlu modeller her iki boyutu yalnızca fen başarısı olarak değerlendirmektedir. Bireylerin okuma becerilerindeki farklılık onların fen başarılarını etkileyeceğinden, bu testin tek boyutlu bir yapıya sahip olduğunu iddia etmek ölçme işleminde geçerlik problemine neden olabilir. Tek boyutluluk varsayımının ihlali veya bu varsayımın bir derece sorunu olduğu iddiası ile yapılan kestirimlerde yalnızca başat yani ölçülen özellik bakımından en yüksek varyansa sahip olan boyutun ölçtüğü yetenek kestirilmekte, diğer boyut veya boyutlarla ilgili olarak ise sınırlı bilgiler elde edildiği vurgulanmaktadır (Kreiter, 1993).

Çok boyutlu MTK, maddenin güçlük ve ayırt ediciliği olarak tanımlanan bir parametre vektörü ve bilgi-beceri olarak tanımlanan bireyin çoklu yeteneğinin bir vektörünü içeren, matematiksel bir fonksiyon üzerine inşa edilmiştir (Reckase, 2009). Tek boyutlu MTK modelleri tek boyutluluk varsayımı sağlandığında güçlü kestirimler yapmasına rağmen, birey ve

madde arasındaki karmaşık etkileşimleri incelemede yetersiz kalmakta ve bu nedenle daha gelişmiş modellere duyulan ihtiyaç açığa çıkmaktadır.

Ölçme ve değerlendirme uygulamalarında Klasik Test Kuramı'na dayalı olarak kestirilen parametrelerin sınırlılıklarının olması verilen kararların da sınırlı olmasına neden olmakta ve bu durum MTK'nın önemini arttırmaktadır. KTK'da testten kestirilen madde ve yetenek parametrelerinin testin geliştirildiği gruba bağımlı olması, testlerin geçerlik ve güvenilirliğinin uygulandıkları gruba bağlı olması sınırlılığını doğurmaktadır. Testi alan bireyin testte nasıl bir performans göstereceğinin bilinmemesi, KTK'nın bir diğer önemli sınırlılığıdır. Bunun yanında, tek boyutlu MTK'nın psikolojik özelliklerin tek boyutlu olması gerekliliği, testler ile ilgili daha az bilgi sağlaması ve bu durumunda bireyler hakkında verilen kararları etkilemesi çok boyutlu kuramların geliştirilmesi gerekliliğini doğurmuştur.

Tek boyutlu MTK, KTK'dan daha iyi bir çerçeve sunmasına rağmen, araştırmalarda kullanılan veri yapılarının tek boyutluluk varsayımına tamamen uymadığı, tek boyutlu MTK'nın bunu bir derece sorunu olarak ifade ettiği ve bir gizil değişkeni ölçtüğünü varsayarak kestirimler yaptığı görülmektedir. Bunun yanında, çok boyutlu veri için tek bir özellik boyutunun olduğunu varsaymak, MTK'nın değişmezlik özelliğini de tehdit etmektedir (Ackerman, 1994). Tek bir örtük özelliği ölçmek için çok boyutlu maddeler kullanmak, testin yapı geçerliğine zarar vermektedir. Bunun yanı sıra test birden fazla boyutu ölçüyorsa, test alanları tek bir boyut altında sıralamak, alanyazınında da belirtildiği gibi, ölçme ile ilgili problemlere neden olmaktadır. Bu problemlerin ortadan kaldırılması çok boyutlu MTK araştırmalarını gerekli kılmaktadır. Çünkü bahsedilen yapıların geçerli bir şekilde ölçülmesi alınan kararların doğruluğunu artırmaktadır.

Yapılan araştırmalara bakıldığında, Wiberg (2012) araştırmasında çok boyutlu veri yapısına sahip beş alt testten oluşan bir üniversite giriş sınavını tek boyutlu MTK ile analiz etmiştir. Tek boyutlu ve çok boyutlu modelleri kullanarak madde ve yetenek parametre kestirimlerini simülatif olarak yapmıştır. Sonuç olarak üniversite sınavı gibi çok boyutlu bir verinin, çok boyutlu MTK'ya göre analiz edilmesinin daha iyi sonuçlar ürettiği belirlemiştir. Testin beş boyutlu olarak kabul edildiği ve çok boyutlu MTK'ya göre analizlerin yapıldığı durumda daha az hata ile kestirimlerin yapıldığı görülmektedir. Fakat testi konu alanına göre iki boyutlu kabul edildiği ve tek boyutlu MTK ile analiz edildiği durumda hem hata miktarının yükseldiği hem de ölçme kesinliğinin düştüğü sonucuna ulaşılmıştır. Eunike ve Benedikt (2014) yaptıkları araştırmada, genellikle tek boyutlu MTK ile analiz edilen "Mesleki İlgi Envanteri"ni çok boyutlu MTK'ya göre incelenmişlerdir. Araştırma sonucunda, çok boyutlu yapısı olan mesleki ilgi envanterinin çok boyutlu MTK'ya göre yapılan kestirimlerinin daha iyi sonuçlar verdiğini belirtmişlerdir. Bashkov (2015) yaptığı araştırmada, Cai (2010) tarafından geliştirilen ve bu araştırmada da kullanılan MH-RM algoritmasının etkililiğini araştırmıştır. Diğer algoritmaların aksine, yüksek çok boyutlu (üç boyut ve üzeri) modellerde MH-RM algoritmasının daha iyi sonuçlar verdiğini ve analizlerin daha az zaman aldığını belirtmiştir.

Çok boyutlu veri yapıları, eğitimde ve psikolojide yıllardır kullanılmasına rağmen, bu veri yapılarının tek boyutluluk varsayımı altında çözümlenmesinin ne tür sıkıntılara neden olabileceği ortaya konamamaktadır. Alanyazın incelendiğinde tek boyutlu MTK'nın, tek boyutluluk varsayımının getirdiği sınırlılıkları çok boyutlu MTK'nın ortadan kaldırıp kaldırmadığına ilişkin araştırmaların sınırlı olduğu ve bu yönüyle araştırmanın çok boyutlu yapıların geliştirilmesinin ve yorumlanmasının gerekli olduğu vurgulanmaktadır. Bireyler hakkında önemli kararların verildiği geniş ölçekli ve çok boyutlu testlerden elde edilen sonuçların doğruluğu ve kesinliği, bireyler açısından çok önemlidir. Bu anlamda bu araştırma geniş ölçekli ve çok boyutlu test uygulamalarında hangi modellerin kullanılacağına dair bilgi vermesi açısından da önemlidir. Eğitimde ve psikolojide karşılaşılabilecek başarı, tutum ve yetenek gibi veri yapılarını benzer yapıları simülasyon ile oluşturarak, kuramın sınırlı ve güçlü

yanları uygulamaya geçilmeden ortaya konmaktadır. Bu araştırma, uygulamada karşılaşılabilecek olumsuz durumları simülasyon ile üretilmiş veri setleri ile ortaya koyması bakımından da önemlidir.

