

## Çok Kategorili Parametrik ve Parametrik Olmayan Madde Tepki Kuramı Modellerinin Karşılaştırılması\*

### Comparison of Polytomous Parametric and Nonparametric Item Response Theory Models

Özge BIKMAZ BİLGİN \*\*

Nuri DOĞAN \*\*\*

#### Öz

Bu araştırmanın amacı çok kategorili maddeler için Parametrik Madde Tepki Kuramı (PMTK) kapsamındaki Aşamalı Tepki Modeli (ATM) ve Parametrik olmayan Madde Tepki Kuramı (PoMTK) kapsamındaki Monoton Homojenlik Modeli (MHM) ile yapılan kestirimlere örneklem büyüklüğü, örneklem dağılımı, testte yer alan madde sayısı, testte yer alan maddelerin yanıt kategorisi sayıları bağımsız değişkenlerinin etkilerini incelemektir. Bu amaca ulaşabilmek için araştırma; örneklem büyüklüğü, örneklem dağılımı, madde sayısı, maddenin kategori sayısı değişkenleri ile belirlenen 192 simülasyon koşulu desenlenen temel bir çalışma olarak gerçekleştirilmiştir. Örneklem büyüklüğü (N=100, 250, 500, 1000), örneklem dağılımı (normal dağılım, çarpıklık katsayısı -1,0 olan dağılım), madde sayısı (10, 20, 40, 80), maddenin yanıt kategorisi sayısı (3, 5, 7) koşulları için ATM ve MHM ile yapılan kestirimler sırasıyla model veri uyumları, güvenilirlik değerleri, madde parametrelerinin standart hataları hesaplanarak incelenmiştir. Araştırma sonucunda ATM’de model veri uyumu hesaplanırken değerlerin değişken artışından etkilenmesi, tek başına yorumlanamaması bu değerlerin karşılaştırılması ve genellenmesini zorlaştırmaktadır. MHM’de model veri uyumunun pratik olarak hesaplanması, başka bir değere ihtiyaç duyulmadan tek başına yorumlanması ATM’ye göre üstünlük sağlamaktadır. Diğer bir araştırma sonucu güvenilirlik değerlerinin iki model için benzer sonuç vermesidir. MHM için hesaplanan parametrelerin, küçük örneklem ve kısa test koşullarında standart hataları ATM kestirimlerine göre oldukça düşüktür ve MHM için hesaplanan parametrelerin standart hataları tüm koşullarda birbirine yakın değer almıştır.

*Anahtar Kelimeler:* Parametrik olmayan madde tepki kuramı, aşamalı tepki modeli, monoton homojenlik modeli

#### Abstract

This research aimed to identify the effects of independent variables as sample size, sample distribution, the number of items in the test, and the number of response categories of items in the test on the estimations of Graded Response Model (GRM) under Parametric Item Response Theory (PIRT) and by Monotone Homogeneity Model (MHM) under Non-Parametric Item Response Theory (NIRT) for polytomously scored items. To achieve this aim, the research was performed as a fundamental study in which 192 simulation conditions were designed by the combination of sample size, sample distribution, the number of items, and the number of categories of items. Estimates by GRM and MHM were examined under different levels of sample size (N= 100, 250, 500, 1000), sample distribution (normal, skewed), the number of items (10, 20, 40, 80), and the number of categories of items (3, 5, 7) conditions, by respectively calculating model-data fit, reliability values, standart errors of parameters. As a result of the research, it was found that since the values used to evaluate model-data fit were influenced by the increase of variable while calculating model-data fit and since they can not be interpreted alone, it is difficult to compare and generalize the results. The practical calculation of model data fit, which can be interpreted without the need for another value, in MHM provides superiority over GRM. Another research result is that the reliability values give similar results for both models. The

\*Bu çalışma, ilk yazarın, ikinci yazar danışmanlığında tamamladığı “Parametrik ve Parametrik Olmayan Madde Tepki Kuramı Modellerinin Farklı Örneklem ve Test Uzunluğunda Karşılaştırılması” isimli doktora tezinden üretilmiştir.

\*\* Arş. Gör. Dr., Adnan Menderes Üniversitesi, Eğitim Fakültesi, Aydın-Türkiye, ozgebikmaz@adu.edu.tr, ORCID ID: orcid.org/0000-0003-2219-2026

\*\*\*Doç. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: [nurid@hacettepe.edu.tr](mailto:nurid@hacettepe.edu.tr), ORCID ID: orcid.org/0000-0001-6274-2016

standard errors of the MHM parameter estimates is lower than the GRM estimates under small sample and few items conditions and the standard errors of the MHM parameter estimates are close to each other in all conditions.

*Keywords:* Nonparametric item response theory, graded response model, monotone homogeneity model

## GİRİŞ

Eğitim alanında bireyler hakkında bilgi toplamak için başarı, ilgi, tutum vb. psikolojik yapılar ölçülmektedir. Psikolojik yapıların doğrudan gözlenememesi ve değişkenlik göstermesi nedeniyle ölçme alanında test kuramları geliştirilmiştir (Hambleton & Jones, 1993). Bu kuramlardan Madde Tepki Kuramı (MTK) günümüzde, araştırmalarda sıklıkla kullanılan bir test kuramıdır (De Ayala, 2009; Ostini & Neing, 2006).

Madde Tepki Kuramı (MTK); Klasik Test Kuramı (KTK)'nın bir uzantısı olmakla birlikte genellikle KTK'ya kıyasla daha modern ve üstün bir alternatif olarak sunulmaktadır (Embretson & Reise, 2000; Nering & Ostini, 2010). KTK'dan farklı olarak MTK örneklemeden bağımsız madde ölçeklemesini ve maddelerden bağımsız yetenek kestirimini matematiksel modellerle olanaklı hâle getirme iddiasındadır (De Ayala, 2009).

MTK'da bireylerin testteki maddelere verdikleri yanıtın altında o test ile ölçülmeye çalışılan yetenekleri ( $\theta$ ) olduğu kabul edilmektedir. Yetenek gözlenemeyen bir özelliktir. MTK'da gözlenemeyen özelliğin ancak gözlenebilir olan teste yanıt verme davranışı ile ölçülebildiği düşünülmektedir. Yani bireyin yetenek düzeyi, bireyin maddelere verdiği yanıtlar ile kestirilmeye çalışılır. Bu kestirimde matematiksel fonksiyonlardan yararlanır. Bu doğrultuda bireylerin yeteneği ile testteki maddelere verdikleri yanıtlar arasındaki ilişki incelenmektedir. Bu ilişki, bireyin belli bir maddeye tepki verme olasılığı ile bireylerin test maddelerinin altında yatan yetenek üzerindeki konumu arasında matematiksel bir bağlantı olan matematiksel fonksiyonlarla kurulmaktadır (Ostini & Neing, 2006).