Bu tartışmalar doğrultusunda, bu araştırmanın amacı, farklı boyutluluk özelliklerine sahip ikili puanlanan (1-0) veri yapılarının tek boyutlu ve çok boyutlu MTK'ya dayalı olarak kestirilen madde ve yetenek parametrelerinden elde edilen standart hataları karşılaştırmaktır. Bu karşılaştırma, değişen koşullar altında simülasyon ile yapılmıştır. Bu koşullar; a-) testin ölçtüğü boyut sayısı, b-) boyutlar arasındaki korelasyon, c-) örneklem büyüklüğüdür.

## 2. YÖNTEM

### 2.1. Araştırmanın Modeli

Bu araştırmada tek boyutlu ve çok boyutlu MTK çerçevesinde, çok boyutlu veri yapıları, bu yapılarda yer alan boyutlar arası ilişkiler ve farklı örneklem büyüklükleri dikkate alınarak incelenmiştir. Farklı test koşullarına uygun veriler üretildiği için, bu araştırma bir simülasyon çalışmasıdır. Araştırmada simülasyon verileri ile yöntemleri etkileyecek değişkenler kontrol edilmekte ve yöntemlerin üstünlükleri ve sınırlılıkları incelenmektedir. Bu yönüyle de yöntemlerin geliştirilmesine katkı sağlayan temel bir araştırma olduğu ifade edilebilir (Karasar, 2010).

### 2.2. Simülasyon Verilerinin Üretilmesi ve Analizi

Alanyazında, simülasyon çalışmalarında farklı programlarının (R, MULTISIM vb) kullanıldığı görülmektedir. Bu araştırmada farklı özelliklerde ikili puanlanan (1-0) veri yapıları üretmek için, Luecht (2004) tarafında geliştirilen MIRTGEN 2.0 yazılımı kullanılmıştır. Veri üretimi iki, üç ve beş boyutlu, basit yapılı, iki kategorili (1-0) ve 3 parametrelili tamamlamalı modele uygun olarak gerçekleştirilmiştir. Araştırmada incelenen koşullar alanyazında yer alan tartışmalar doğrultusunda oluşturulmuştur. Way ve diğ. (1988) simülasyon veri setlerinin test maddelerine verilecek yanıtları yansıtacak biçimde ele alınması gerektiğini vurgulamışlardır. Bu nedenle, tüm boyutlarda yer alan maddelerin a (madde ayırıcılık) değerleri ortalaması 0.00 ve standart sapması 1.00 olan lognormal dağılımdan ve b (madde güclüğü) değerleri ise, minimum değeri -1.75, ile maksimum değeri +1.75 aralığında üretilmiştir (Reckase, 2009). Şans parametre değerleri  $c=0.15$ 'e sabitlenmiştir. Yetenek parametreleri ise standart normal dağılıma göre üretilmiştir. Veri yapısında boyut sayısı, boyutlararası korelasyon ve örneklem büyüklüğü değiştirilerek 25 tekrar yapılmıştır. Harwell ve diğ. (1996), tekrar sayısının üretilen hatalar üzerindeki etkisini araştırmış, 25 tekrardan sonra hataların azaldığı ve etkinin 1.0 düzeyine yaklaştığı rapor edilmiştir. Bu nedenle bu araştırmada da 25 tekrar kullanılması uygun görülmüştür. Böylelikle,  $3 \times 3 \times 3 = 27$  farklı deneysel koşul için  $27 \times 25 = 675$  farklı veri seti üretilmiştir. Bu araştırmada incelenmesine karar verilen en fazla boyut sayısı beştir. Dolayısıyla her boyutta en az 6 madde yer aldığındaki toplam test uzunluğunun 30 maddeye sabitlenmesine karar verilmiştir. Bir diğer deyişle, iki boyutlu veri yapısı için, her boyutta 15 madde; üç boyutlu veri yapısı için her boyutta 10 madde ve beş boyutlu veri yapısı için ise her boyutta altı madde olacak şekilde tüm test uzunlukları 30 olarak belirlenmiştir. Çalışmada kullanılan tüm deneysel koşullar Tablo 1'de sunulmuştur.

**Tablo 1: Araştırmada kullanılan koşullar**

Örneklem büyüklüğü	Boyut Sayısı	Boyutlararası Korelasyon	Analiz
500	2	0.10	Tek Boyutlu
			Çok Boyutlu
	3	0.40	Tek Boyutlu
			Çok Boyutlu
	5	0.70	Tek Boyutlu
			Çok Boyutlu
1000	2	0.10	Tek Boyutlu
			Çok Boyutlu
	3	0.40	Tek Boyutlu
			Çok Boyutlu
	5	0.70	Tek Boyutlu
			Çok Boyutlu
2000	2	0.10	Tek Boyutlu
			Çok Boyutlu
	3	0.40	Tek Boyutlu
			Çok Boyutlu
	5	0.70	Tek Boyutlu
			Çok Boyutlu

### 2.3. Verilerin Analizi

Çalışmanın analizleri flexMIRT 2.0 yazılımı kullanılarak yapılmıştır. flexMIRT çok düzeyli ve çok boyutlu MTK ile tek boyutlu MTK analizlerini yapan bir yazılımdır. flexMIRT 2.0 yazılımı kestirimleri yapmak için Bock-Aitkin Expectation-Maximization (BA-EM), Adaptive Quadrature ve Metropolis-Hastings Robbins-Monro (MH-RM) algoritmaları kullanmaktadır (Cai, 2013). Tek boyutlu analizler flexMIRT programında hızlı bir şekilde yapılmaktadır. Çok boyutlu analizlerin süresi, boyut sayısı ve örneklem büyüklüğü arttıkça uzamakta ve beş boyutlu ve 2000 örneklem büyüklüğüne sahip veri setinde maksimum süreye ulaşmaktadır. Bock-Aitkin Expectation-Maximization (BA-EM) algoritması, veri yapısının üç boyutu geçmediği durumlarda önerilmektedir. Fakat beş boyutlu yapıların incelenmesinde Metropolis-Hastings Robbins-Monro (MH-RM) algoritmasının kullanılması önerilmektedir.

Tek boyutlu ve çok boyutlu MTK'ya dayalı olarak yapılan analizler sonucunda elde edilen birey parametrelerine ait hata miktarı RMS (Root Mean Square) değerleri hesaplanarak karşılaştırılmıştır. RMS, sürekli olarak değişen bir fonksiyonun sürekli olmayan değer serisi için, hata kareleri ortalamasının karekökü alınarak hesaplanmaktadır (Toit, 2003). RMS hesaplamaları için Eşitlik 2 kullanılmıştır:

$$RMS = \sqrt{\frac{\sum_{i=1}^n se^2}{n}} \dots \text{Eşitlik 2}$$

*se*: Her bir parametre kestiriminden elde edilen standart hata miktarları

*n*: Kestirilen parametre sayısı

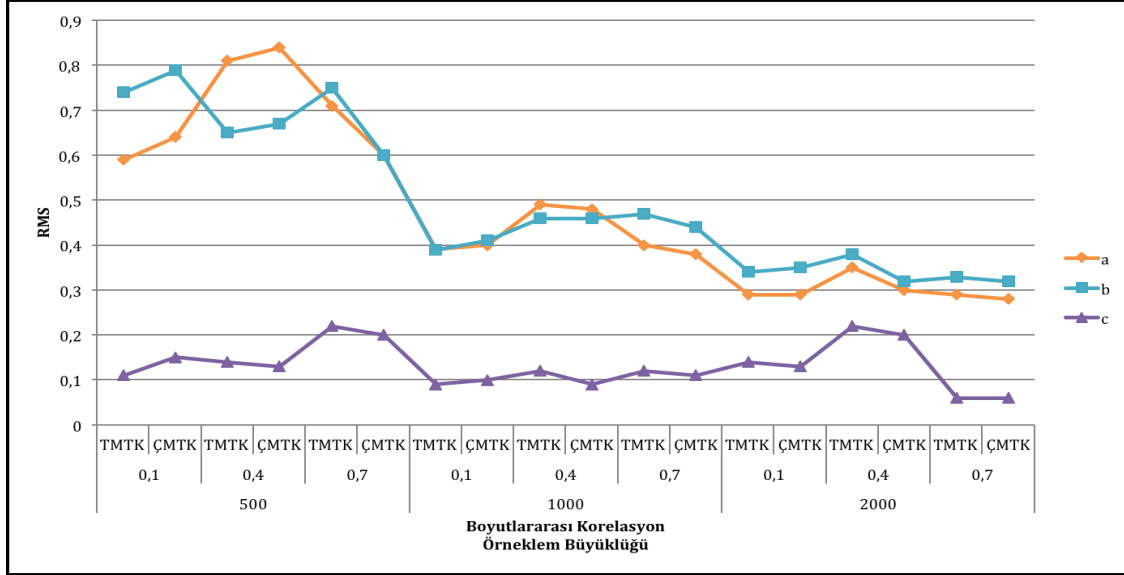
RMS: Root Mean Square

### 3. BULGULAR

Araştırmanın bu bölümünde, iki, üç ve beş boyutlu veri yapılarında tek boyutlu ve çok boyutlu MTK'ya dayalı kestirimlerden elde edilen standart hatalara ilişkin bulgulara yer verilmiştir.

### 3.1. İki Boyutlu Veri Yapısına İlişkin Bulgular

İki boyutlu veri yapısı için, tek boyutlu ve çok boyutlu MTK'ya göre kestirilen madde parametre değerlerine ilişkin RMS'ler Şekil 2'de sunulmuştur.