MTK'da bireyin gözlenemeyen özelliği ile gözlenen özelliği arasındaki matematiksel fonksiyonu veren bir madde karakteristik fonksiyonu (MKF) ve fonksiyona dayalı olarak elde edilen madde karakteristik eğrisi (MKE) bulunmaktadır. MKE madde puanının yetenek üzerindeki regresyonudur ve MKE'nin şekli, yetenek düzeyindeki değişim ile doğru cevaplama olasılığı arasındaki ilişkiyi ortaya koymaktadır. MKE madde parametrelerine bağlıdır ki elde edilen parametreler eğrinin şeklini belirlemektedir (Doğan, 2002). MKE eğrisinin şekline dayalı olarak MTK modelleri parametrik ve parametrik olmayan modeller olarak iki ana kategoride incelenmektedir (Sijtsma ve Molenaar, 2002). Parametrik MTK modellerinde MKE'nin normal veya lojistik olması beklenirken, parametrik olmayan MTK modellerinde eğrinin şekli belirli bir formla sınırlı değildir (Sijtsma, 2005).

Alan yazın incelendiğinde MTK çerçevesinde yapılan çalışmalarda daha çok parametrik modeller (PMTK) üzerinde odaklanıldığı görülmektedir (Molenaar, 2001). Bu durumun gerekçeleri arasında parametrik modellerin daha önce önerilmesi, kestirimlerde kullanılan istatistiksel yazılımlarının daha fazla olması sayılabilir (Sijtsma, Emon, Bouwmeester, Nyklicek ve Roorda, 2008; Junker & Sijtsma, 2001).

MTK temel birtakım varsayımlara sahiptir. Temel varsayımlardan biri maddenin tek bir değişkeni ölçmesi (Hambleton ve Swaminathan, 1985), bir testteki tüm maddelerin tek bir özellikte açıklanması (Sijtsma, 2005) anlamına gelen tek boyutluluktur. MTK'nın diğer bir varsayımı yerel bağımsızlıktır. Varsayım bireylerin maddeyi yanıtlarken verdikleri tepkilerin birbirinden bağımsız olduğunu temel almaktadır. Kuramlardan elde edilen sonuçların nitelikli olması kuramların varsayımlarının karşılanmasına bağlıdır. MTK çalışmalarında tekboyutluluk, yerel bağımsızlık varsayımlarının karşılanmasının yanısıra parametrik modellerle çalışılıyorsa PMTK ile çalışma sonuçlarının nitelikli olması için birtakım gerekliliklerin olduğu belirtilmektedir (Molenaar, 2001).

Bu gerekliliklerden biri PMTK ile çalışmada geniş örnekleme ihtiyaç duyulmasıdır (Demars, 2010). PMTK ile geniş örneklem ile çalışılmasının önerilmesi MTK'nın erken dönem çalışmalarına dek

uzanmaktadır (Hulin, Lissak & Drasgow, 1982; Thissen & Wainer, 1982). Geniş birey kitlesiyle çalışan araştırma sonuçlarının beklenen düzeyde olması (Zenisk, Hambleton & Sireci, 2002) diğer yandan küçük örneklem üzerinden elde edilen sonuçların istenen nitelikte olmaması (De Ayala, 2009) araştırmacıların önerilerini büyük örneklemin gerekliliği noktasında birleştirmektedir. Parametrik modellerle MTK çalışmasında iyi sonuçlar elde edilmesi için önerilen bir başka koşul geniş madde havuzu kullanımımızdır (Demars, 2010). Çalışmanın güvenilirliğinin ve geçerliğinin istenen nitelikte olmasında madde sayısı artırılmasının olumlu etkiye sahip olduğu bilinmektedir. Madde sayısı değişiminin sonuçların doğruluğu üzerindeki etkisinin incelendiği çalışmalarda daha çok maddeli testlerle çalışmanın sonuçların doğruluğunu arttırdığına (ölçmenin standart hatasının düştüğüne) yönelik bulgular elde edilmiştir (Ankenmann & Stone, 1992; Hulin, Lissak & Drasgow, 1982). Parametrik modellerle MTK çalışmalarında karşılanması beklenen diğer bir nokta, madde yapısının kullanılan modelle uyum sağlaması gerekliliğidir. Parametrik MTK modelleri genel olarak iki ve çok kategorili modeller olarak iki ana başlık altında incelenmektedir (Thissen & Steinberg, 1986). Maddenin yanıt kategori sayısının değişimi farklı model kullanımını gerektirmektedir. Çok kategorili yanıtlanan maddelerin iki kategorili yanıtlananlara göre daha kapsamlı bilgi verdiği düşünülerek (Ostini & Neing, 2006) çalışmalarda sıklıkla beş kategorili maddeler kullanılmaktadır. Yapılan çalışmalar model veri uyumu, güvenilirlik gibi niteliklerin sağlanmasında kategori sayısının etkili olduğuna işaret etmektedir (Zenisky, Hambleton & Sireci, 2002). Belirtilen PMTK'ya ait güçlü varsayımların ve zahmetli gerekliliklerin olmasının yarattığı problemlerin çözümünde PoMTK modelleri, parametrik modellere alternatif olarak sunulmuştur (Mokken, 1971; Molenaar, 2001; Sijtsma, 1998; Stout, 1987).

PoMTK modellerinde tek boyutluluk parametrik modellerde olduğu gibi temel varsayımdır. Bunlara ek olarak PoMTK modelleri çalışmalarında önemli bir yeri olan diğer bir varsayım monotonluktur. Monotonluk PMTK için de geçerli olan MKE veya çok kategorili yanıtlanan maddeler için madde adımı yanıt eğrisinin monoton azalmayan bir yapı sergilemesi anlamına gelmektedir. PoMTK modellerini, PMTK'dan ayıran nokta MKE'lerin monoton artışının (azalmayışının) belli şekle dayanmamasıdır. Yani PMTK'nın lojistik ya da normal eğri şeklinde yapı göstermesi beklenirken bu modelde monoton azalmayan yapının belli bir şekle bağlı olması beklenmemektedir (Sijtsma & Molenaar, 2002). Hatta eğrinin şekli oldukça düzensiz, kesikli olabilmektedir. Yani  $P(\theta)$  ve  $\theta$  arasındaki ilişki monoton olarak azalmayan ise herhangi bir fonksiyonel formda olması kabul edilebilir.