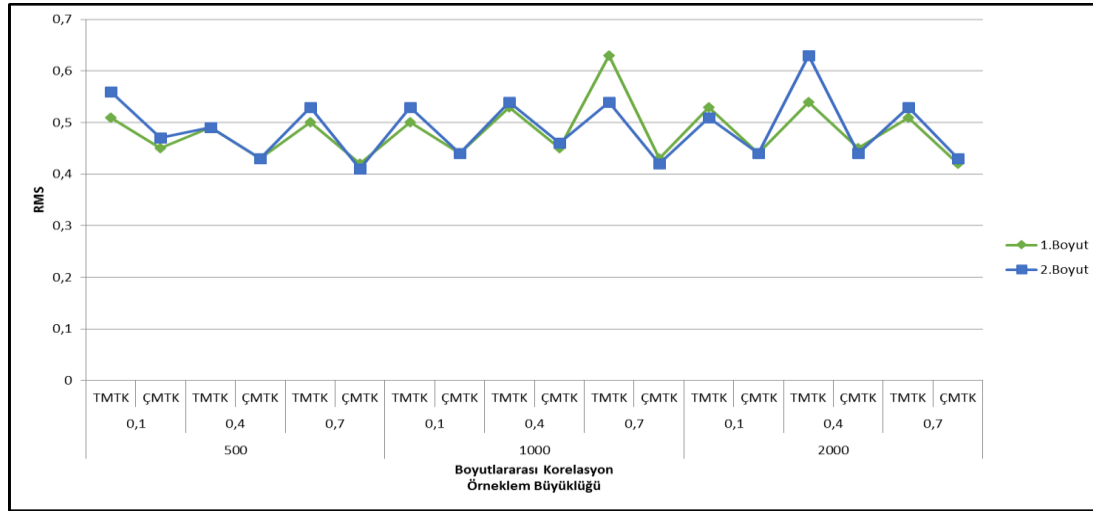


Şekil 2. Madde parametreleri standart hata kareleri ortalamaları

Şekil 2 incelendiğinde, tek boyutlu MTK'ya göre kestirilen hata kareleri ortalaması (RMS) a parametresi için, 0.29 ile 0.81; b parametresi için, 0.33 ile 0.75 ve c parametresi için ise 0.06 ile 0.22 aralığında değişmektedir. Çok boyutlu MTK'ya göre kestirilen hata kareleri ortalamaları a parametresi için, 0.28 ile 0.84; b parametresi için, 0.32 ile 0.79 ve c parametresi için, 0.06 ile 0.20 aralığında değişmektedir. Şekil 1'e göre, örneklem büyüklüğü arttıkça, hata miktarının hem tek boyutlu, hem de çok boyutlu MTK'da azaldığı görülmektedir. Bu bulgu, örneklem büyüklüğü arttıkça, RMS değerlerinin düştüğünü göstermektedir. Bu bulgular Zhang (2012), Bolt ve Lall (2003) ve Finch (2010) tarafından yapılan araştırmaların sonuçları ile benzerdir. RMS değerleri incelendiğinde, örneklem büyüklüğü 500 ve boyutlararası korelasyon 0.10 ve 0.40 iken tek boyutlu MTK'nın; örneklem büyüklüğü 1000 ve 2000 iken ise çok boyutlu MTK'nın daha az hatalı sonuçlar ürettiği belirlenmiştir. Bir diğer deyişle, örneklem büyüklüğü arttıkça, çok boyutlu modeller daha düşük hata ile madde parametresi kestirmektedirler. Babcock (2009) çok boyutlu MTK modelleri ile ilgili yaptığı araştırmada, örneklem büyüklüğünün 4000 olduğu durumda parametre kestirimlerinin iyileştiğini belirlemiştir. Fakat gerçek yapılarda 4000 örneklem büyüklüğüne ulaşmak çok zordur, bu nedenle bu araştırmada en çok 2000 örneklem büyüklüğü kullanılmıştır. Bu bulgu, büyük örneklemelerde daha düşük standart hata ile kestirim yapıldığını gösterdiğinden, bu araştırmanın bulgusu ile paralellik göstermektedir.

Tek boyutlu ve çok boyutlu MTK'ya dayalı iki boyutlu veri yapısında kestirilen birey parametrelerine ilişkin standart hata kareleri ortalamaları Şekil 3'te sunulmuştur.



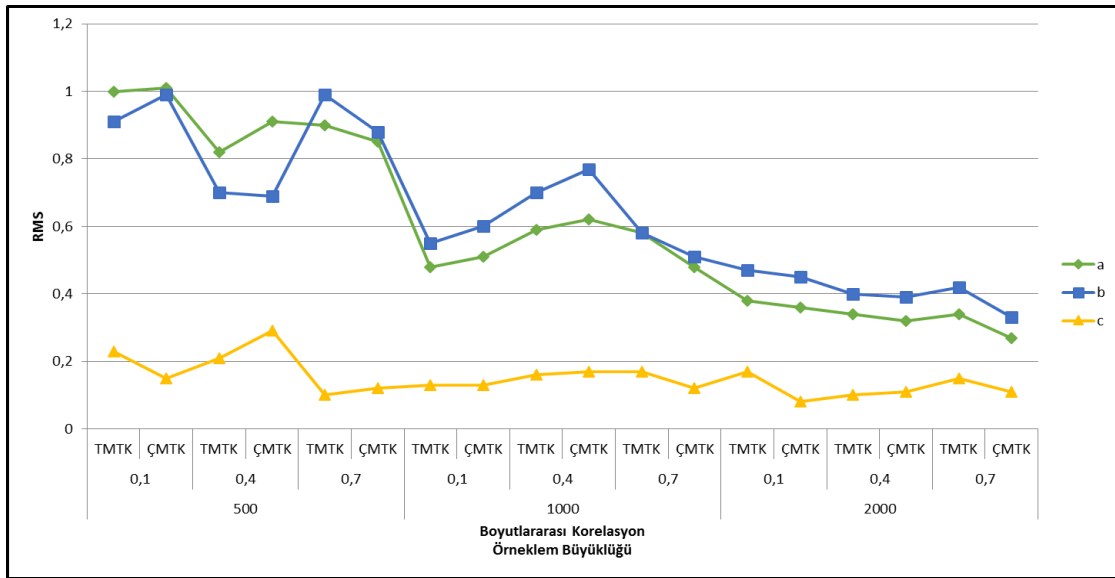


Şekil 3. Birey parametreleri standart hata kareleri ortalamaları

Şekil 3 incelendiğinde, tek boyutlu MTK'ya dayalı kestirilen yetenek parametrelerine ait hata kareleri ortalamasının 1.boyutta 0.49 ile 0.63; 2.boyutta ise, 0.49 ile 0.64 aralığında değiştiği görülmektedir. Çok boyutlu MTK'ya dayalı kestirilen yetenek parametre değerlerine ilişkin hata kareleri ortalamasının, 1.boyutta 0.42 ile 0.45; 2. boyutta ise 0.41 ile 0.47 aralığında değiştiği görülmektedir. Şekil 2'de, çok boyutlu MTK'nın, tek boyutlu MTK'ya göre daha dar aralıklarda hatalar ürettiği belirlenmiştir. Şekil 2'ye göre çok boyutlu MTK, kestirilen tüm birey parametreleri ve tüm araştırma koşullarında daha düşük hatalar üretmiştir. Boyutlararası korelasyon düzeyinin artması veya azalması, elde edilen RMS değerlerinde herhangi bir örüntüye neden olmamıştır. Luecht (2004) tarafından yapılan çalışmada KTK ve farklı MTK modelleri ile yapılan yetenek kestirimlerinde, çok boyutlu MTK'nın diğer kuramlara göre daha az hatalı kestirimler yaptığı sonucuna ulaşılmıştır. Bu araştırmadan elde edilen bulgu da Luecht tarafından yapılan çalışma ile paraleldir. Walker ve Bretvas (2003) tarafından yapılan çalışmada, veri yapısının tek boyutlu olduğunu varsayarak yapılan analizlerle, bireyin yeteneği ile ilgili yanlış kararlar verilebileceği belirtilmiştir. Ayrıca bu bulgu, Özer-Özkan'nın (2012) gerçek veriler üzerinde KTK, tek boyutlu ve çok boyutlu MTK modellerini kullanarak yaptığı araştırmada, çok boyutlu MTK'nın yetenek kestiriminde daha az hata ürettiği sonucu ile de tutarlıdır.