Junker ve Sijtsma (2001) PoMTK modellerinin PMTK modellerinin daha iyi anlaşılmasını sağladığı, parametrik modellerin zayıf model-veri uyumu sağladığı durumlarda model veri uyumu sağlamada daha esnek bir çerçeve sunduğu, daha az sayıda madde ve birey ile çalışıldığı durumlarda model veri uyumu, güvenilirlik, geçerlik konularında daha iyi sonuç sağladığı gerekçeleriyle daha kullanışlı olduğunu öne sürmüştür (Emons, 2008). PoMTK modelleri daha esnek varsayımlara sahiptir (Molenaar, 2001; Emons, 2008). Araştırmacılar bu modellerin en önemli avantajının belli bir madde yanıt fonksiyonuna (lojistik ya da normal gibi) gerek duyulmaması olduğu noktasında birleşmektedir (Sijtsma, 2005). Bu yönüyle modeller madde yanıt fonksiyonun maddeye özgü olmasını sağlayarak madde yapısının özgün haliyle incelenmesine fırsat tanımaktadır. Oysa PMTK'da maddenin modelle birlikte belirlenen madde karakteristik eğrisine (örneğin lojistik) uyup uymadığına bakılarak kestirim yapılmaktadır. Bu yönüyle PoMTK'nın keşfedici; PMTK'nın ise doğrulayıcı bir yapı sergilediği belirtilmektedir (Junker ve Sijtsma, 2001). Parametrik ve parametrik olmayan madde tepki kuramı modelleri sırasıyla açıklanmıştır.

### ***Çok Kategorili Maddeler İçin Parametrik Madde Tepki Kuramı Modelleri***

Parametrik modeller kendi içinde iki ve çok kategorili maddeler için geliştirilen modeller olarak ikiye ayrılmaktadır. İki kategorili modellerde, içerdikleri parametre sayısına bağlı olarak adlandırılma yoluna gidilmektedir. Alan yazında 1 parametrelili model, 2 parametrelili model, 3 parametrelili ve 4 parametrelili modellerin olduğu görülmektedir (Embretson & Reise; 2000). Çok kategorili yanıtlanan maddelerin iki kategorili yanıtlananlara göre daha kapsamlı bilgi verdiği düşünülerek (Ostini & Neing, 2006) çalışmalarda sıklıkla ikiden fazla kategorili maddeler

kullanılmaktadır. Böyle maddelerin analizinde iki kategorili modellerin uzantısı olarak geliştirilen modellerden yararlanılmaktadır. Kısmi puan modeli, dereceleme ölçeği, nominal tepki ve aşamalı tepki modelleri çok kategorili maddeler için geliştirilen modellerdir (De ayala, 2009). Çok kategorili modeller arasında sıklıkla kullanılan Likert ölçeği gibi sıralı kategorilere sahip çok kategorili maddeler için uygun parametrik bir model Aşamalı Tepki Modeli (ATM)'dir (Hemker, Sijtsma, Molenaar ve Junker, 1997). ATM, Samejima (1969) tarafından önerilen, iki kategorili maddeler için uygulanabilen 2PLM'nin Likert tipi ölçek gibi çok kategorili verilere genişletilmiş halidir (De Ayala, 2009).

Çok kategorili maddelerde bireylerin yetenek düzeyi ile maddelere verdikleri yanıtlar arasındaki ilişki Madde Adımı Yanıt Fonksiyonu (MAYF) ile incelenmektedir. MAYF, iki kategorili maddelerdeki madde karakteristik fonksiyonu yerine kullanılmaktadır. Şans parametresinin sıfır olduğu varsayımına dayanan ATM ile ayırıcılık ve eşik olarak iki tür parameter kestirilmektedir. Kestirimde, bütün maddeler için bir ayırt edicilik ve kategori sayısının bir eksiği kadar eşik parametresi elde edilmektedir. "a" ile sembolize edilen maddelerin ayırıcılık parametresidir. Ayırt edicilik MAYF'nin doğru yanıtlama olasılığının 0,5 olduğu noktadaki eğimine karşılık gelmektedir. Bu bağlamda a parametresinin "eğim parametresi" olarak da geçtiği kaynaklar mevcuttur. Eğri ne kadar dikse, madde o kadar ayırıcıdır. Teorik olarak a parametresi  $\pm$  sonsuz arasında değer almakta, uygulamada ise sıklıkla -2,8 ile 2,8 arasında kestirilmektedir. ATM ile kestirilen ikinci parametre "b" ile sembolize edilen eşik parametresidir. MTK'da eşik, madde adımı yanıt fonksiyonunda maksimum eğime sahip olduğu nokta olarak tanımlanmaktadır. Yani eğrinin büküm noktasının izdüşümündeki yetenek değeridir. Bu parametre maddenin en iyi ölçtüğü yetenek düzeyini göstermektedir. Eşik değerleri -3 ile +3 arasında değişmektedir, -3'e yakın değerler kolay; +3'e yakın değerler ise zor maddeleri temsil etmektedir (De Ayala, 2009; Demars, 2010).

### **Çok Kategorili Maddeler İçin Parametrik Olmayan Madde Tepki Kuramı Modelleri**

Parametrik olmayan modeller de parametrik modeller gibi kendi içinde iki ve çok kategorili maddeler için geliştirilen modeller olarak ikiye ayrılmaktadır. İki kategorili modeller Monoton Homojenlik Modeli ve İkili Monotonluk Modeli'dir. Çok kategorili maddeler için Oranlı Eğri Modeli, Kernel Smoothing Modeli ve Monoton Homojenlik Modeli modelleri önerilmiştir. Bu modellerden Monoton Homojenlik Modeli parametrik ATM'nin parametrik olmayan uzantısıdır (Sijtsma ve Molenaar, 2002). MHM, eğitim alanında sıklıkla kullanılan Likert türü ölçeklere uygundur.

MHM'de bütün maddeler için madde ayırt ediciliğini veren bir ölçeklenebilirlik (H) katsayısı ve kategori sayısının bir eksiği kadar güçlük değeri kestirilmektedir. H katsayısı teorik olarak negatif değer de almaktadır; ancak pozitif H'ye sahip olan tüm maddeler kabul edilebilirdir. Uygulamada düşük düzeyde pozitif H'ye sahip olanlar örneğin ile 0,0 ve 0,3 gibi pozitif ancak düşük ayırıcılığa sahiptir ki bu durum testin çok geçerli olmadığını düşündürülebilir (Sijtsma ve Molenaar, 2002). Diğer yandan bireylerin yetenek üzerinde güvenilir şekilde sıralanmasında çok küçük etkiye sahip olduğu söylenebilir. Mokken (1971) H için değer aralığı önermiştir: maddenin H katsayısı 0,30 ile 0,40 arasında ise zayıf; 0,40 ile 0,50 arasında ise orta; 0,50 üzerindeyse güçlü ayırıcı olduğu söylenebilir. Madde güçlüğü bireylerin maddelere doğru yanıt verme olasılıklarını ifade etmektedir (Sijtsma vd, 2008). Testin geneli için hesaplanan H katsayısı ise model veri uyumunun incelenmesinde kullanılmakta, maddeler için hesaplanan H katsayısıyla benzer şekilde yorumlanmaktadır. PoMTK'daki güçlük (P), PMTK'daki b parametresinden farklı yorumlanmaktadır. Daha düşük değerli maddeler daha zor, yüksek değerli maddeler ise daha kolaydır. 1'e yaklaşan maddeler kolayken, 0'a doğru gidildikçe maddeler zorlaşmaktadır. Madde kategorilerinin olasılıklarının kategori sırası arttıkça giderek azalması beklenmektedir. Örneğin üç kategorili bir maddede birinci P değerinin ikinciden daha yüksek olması istenen bir durumdur.

PoMTK daha esnek varsayımına sahip olması ve keşfedici yapısıyla PMTK'nın güçlü varsayımlarının karşılanamadığı durumlar için alternatif olduğu belirtilmektedir. Aynı koşullarda PMTK ve PoMTK modellerinden elde edilen sonuçların incelendiği çalışmalara ihtiyaç duyulmaktadır. Parametrik

olmayan MHM, parametrik ATM'nin genel bir hali ve uzantısı olmasıyla iki modelden elde edilen sonuçlar kıyaslanabilir.

### **Araştırmanın Amacı**

Bu araştırmanın bir amacı parametrik ve parametrik olmayan MTK modeliyle yapılan kestirimlere örneklem büyüklüğü, örneklem dağılımı, testteki madde sayısı, testte yer alan maddelerin yanıt kategori sayıları değişkenlerinin etkilerini incelemektir. Diğer bir amaç iki modelden elde edilen bulguların model-veri uyumu hata, güvenilirlik değerlerini karşılaştırarak ileride yapılacak araştırmalar için model seçimi konusunda bir öneri sunmaktır.

Buradan hareketle araştırmada “Aşamalı Tepki Modeli (PMTK) ve Monoton Homojenlik Modeliyle (PoMTK) yapılan kestirimlerde model veri uyumları, güvenilirlik değerleri, madde parametrelerinin hataları farklı örneklem dağılımlarında (normal, -1,0 düzeyinde çarpık), farklı örneklem büyüklüklerinde (N=100, 250, 500, 1000), farklı test uzunluklarında (k=10, 20, 40, 80), ve kategori sayısı farklı olan maddeler söz konusu olduğunda (x=3, 5, 7) nasıldır?” sorusuna yanıt aranmıştır. İlgili problem çerçevesinde iki alt problem oluşturulmuştur.

- 1) Örneklem dağılımı normal olduğunda Aşamalı Tepki Modeli ve Monoton Homojenlik Modeliyle yapılan kestirimlerde model veri uyumları, güvenilirlik değerleri, madde parametrelerinin hataları farklı örneklem büyüklükleri, farklı test uzunlukları ve kategori sayısı farklı olan maddeler söz konusu olduğunda nasıldır?
- 2) Örneklem dağılımı -1,0 düzeyinde çarpık olduğunda Aşamalı Tepki Modeli ve Monoton Homojenlik Modeliyle yapılan kestirimlerde model veri uyumları, güvenilirlik değerleri, madde parametrelerinin hataları farklı örneklem büyüklükleri farklı test uzunlukları ve kategori sayısı farklı olan maddeler söz konusu olduğunda nasıldır?

### **YÖNTEM**

Bu araştırma parametrik ve parametrik olmayan madde tepki kuramı kapsamında modellere ilişkin model veri uyumu, güvenilirlik, madde parametrelerinin hatalarıyla benzerlik ve farklılıklarının belirlenmesine, sayıltı ve sınırlılıklarının incelenmesine, hangisinin daha fazla bilgi sağladığının saptanmasına dayanmaktadır. Çalışma farklı modellere ilişkin olarak değişen koşullarda model uyumunu, güvenilirlik değerlerini, standart hataları belirlemeyi ve karşılaştırmayı amaçlaması, kuramları test etmeye dayanması bakımından temel bir araştırma niteliğindedir.

### **Veri Üretimi**

Araştırmanın verileri simülasyon tekniği ile üretilmiştir. İstatistiksel problemlere ilişkin olarak simülasyon verisi, eğitim alanındaki problemlerin çözümü için yürütülen çalışmalarda giderek artan bir önem kazanmaktadır (Davey, Nering & Thompson, 1997). Bu araştırmada kullanılan simülasyon verileri Wingen'3 bilgisayar programları aracılığıyla üretilmiştir. Wingen'3 yazılımı hem pratik olması hem de farklı tür dağılımlar (normal, uniform, çarpık/beta) ve modeller (iki kategorili, çok kategorili, parametrik olmayan gibi) için simülasyon verisi üretmeye imkan tanınmasıyla (Han, 2007; Han ve Hambleton, 2007) tercih edilmiştir.

Araştırmada, eğitim alanındaki çalışmalarda sıklıkla kullanılan veri türü çok kategorilidir. Çok kategorili yanıtlanan maddelerin iki kategorili yanıtlananlara göre daha kapsamlı bilgi verdiği düşünülmektedir (Ostini & Neing, 2006). Eğitim alanında sıklıkla kullanımından yola çıkarak bu araştırma kapsamında çok kategorili veriler tercih edilmiştir. Verilerin analizinde parametrik ATM ve parametrik olmayan MHM kullanılmıştır. MHM, ATM'nin parametrik olmayan karşılığıdır (Sjitsma ve Molenaar, 2002). Dolayısıyla üretilen veri seti iki modelin de kökeni ortak olduğu için iki model de analize de uygundur.