### 3.2. Üç Boyutlu Veri Yapısına İlişkin Bulgular

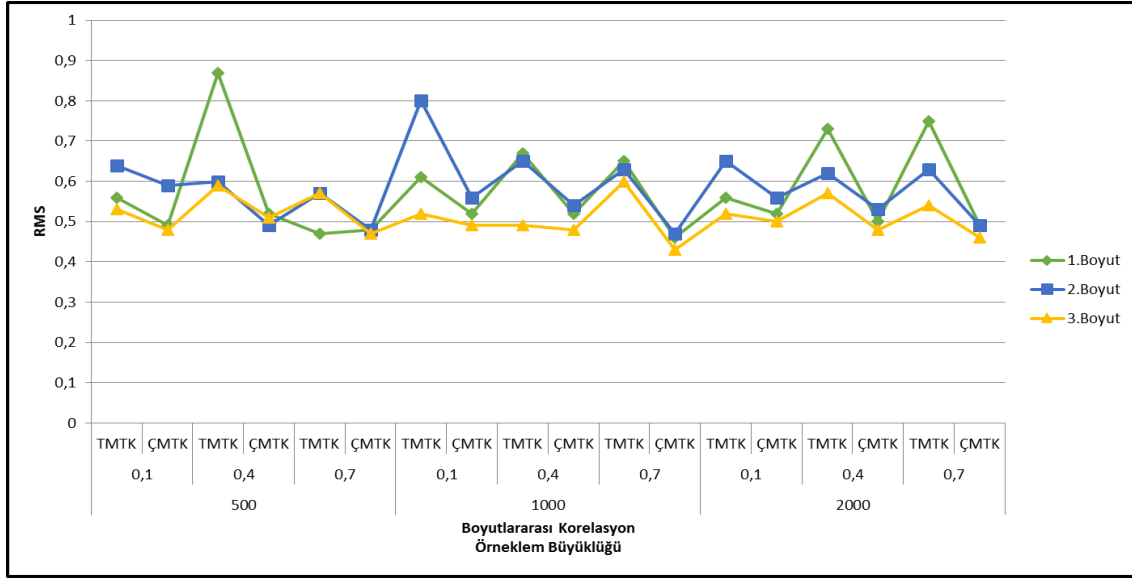
Tek boyutlu MTK ve çok boyutlu MTK'ya göre kestirilen madde parametre değerlerine ilişkin standart hata kareleri ortalamaları Şekil 4'te sunulmuştur. Şekil 4 incelendiğinde, tek boyutlu MTK'ya göre kestirilen hata kareleri ortalaması değerlerinin a parametresi için 0.34 ile 1.00; b parametresi için, 0.40 ile 1.36 ve c parametresi için ise 0.10 ile 0.23 aralığında olduğu görülmektedir. Çok boyutlu MTK'ya göre kestirilen hata kareleri ortalaması a parametresi için, 0.27 ile 1.01; b parametresi için, 0.33 ile 0.99 ve c parametresi için ise, 0.08 ile 0.29 aralığında değiştiği görülmektedir.



Şekil 4. Madde parametreleri standart hata kareleri ortalamaları

Şekil 4'te her iki kurama dayalı yapılan analizlerde, örneklem büyüklüğü ve boyutlararası korelasyon yükseldikçe, hata miktarının azaldığı gözlenmektedir. Tek boyutlu ve çok boyutlu MTK ile elde edilen RMS değerleri incelendiğinde belirli bir örüntü olmaksızın, örneklem büyüklüğünün 1000'i aştığı durumlarda çok boyutlu MTK'nın daha az hatalı sonuçlar ürettiği belirlenmiştir. Boyutlararası korelasyon sabit iken, örneklem büyüklüğünün artması, her iki kurama göre, kestirilen hata miktarlarının düşmesine neden olmaktadır. Araştırma bulgularına göre örneklem büyüklüğünün artması RMS değerlerinde düşüşe neden olmaktadır. Boyutlararası korelasyonun değişmesi, RMS değerleri ile ilgili herhangi bir örüntü oluşturmamıştır. Bu bulgular, Ansley ve Forsyth'in (1985) yaptıkları araştırmadan elde ettikleri sonuçlarla uyumludur.

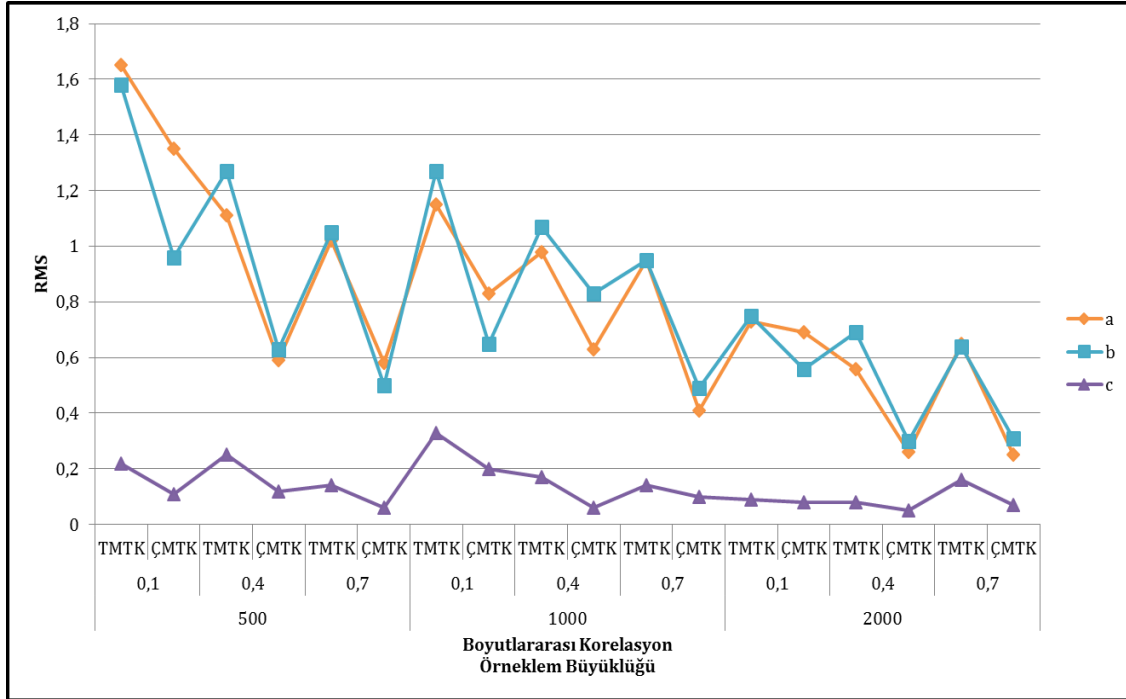
Tek boyutlu ve çok boyutlu MTK'ya dayalı kestirilen birey (yetenek) parametrelerine ait standart hata kareleri ortalamaları Şekil 5'te verilmiştir. Çok boyutlu MTK'ya göre kestirilen yetenek parametre değerlerine ilişkin hata kareleri ortalamasının, 1.boyutta 0.46 ile 0.52; 2. boyutta 0.47 ile 0.59 ve üçüncü boyutta, 0.43 ile 0.51 aralığında değiştiği görülmektedir. Bu değerlere göre çok boyutlu MTK'nın daha dar aralıklarla standart hata kestimi yaptığını göstermektedir. Çok boyutlu MTK'ya göre kestirilen tüm birey parametreleri, tüm deneysel koşullarda daha düşük hatalar üretmiştir. Bunun yanında 2000 örneklem büyüklüğünde, boyutlararası ilişki düzeyi arttıkça tek boyutlu MTK ile çok boyutlu MTK'ya dayalı yapılan birey parametre kestirim hataları arasındaki fark artmaktadır. Bu bulgu Luecht'in (2004) ve Özer-Özkan'ın (2012) yaptıkları araştırmalardan elde ettikleri bulgular ile benzerdir. Wiberg (2012) araştırmasında çok boyutlu veri yapısına sahip bir beş alt testten oluşan üniversite giriş sınavının tek boyutlu MTK ile analiz etmiştir. Tek boyutlu ve çok boyutlu modelleri kullanıldığı bu araştırmada, yetenek parametre kestirimlerinin sonucu olarak üniversite sınavı gibi çok boyutlu alt testlerden oluşan bu verinin çok boyutlu MTK'ya göre analiz edilmesinin daha iyi ve daha az hatalı sonuçlar ürettiği bulgusu da bu araştırmanın bulgusunu desteklemektedir.