Parametrelerin üretilmesinde ayırt edicilik ve eşik parametreleri için ayrı dağılım özellikleri dikkate alınmıştır. Ayırt edicilik parametrelerinin üretilmesinde Log-N $\sim$ (0,5; 0,4) dağılımdan, güçlük parametrelerinin üretilmesinde ise N $\sim$ (0; 1) standart normal dağılımdan yararlanılmıştır. Ayırt edicilik parametrelerinin ortalaması (1,6403) ve güçlük parametrelerinin ortalaması (-0,117) eşit ve tüm koşullarda sabit olarak belirlenmiştir. Yetenek ( $\theta$ ) parametreleri N $\sim$ (0; 1) standart normal dağılım; -1,0 düzeyinde çarpık dağılım referans alınarak üretilmiştir.

Araştırmada simülasyon çalışmasının aşamaları uygulanmıştır. Simülasyon çalışmasında sırasıyla araştırmacının amacı belirlenmiş, amaca uygun problemler tanımlanmış, araştırma koşulları düzenlenmiş (bağımsız değişkenler), veri üretilmiş (3, 5, 7 yanıt kategorili madde), madde yanıtlarıyla madde parametreleri, hataları, güvenilirlik ve model-uyum değerleri elde edilmiştir. Tablo 1'de verilen araştırma desenine uygun olacak şekilde veri üretilmiştir. Araştırma sonuçları her bir koşul için 25 tekrarlı verinin analiziyle elde edilmiştir.

### ***Araştırmanın Deseni***

Çalışmada örneklem dağılımının özelliği, örneklemin büyüklüğü, madde sayısı ve maddelerin yanıt kategori sayısı araştırmanın bağımsız değişkenleri olarak ele alınmıştır. Bağımsız değişkenlere ilişkin açıklamalar sırasıyla verilmiştir.

**Örneklem büyüklüğü:** Bağımsız değişkenlerden biri örneklem büyüklüğüdür. Alan yazında PMTK için büyük örneklem önerildiği; örneklem sayısı az olduğu durumlarda alternatif olarak PoMTK modellerinin kullanılabileceği önerilmiştir. Hem küçük hem de büyük örneklemde her iki modelden kestirimlerin nasıl etkilendiğini belirlemek adına 100, 250, 500, 1000 kişi olmak üzere dört farklı örneklem büyüklüğü koşulu ele alınmıştır. PMTK ile PoMTK modelleriyle elde edilen bulgulara örneklem büyüklüğünün nasıl etkilendiğini sınamak ve araştırmanın sonucunda gerekli minimum örneklem büyüklüğünün ne kadar olması gerektiğine ilişkin sonuç elde etmek amacıyla bu örneklem büyüklükleri çalışmaya dahil edilmiştir.

**Örneklem dağılımı:** Araştırmada diğer bir bağımsız değişken örneklem dağılımının şeklidir. Normallik PMTK'nın bir varsayımı olmasına karşın araştırmalarda dağılımın normalden saptığı durumlarla sıklıkla karşılaşmakta ve bu tür durumlarda analizlere devam edildiği görülmektedir (Doğan, 2002). Bu normalden uzaklaşmanın elde edilen kestirimlerin üzerine etkisini test etmek amacıyla normal (Çarpıklık Katsayısı=0,0,) çarpık (Çarpıklık Katsayısı=-1,0), olmak üzere iki farklı örneklem dağılımları ile çalışılmıştır.

**Madde Sayısı:** Araştırmanın diğer bir bağımsız değişkeni madde sayısı değişimidir. Bu değişkenin parametre iyiliği üzerine etkilerini sınamak amacıyla kısa, orta, uzun testler olarak koşullar oluşturulmuştur. 10, 20, 40, 80 maddelik testler oluşturulmuştur. Maddelerin sayısı artırılırken daha geniş madde setinin kendinden önce gelen madde setlerini içermesi sağlanmıştır. Örneğin 20 maddelik testteki ilk 10 madde, 10 maddelik testteki değerlerle aynıdır. Benzer şekilde 20 madde 40 maddelik testin ilk 20 maddesini, 40 maddelik test 80 soruluk testin ilk 40 maddesini içermektedir. Farklı uzunluktaki testlerin ortak maddeleri içermesinin karşılaştırmaları daha anlamlı kılacağı düşünülmüştür.

**Kategori Sayısı:** Son bağımsız değişken maddelerin sahip olduğu kategori sayısıdır. Sosyal bilim çalışmalarında genellikle beş kategorili maddeler sıklıkla kullanılmaktadır. Bu sayının artması ya da azalmasının sonuçların güvenilirliği ve hatasızlığı üzerindeki etkisinin incelenmesi amaçlanarak 3, 5, 7 olmak üzere her kategori için madde üretilmiştir. 3, 5, ve 7 kategorili maddelerin için a parametreleri her koşulda birebir aynı, b parametrelerinin maddedeki kategori artmasıyla artmasına rağmen ortalamaları aynı olacak şekilde simülasyon yapılmıştır.

Üretilen gerçek yetenek ve madde parametreleri üzerinde araştırılacak olan bağımsız değişkenler ile araştırmanın deseni belirlenmiştir. Araştırma deseni Tablo 1'de özetlenmiştir.

Tablo 1. Araştırmanın Deseni

Modeller	Dağılım biçimi	Kategori sayısı	Soru sayısı	Örneklem büyüklüğü	Tekrar sayısı	Toplam
PMTK (Aşamalı Tepki Modeli)	2	3	4	4	25	2400
PoMTK( Monoton Homojenlik Modeli)	2	3	4	4	25	2400

Çalışmada parametrik ve parametrik olmayan model 2, örneklem dağılımı 2, örneklem büyüklüğü 4 ve test uzunluğu 4 koşul, maddelerin kategori sayısı 3 koşul olmak üzere ( $2 \times 2 \times 4 \times 4 \times 3 = 192$ ) veri elde edilmiştir. Diğer yandan her bir koşul için 25 tekrar yapıldığından araştırma toplamda ( $192 \times 25$ ) 4800 veri seti üzerinden yürütülmüştür.

### Verilerin Analizi

Verilerin analizi parametrik ve parametrik olmayan modeller için ayrı ayrı yürütülmüş ve elde edilen sonuçlar birleştirilerek karşılaştırmalar yapılmıştır.

Parametrik ATM'de model veri uyumunun belirlenmesi için  $-2\log$  benzerlik değerleri, güvenilirlik katsayısı, madde parametreleri ve parametrelerin standart hataları kestirilmiştir. ATM'de her bir madde için bir ayırıcılık ve kategori sayısının bir eksiği kadar eşik parametresi belirlenmektedir. İlgili parametrelere ilişkin standart hata değerleri hesaplanmıştır. İlgili hesaplamalar için MULTILOG paket programı ve Microsoft Excelden yararlanılmıştır.

Parametrik olmayan MHM'de model veri uyumunun belirlenmesi için ölçeklenebilirlik katsayısı, güvenilirlik katsayısı, madde parametreleri ve parametrelerin hataları kestirilmiştir. MHM'de her bir madde için H katsayısı (ayırıcılık) ve kategori sayısının bir eksiği kadar P (güçlük) değeri belirlenmektedir. İlgili hesaplamalar için MSP (Mokken Scale Analysis Program) ve Microsoft Excelden yararlanılmıştır.

## BULGULAR

Birinci alt problem kapsamında örneklem dağılımının normal olduğu koşulda çeşitli örneklem büyüklüğü, test uzunluğu ve farklı yanıt kategorili maddeler olduğunda Aşamalı Tepki Modeli ve Monoton Homojenlik yapılan kestirimlerde sırasıyla model veri uyumları, güvenilirlik değerleri, madde parametrelerinin hataları incelenmiştir.

### Model veri uyumuna yönelik bulgular

Model veri uyumunun değerlendirilmesinde kestirimlerin yapıldığı parametrik model için MULTILOG, parametrik olmayan model için MSP'den elde edilen değerler baz alınmıştır. MULTILOG program model veri uyumunun değerlendirilmesine ilişkin  $2 \times \log$ -benzerlik değerlerini vermektedir. Bu değerler örneklem ve parametre sayısına dayalı hesaplanması ve hesaplamaya dahil edilen parameter sayısından etkilenmesi sonuçların tek başına yorumlanmasını güçleştirmektedir (Pampel, 2000). Tablo 2'de verilen parametrik ATM model veri uyumu için  $-2 \times \log$ -benzerlik değerleri incelendiğinde madde sayısı aynı iken örneklem büyüklüğü arttıkça  $-2 \times \log$ -benzerlik değeri de beklendiği üzere artmıştır. Örneklemden bağımsız olarak yani aynı örneklem büyüklüğünde madde sayısı arttığında değer artış göstermiştir. Maddelerin kategori sayısı arttığında  $-2 \times \log$ -benzerlik değerlerinin arttığı tespit edilmiştir. Üç bağımsız değişkene ilişkin bulgu, değerlerin tek başına yorumlanmasının güçlüğüne işaret etmektedir. ATM ile model veri uyumunu incelemek için programdan elde edilen sonuçlar tek başına yorumlanamamaktadır. Bu bağlamda araştırmada ele alınan bağımsız değişkenlerin parametrik ATM model uyumuna etkisine ilişkin genelleme yoluna gidilmemiştir.

Tablo 2. Örneklem Dağılımı Normal Olduğunda Model Veri Uyumuna Ait Bulgular

Madde	Aşamalı Tepki Modeli (PMTK)				Monoton Homojenlik Model (PoMTK)		
	-2log benzerlik				H Katsayısı		
	Birey	3 kategori	5 kategori	7 kategori	3 kategori	5 kategori	7 kategori
10	100	699,2	1713,8	2225,9	0,42	0,41	0,41
	250	1448,0	3839,9	5371,8	0,42	0,41	0,40
	500	2176,7	6980,0	10433,2	0,41	0,41	0,40
	1000	3341,0	12833,0	19494,6	0,41	0,40	0,40
20	100	2187,0	4239,4	5371,0	0,42	0,41	0,40
	250	5300,9	10169,8	13078,3	0,42	0,41	0,40
	500	10053,9	19943,1	25690,3	0,40	0,40	0,40
	1000	19226,0	38917,5	50572,0	0,40	0,40	0,40
40	100	5382,7	8872,4	9800,5	0,42	0,41	0,40
	250	12983,5	21592,6	25200,2	0,42	0,41	0,40
	500	26325,0	43591,6	51800,1	0,40	0,40	0,40
	1000	51108,0	86639,7	106400,5	0,39	0,39	0,39
80	100	8472,7	14534,5	19600,0	0,42	0,41	0,40
	250	18562,7	36006,6	50462,8	0,41	0,40	0,40
	500	37855,1	74087,5	130661,0	0,40	0,40	0,40
	1000	74311,4	15282,9	212896,6	0,39	0,39	0,39

Parametrik olmayan model veri uyumunun incelenmesi ise tek başına yorumlanabilen H ölçeklenebilirlik katsayısına dayalı olarak yapılmaktadır (Sijtsma & Molenaar, 2002). Tablo 2’deki bulgular parametrik olmayan MHM için incelendiğinde, model uyumu için kullanılan ölçeklenebilirlik (H) katsayısı madde sayısı örneklem artışı ve maddenin yanıt kategorisi artışıyla istatistiksel olarak önemli ölçüde değişmemiştir. Tüm koşullarda birbirine ve 0,4’e yakın değerler elde edilmiştir. Katsayının 0,3 ve üzerinde olması kabul edilebilir model uyumu olduğunu göstermektedir (Sijtsma & Molenaar, 2002). Bu bilgiden yola çıkarak ele alınan tüm durumlarda model uyumu düşük düzeyde ve kabul edilebilirdir yorumu yapılabilir.

Model veri uyumunun değerlendirilmesinde analiz için kullanılan ATM’de MULTILOG, MHM’de MSP çıktılarında yer alan model veri uyum indeksleri temel alınmıştır. ATM için kullanılan model veri uyum katsayılarının örneklem büyüklüğünden ve analize dahil edilen parametre sayısından etkilenmesi doğrudan yorumlanmasını güçleştirmektedir. (Pampel, 2000; De Ayala, 2009). MHM sonuçları ise farklı koşullarda kararlı bir yapı sergilemiştir. H katsayısının tek başına yorumlanabilmesi farklı koşulların karşılaştırılabilmesine imkân tanımaktadır (Sijtsma & Molenaar, 2002). Bu bağlamda MHM’nin model veri uyumu kestiriminde avantajlı olduğu düşünülebilir.

#### Güvenirlilik değerlerine ilişkin bulgular

Normal dağılım koşulları için güvenirlilik katsayılarının değerlendirilmesine yönelik bulgular Tablo 3’te sunulmuştur.