Şekil 5. Birey parametreleri standart hata kareleri ortalamaları

### 3.3. Beş Boyutlu Veri Yapısına İlişkin Bulgular

Tek boyutlu MTK ve çok boyutlu MTK' ya göre kestirilen madde parametre değerlerine ilişkin RMS değerleri Şekil 6'da sunulmuştur.

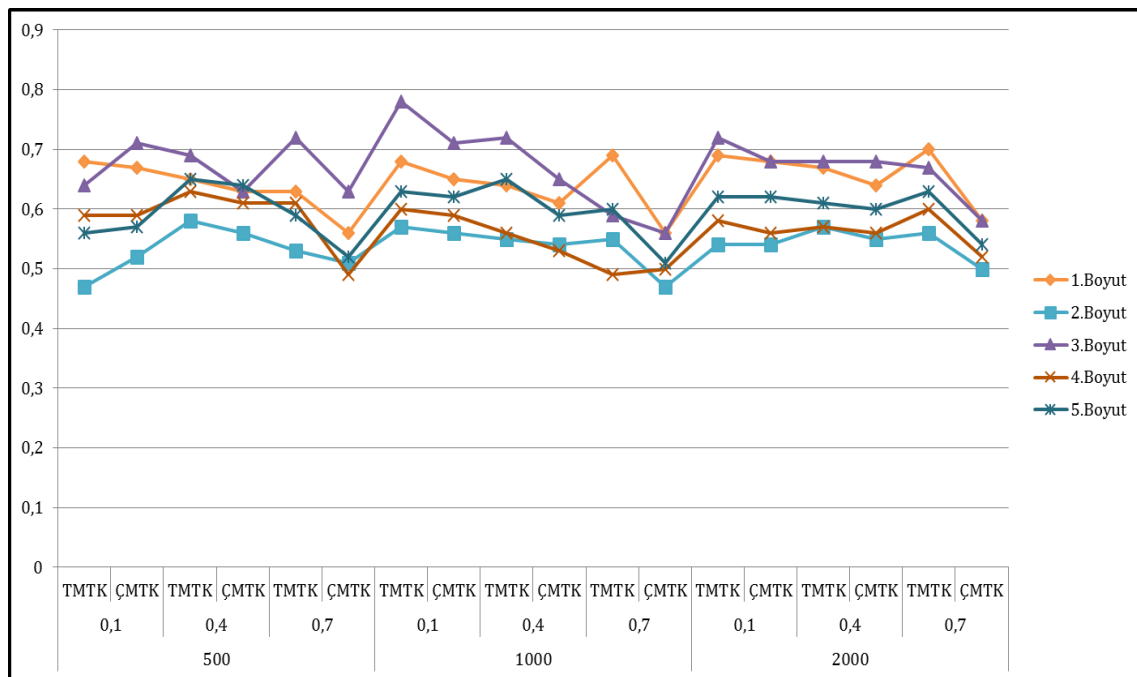


Şekil 6. Madde parametreleri standart hata kareleri ortalamaları

Şekil 6'da araştırma koşullar dikkate alındığında örneklem büyüklüğü ve boyutlararası korelasyon artıça hata miktarının azaldığı gözlenmektedir. Tek boyutlu ile çok boyutlu MTK'dan elde edilen RMS değerlerine bakıldığında belirli bir örüntü olmaksızın, örneklem büyüklüğünün artması ile hem tek hem de çok boyutlu MTK'nın daha az hatalı sonuçlar ürettiği görülmektedir. Araştırma bulgularına göre örneklem büyüklüğünün artması RMS değerlerinde

düşüşe neden olmuştur. Örneklem büyüklüğü sabit tutulduğunda, boyutlararası korelasyon değerlerinin değişmesi hem tek boyutlu hem de tek boyutlu MTK'ya göre yapılan kestirim hatasında azalmaya neden olmuştur. Boyutlararası korelasyonun artması ile veri yapısının tek boyutluluğa yönelmesi kestirilen hatanın düşmesine neden olmuştur. Bu bulgu, Ansley ve Forsyth (1985) yaptığı çalışmanın bulgularına paralellik göstermektedir. Tüm araştırma koşullarında da çok boyutlu MTK'nın tek boyutlu MTK'ya göre daha az hatalı kestirimler yaptığı görülmektedir. Bunun yanında örneklem büyüklüğünün artması madde parametre kestiriminde önemli düzeyde bir düşüşe neden olmuştur. Yüksek boyutlarda ve düşük örneklemelerde çok boyutlu MTK'nın madde parametre kestiriminde daha iyi sonuçlar verdiği görülmektedir.

Tek boyutlu ve çok boyutlu MTK'ya dayalı kestirilen birey parametrelerine ait standart hata kareleri ortalamaları Şekil 7'de verilmiştir.



Şekil 7. Birey parametreleri standart hata kareleri ortalamaları

Şekil 7'de görüldüğü gibi, çok boyutlu MTK'ya göre kestirilen tüm birey parametreleri için tüm deneysel koşullarda daha düşük hatalar üretmiştir. Bunun yanında boyutlararası ilişki düzeyi arttıkça tek boyutlu ile çok boyutlu MTK'ya dayalı yapılan birey parametre kestirim hataları arasındaki fark artmaktadır. Boyutlararası ilişki düzeyi sabit iken, örneklem büyüklüğünün artması, elde edilen standart hatalarda değişikliğe neden olmamaktadır. Bu bulgu Luecht (2004) ve Özer-Özkan (2012) yaptıkları araştırma bulguları ile örtüşmektedir.