Tablo 3 incelendiğinde, parametrik model için kestirilen marjinal güvenirliliğin elde edilen en düşük değeri en küçük örneklem, en az sayıda ve üç kategorili maddelerin olduğu koşulda elde edilmiştir. Madde sayısı aynı iken örneklem büyüklüğü arttıkça değerler önemli bir değişim göstermemiştir. Aynı madde sayısında ve örneklem büyüklüğünde maddenin yanıt kategorisi arttığında güvenirlilik değerleri artmıştır. Ancak bu fark Kruskal Wallis testine göre anlamlı değildir [ $\chi^2=2,33$  (2),  $p>.05$ ]. Madde sayısı değişimi diğer faktörlere göre daha belirgin değişime işaret etmiştir. Aynı örneklem büyüklüğü ve kategori sayısında madde sayısının artışı güvenirliliği artırmıştır.

Tablo 3’e göre MHM’de rho güvenirlilik değerleri virgülden sonra yüzde bir basamağı olacak şekilde hesaplanmakta dolayısıyla parametrik hesaplama sonuçlarına göre 0,01’i geçemeyecek düzeyde küçük değişiklikler içermektedir. Bu modelden kestirilen güvenirlilik değerleri madde sayısı artışıyla etkilenmekte, madde sayısı arttıkça güvenirlilik artmaktadır. Maddenin yanıt kategorisi artışı madde sayısı ve örneklem büyüklüğü aynıyken değerlerde anlamlı olmayan artışa işaret etmiş



$[\chi^2=2,25 (2), p>.05]$ , 40 madde ve üzerinde 5 ve 7 kategorili koşullar için aynı değerler elde edilmiştir.

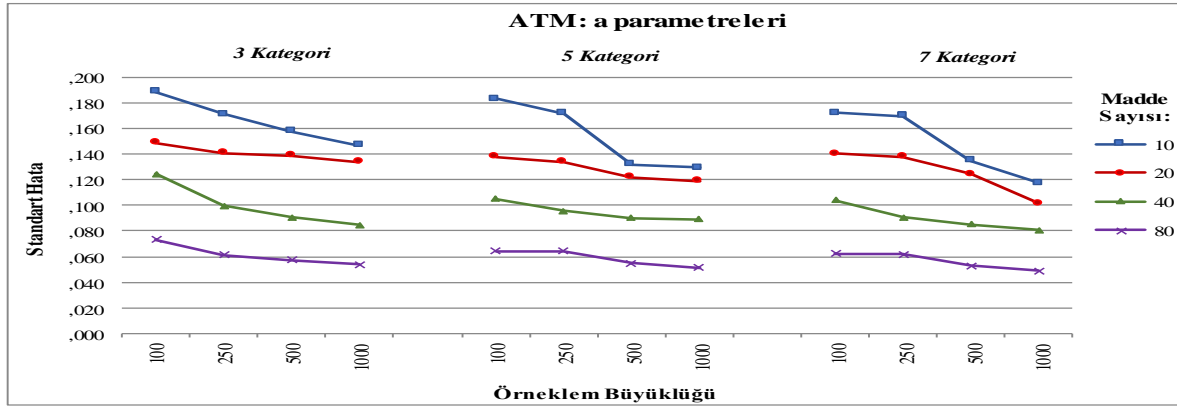
Tablo 3: Örneklem Dağılımı Normal Olduğunda Güvenirlik Değerlerine Ait Bulgular

Madde	Aşamalı Tepki Modeli (PMTK)			Monoton Homojenlik Model (PoMTK)			
	Birey	r			Rho Katsayısı		
		3 kategori	5 kategori	7 kategori	3 kategori	5 kategori	7 kategori
10	100	0,847	0,874	0,887	0,850	0,860	0,860
	250	0,845	0,873	0,885	0,850	0,860	0,860
	500	0,845	0,871	0,885	0,850	0,860	0,860
	1000	0,846	0,875	0,886	0,850	0,860	0,860
20	100	0,902	0,927	0,934	0,910	0,930	0,940
	250	0,901	0,930	0,936	0,910	0,930	0,940
	500	0,904	0,927	0,934	0,910	0,930	0,940
	1000	0,902	0,930	0,938	0,910	0,930	0,940
40	100	0,952	0,964	0,966	0,960	0,970	0,970
	250	0,955	0,964	0,966	0,960	0,970	0,970
	500	0,954	0,962	0,965	0,960	0,970	0,970
	1000	0,955	0,964	0,965	0,960	0,970	0,970
80	100	0,977	0,979	0,979	0,980	0,980	0,980
	250	0,973	0,979	0,978	0,980	0,980	0,980
	500	0,972	0,978	0,979	0,980	0,980	0,980
	1000	0,976	0,979	0,979	0,980	0,980	0,980

Parametrik MTK'da güvenilirlik madde ve test bilgi fonksiyonları açıklanmakta, test bilgi fonksiyonlarının ortalamasına dayalı hesaplanan marjinal güvenilirlik katsayısı ile ifade edilmektedir (Thissen, 1991). Ölçme sonuçlarının hatasız olması beklendiği için güvenilirliğin yüksek olması istenen durumdur. Çalışılan kuramdaki modeller güvenilirlik bağlamında paralel sonuçlar vermiştir. Bu durum araştırma koşullarının MTK varsayımlarını karşıladığı durumlarda ATM ve MHM'nin özellikle güvenilirlik değeri bakımından benzer sonuçlar üretmelerinin beklendiği bulgusuyla örtüşmektedir (Dyehouse, 2009). Güvenirliğe ilişkin elde edilen sonuçlar MHM'nin ATM'nin daha genel bir formu olması bilgisini doğrularken (Sjitsma vd., 2008), örneklem ve test koşulları parametrik bir kestirime uygun olmadığı durumlar için MHM'nin kullanılmasının önerilebileceğini göstermektedir.

#### Madde Parametrelerinin Hatalarına İlişkin Bulgular

Örneklem normal dağılım gösterdiği koşullarda ATM'den elde edilen ayırıcılık ( $a$ ) parametrelerinin standart hatalarının araştırmanın bağımsız değişken koşullarına göre değişimini veren grafik Şekil 1'de sunulmuştur.