Walker ve Beretvas (2003) çok boyutlu olduğu bilinen bir verinin, parametre kestirimlerini tek boyutlu MTK'ya göre yaptıklarında ne tür sonuçlar doğurduğu ile ilgili yaptıkları çalışmada, çok boyutlu verinin tek boyutlu kuramla incelenmesinin, bireyin yeteneğinin ortaya konmasında yanlış sonuçlar doğurabileceğini belirtmişlerdir. Bu sonuç, bu araştırmanın bulgusunu destekler niteliktedir.

Şekil 7'de, açık bir biçimde görüldüğü gibi birey parametre değerlerine ait hata miktarlarının çok boyutlu MTK'da daha düşük olduğu görülmektedir. Tüm araştırma koşullarında da çok boyutlu MTK'nın tek boyutlu MTK'ya göre daha az hatalı kestirimler yaptığı bulunmuştur. Buna ek olarak, örneklem büyüklüğünün artması tüm yetenek

düzeylerinde hesaplanan hata miktarını değiştirmemiştir. Bu bulgu, Çakıcı-Eser (2015) yaptığı araştırmada elde ettiği bulgu ile paraleldir.

#### 4. TARTIŞMA ve SONUÇ

Bu araştırmada, farklı boyutluluk özelliklerine sahip ikili puanlanan (1-0) veri yapılarının tek boyutlu ve çok boyutlu MTK'ya dayalı olarak incelenmiştir. Bu karşılaştırmalar, değişen koşullar altında simülasyon olarak gerçekleştirilmiştir. Bu koşullar, testin ölçtüğü boyut sayısı, boyutlararası korelasyon ve örneklem büyüklüğüdür.

Madde parametrelerine ilişkin sonuçlara bakıldığında, düşük boyut sayısında tek boyutlu ve çok boyutlu MTK'dan elde edilen standart hataların yüksek oranda farklılaşmadığı görülmüştür. Fakat boyut sayısının artması ile birlikte çok boyutlu MTK'nın madde parametresi kestiriminde daha az hatalı sonuçlar üretmiştir. Bununla birlikte örneklem büyüklüğünün artması, alan yazında da belirtildiği gibi, her iki kurama dayalı yapılan madde parametre kestirimlerinin daha kesin ve az hata ile olduğunu göstermektedir. Boyutlararası korelasyonların farklılaşması kestirilen hata miktarlarını da etkilemektedir. Tek boyutlu ve çok boyutlu MTK'ya dayalı olarak yapılan kestirimlerin boyutlararası korelasyonun artması ile daha az hata ile kestirildikleri sonucuna ulaşılmıştır. Birey parametrelerine ilişkin sonuçlar incelendiğinde, araştırma koşullarının çoğunluğunda, çok boyutlu MTK'nın daha az hata ile birey parametresi kestirdiği görülmüştür. Özellikle, beş boyutlu veri yapısı için örneklem büyüklüğü ve boyutlararası korelasyon arttıkça, elde edilen hata miktarının azaldığı görülmüştür. Araştırma koşulları dikkate alındığında örneklem büyüklüğü ve boyutlararası korelasyon yükseldikçe, hata miktarının azaldığı gözlenmektedir. Araştırılan veri yapısı çok boyutlu olduğunda örneklem büyüklüğünün kestirimlerden çok etkilendiği görülmektedir. Bu nedenle araştırmacıların, çok boyutlu veri yapılarında çok boyutlu MTK ile kestirimler yapması ve geniş örneklem büyüklüğü ile çalışmalarını önerilebilir. Çok boyutlu veri yapısında boyutlararası korelasyon değerlerinin farklılaşması tek boyutlu ve çok boyutlu kestirimleri etkilediğinden uygun yöntemin belirlenmesi için yalnızca boyutluluk analizinin değil, boyutlararası ilişkilerin de incelenmesi gereklidir. İki boyutlu veri yapıları ve boyutlararası korelasyon değerinin yüksek olduğu durumlarda tek boyutlu ve çok boyutlu MTK'dan elde edilen madde parametre değerlerine ait hatalar yüksek oranda farklılaşmadığından analizler her iki kurama dayalı olarak da yapılabilir. Veri yapısının üç ve beş boyutlu olduğu durumda tek boyutlu MTK'dan elde edilen madde parametre değerlerine ait hata miktarları artmaktadır. Bu nedenle bu tür veri yapılarında çok boyutlu modellere göre kestirim yapılması daha doğru sonuçlar üretecektir. Bu araştırmanın bulgularına göre, birey parametre değerleri tüm boyutlar için çok boyutlu yapılan analizler sonucunda daha düşük hata ürettiğinden, bireyler hakkında karar verilirken çok boyutlu modellerin kullanılması önerilebilir. Test geliştiricilerin alt boyutlar ve alt boyutlar arasındaki ilişkileri dikkate alarak kestirimler yapmaları önerilebilir. Çok boyutlu modeller bu araştırmada ele alınan tüm koşullarda birey parametre değeri kestirmede daha az hatalı sonuçlar üretmiştir. Ulusal ve uluslararası düzeyde uygulanan geniş ölçekli sınavlar, birçok alt boyuttan oluştuğundan, bu sınavlarda bireylerin yetenek kestirimleri için çok boyutlu modellerin kullanılması önerilebilir.

#### 5. KAYNAKLAR

- Ackerman, T. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, 7(4), 255-278.
- Ackerman, T. (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: MIRT Instructional Module*.
- Ansley, T. M., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement*, 9(1), 39-48.

- Babcock, B. (2009). *Estimating a Noncompensatory IRT Model Using a Modified Metropolis Algorithm*. Unpublished Doctoral Dissertation, University Of Minnesota
- Baker, F. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland College Park, MD.
- Bashkov, B. M. (2015). *Examining the Performance of the Metropolis-Hastings Robbins-Monro Algorithm in the Estimation of Multilevel Multidimensional IRT Models*. Unpublished Doctoral Dissertation, University of James Madison.
- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test teorisi ve Uygulaması*. Ankara: ÖSYM Yayınları
- Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and non-compensatory multidimensional item response models using markov chain monte carlo. *Applied Psychological Measurement*, 27, 395. <http://apm.sagepub.com> adresinden 8 Haziran 2015'te alınmıştır.
- Cai, L. (2013). *flexMIRT R version 2: Flexible multilevel multidimensional item analysis and test scoring [Computer software]*. Chapel Hill, NC: Vector Psychometric Group.
- Çakıcı-Eser, D. (2015). *Çok boyutlu madde tepki kuramının farklı modellerinden çeşitli koşullar altında kestirilen parametrelerin incelenmesi*. Yayımlanmamış Doktora Tezi. Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü.
- Crocker L., & Algina J. (1986). *Introduction to classical and modern test theory*. New York: CBS Collage Publishing,
- Eunike, W., & Benedikt, H. (2014). Multidimensional item response theory models in vocational interest measurement an illustration using the AIST-R. *Journal of Psychoeducational Assessment*, 32(4), 342-355.
- 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.
- Gibbons, R. D., Immekus, J., & Bock, R. D. (2007). *Didactic workbook: The added value of multidimensional IRT models*. National Cancer Institute Technical Report.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications Inc.
- Hambleton, R.K., & Swaminathan, H. (1989). *Item response theory: Principles and applications*. USA: Kluwer Nijhoff Publishing
- 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.
- Karasar, N. (2010). Bilimsel araştırma yöntemleri. Ankara: Nobel Yayıncılık
- 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.
- Kreiter, C. D. (1993). *An empirical investigation of compensatory and noncompensatory test items in simulated and real data*. Unpublished Doctoral Dissertation. The University of Iowa.
- Luecht, R. (2004) MIRTGEN 2.0 [Bilgisayar Yazılımı]. Yazardan alınmıştır.
- Özer-Özkan, Y. (2012). *Öğrenci başarılarının belirlenmesi sınavından (ÖBBS) klasik test kuramı, tek boyutlu ve çok boyutlu madde tepki kuramı modelleri ile kestirilen başarı puanlarının karşılaştırılması*. Yayımlanmamış Doktora Tezi. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü.
- Reckase, M. D. (2007). *Multidimensional item response theory*. C. R. Rao, S. Sinharay, (Ed), Handbook of statistics 26: psychometrics (607-642). North Holland: Elsevier B. V.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: 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.
- Toit, M. (2003). IRT from SSI. BILOG-MG MULTILOG PARSCALE TESTFACT. Scientific Software International Inc.
- Turgut, F. (1995) *Eğitimde ölçme ve değerlendirme metotları*. Ankara: Saydam Matbaacılık
- Walker, C. M. & Beretvas, S. N. (2003). Comparing multidimensional and unidimensional proficiency classifications: multidimensional irt as a diagnostic aid. *Journal of Educational Measurement*, 40 (3), 255-275.
- Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). The comparative effects of compensatory and non-compensatory two dimensional data on unidimensional IRT estimates. *Applied Psychological Measurement*, 12(3), 239-252.
- Wiberg, M. (2012). Can a multidimensional test be evaluated with unidimensional item response theory? *Educational Research and Evaluation*, 18(4), 307-320
- Yıldırım, C. (1973). *Bilim Felsefesi*. İstanbul: Remzi Kitabevi
- Yoo, H. (2011). *Evaluating several multidimensional adaptive testing procedures for diagnostic assessment*. Unpublished Doctoral Dissertation. University of Massachusetts Amherst.
- Zhang, J. (2012). Calibration of response data using MIRT models with simple and mixed structures. *Applied Psychological Measurement*, 36(5), 375-398.

### Extended Abstract

The purpose of the study is to compare item-person parameters of two categorical (1-0) multidimensional constructs that are estimated based on unidimensional and multidimensional Item Response Theory (IRT) under the following conditions: different sample sizes (500, 1000 and 2000), inter-dimensional correlation (0.10, 0.40 and 0.70) and the number of dimensions (two, three and five). The item-person parameter values are estimated according to the two model and the standard errors of these parameter values are evaluated according to the model with lower error production and with Root Mean Squared (RMS).

MIRTGEN computer software is used for data simulation in the study. The data fitted to the two categorical (1-0) complementary logistic model with three parameters in a simple construction of two, three and five dimensions through simulation. The item number is fixed ( $k=30$ ) for the data simulation. The data sets are simulated in the following fashion: a parameter from the lognormal distribution with a mean  $a$  (discrimination) values of 0.00 for the items in all dimensions and standard deviation of 1.00 and  $b$  parameter (item difficulty) from the interval values with a minimum of -1.75 and maximum of +1.75. The guessing parameter values are fixed to  $c=0.15$ . Standard normal distribution is employed for ability parameters. The data has been simulated 25 times, with changes in the number of dimensions, inter-dimensional correlation and the sample size. Thus,  $3 \times 3 \times 3 = 27$  different research conditions and  $27 \times 25 = 675$  data sets are obtained. For each data set, eligible methods are employed; BA-EM algorithm for the two and three dimensional data construct, MH-RM algorithm for the five dimensional data construct analysis with flexMIRT program.

When the RMSs, obtained from the estimation of item parameter values in the two, three and five dimensional data constructs are compared and it is concluded that multi-dimensional IRT estimates with lower errors than one dimensional IRT. Additionally, when standard errors obtained from the ability parameter estimations are considered, it is seen that multi-dimensional IRT estimates with low errors under any conditions. As a result of the study, it is concluded that multidimensional IRT has been found to give better results in the analysis of multidimensional constructs, and more accurate results, particularly in person parameter estimations and decision making, when compared to unidimensional IRT. When the number of dimensions of one dimensional IRT is low and the sample size is large, it is found to estimate better person parameter values, whereas multidimensional IRT estimates better than the former when the number of dimensions is greater than two and the sample size is more than 1000. It is seen in the analysis with unidimensional and multidimensional IRT that an increase in the relationship between the dimensions improves estimations based on the two models but multidimensional IRT gives lower errors in estimations.

When the examined data construct is multidimensional, sample size is found to be greatly affected by estimations. Therefore, it is advisable for researchers that they should estimate in multidimensional data constructs with multidimensional IRT and study with larger sample sizes. As the differentiation of inter-dimensional correlation values affects one and multi-dimensional estimations in multidimensional data constructs, it is essential to examine not only dimensionality analysis, but also inter-dimensional relationships to decide the appropriate method. Analyses could be based on the two models because errors obtained from item parameter values do not considerably vary in the two dimensional data constructs or unidimensional and multidimensional IRT when inter-dimensional correlation value is high. When the data construct has three and five dimensions, the number of errors in item parameter values obtained from unidimensional IRT increases. Hence, estimations according to multidimensional models in such data constructs will give better results. According to the research findings, multi-dimensional models are recommended in decision making since person parameter values give lower errors as a result of multi-dimensional analyses in all dimensions. It is also advisable for test developers to estimate in consideration with sub-dimensions and relationships between sub-dimensions. Multidimensional models have been found to give results with lower errors in person parameter estimation value, under all conditions in the research. Since large scale assessments at national and international levels consist of various sub-dimensions, it could be suggested using multidimensional models for ability estimations of test takers.