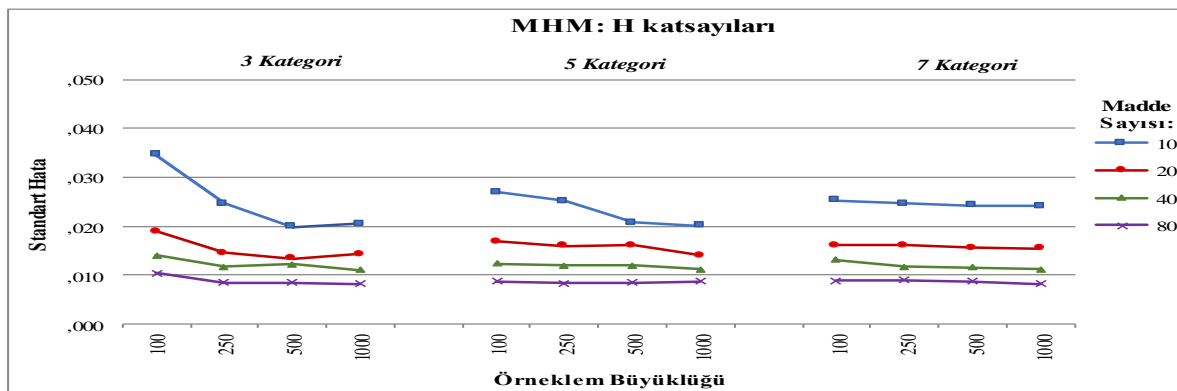


Şekil 1. Örneklem Dağılımı Normal Olduğunda Ayırıcılık Parametrelerinin Standart Hata Değerleri

Şekil 1’de verilen üç kategorili maddelerde ayırıcılık (a) parametrelerinin standart hata değerlerine ilişkin bulgular incelendiğinde, değerlerin 0,19 ile 0,05 arasında olduğu görülmektedir. Genel olarak örneklem büyüklüğü arttığında beklentiye uygun olarak hata azalmıştır. Bu azalma 10 maddeli koşullarda daha belirgindir. Madde sayısı etkisi incelendiğinde, en hatalı değerler 10 maddeli koşula aittir. Hatadaki azalma en belirgin olarak 10 maddeden 20 maddeye çıkıldığında gözlenmiştir. 40 madde ve üzerinde hem hatalar birbirine yakın hem de değerleri düşüktür. Bu bulgudan hareketle madde sayısı artışı hatada örneklem büyüklüğü artışından daha fazla etkili olmuştur yorumu yapılabilir. a parametrelerinin standart hatasının örneklem artışı ve madde sayısı artışıyla birlikte azalması diğer çalışmalarla paralellik göstermektedir (Koğar, 2015).

Beş ve yedi kategorili maddeler için standart hata bulguları üç kategorili durumla paraleldir. Ancak beş ve yedi kategorili maddelerde üç kategorili yanıtlanan maddelerden farklı olarak görece daha düşük hatalı kestirimler yapıldığı görülmektedir. Bu durum madde sayısı değişiminin mümkün olmadığı durumlarda kategori sayısı artırılmasının hatayı azaltma yönünde etkisi olacağı şeklinde yorumlanabilir.

MHM’den elde edilen ayırıcılık H katsayılarının standart hatalarının araştırmanın bağımsız değişken koşullarına göre değişimini veren grafik Şekil 2’de sunulmuştur.



Şekil 2. Örneklem Dağılımı Normal Olduğunda Ayırıcılık Parametrelerinin Standart Hata Değerleri

Şekil 2’de verilen MHM’den kestirilen H katsayılarının hataları ise 0,035 ile 0,01 arasındadır. En yüksek hata değerleri az maddeli koşullarda görülmüş, madde sayısı arttığında hata sayısı azalmıştır. Örneklem arttığında beklentiye uygun olarak hata azalmıştır. Hatadaki azalma en belirgin olarak 10 maddeden 20 maddeye çıkıldığında görülmektedir. Örneklem büyüklüğünün artışı hata değerinde

belirgin bir değişime işaret etmemiştir. Bu bulgu MHM'nin küçük örneklerde kullanılabilir olması bilgisiyle tutarlıdır (Sijtsma ve Molenaar, 2002). Bulgular maddenin yanıt kategorisi açısından yorumlandığında beş ve yedi kategorili maddelerin üç kategorili maddelere göre görece daha düşük hatalı kestirildiği ve paralel sonuçlar verdikleri görülmektedir. İki modele ait bulgular karşılaştırıldığında, MHM'nin en yüksek hata değerinin ATM'nin en düşük hata değerinin altında olması MHM kestirimlerinin daha az hatalı olduğu şeklinde yorumlanabilir. ATM'den kestirilen  $b$  parametreleri ile MHM'den kestirilen  $P$  değerlerinin standart hatalarına ilişkin bulgular Tablo 4'te özetlenmiştir.

Tablo 4. Örneklem Dağılımı Normal Olduğunda Madde Parametrelerinin Standart Hataları

		<i>Parametrik Model</i>						<i>Parametrik Olmayan Model</i>					
*	**	SH <sub>b1</sub>	SH <sub>b2</sub>	SH <sub>b3</sub>	SH <sub>b4</sub>	SH <sub>b5</sub>	SH <sub>b6</sub>	SH <sub>P1</sub>	SH <sub>P2</sub>	SH <sub>P3</sub>	SH <sub>P4</sub>	SH <sub>P5</sub>	SH <sub>P6</sub>
10	100	0,19	0,18					0,06	0,06				
	250	0,17	0,17					0,06	0,05				
	500	0,16	0,17					0,06	0,05				
	1000	0,16	0,17					0,06	0,05				
20	100	0,18	0,17					0,04	0,04				
	250	0,16	0,16					0,04	0,04				
	500	0,15	0,16					0,04	0,04				
	1000	0,15	0,16					0,04	0,04				
40	100	0,12	0,15					0,03	0,03				
	250	0,11	0,13					0,03	0,03				
	500	0,11	0,11					0,03	0,03				
	1000	0,10	0,11					0,03	0,03				
80	100	0,10	0,10					0,02	0,02				
	250	0,09	0,09					0,02	0,02				
	500	0,09	0,09					0,02	0,02				
	1000	0,08	0,09					0,02	0,02				
10	100	0,16	0,15	0,15	0,14			0,02	0,03	0,03	0,02		
	250	0,15	0,12	0,14	0,12			0,02	0,02	0,03	0,02		
	500	0,12	0,11	0,11	0,10			0,02	0,02	0,02	0,02		
	1000	0,11	0,10	0,09	0,09			0,02	0,02	0,02	0,02		
20	100	0,13	0,12	0,12	0,11			0,02	0,03	0,02	0,02		
	250	0,12	0,11	0,09	0,10			0,02	0,02	0,02	0,01		
	500	0,11	0,09	0,08	0,09			0,02	0,02	0,02	0,01		
	1000	0,10	0,07	0,07	0,08			0,02	0,02	0,02	0,01		
40	100	0,12	0,11	0,10	0,10			0,02	0,02	0,02	0,01		
	250	0,11	0,09	0,07	0,09			0,02	0,02	0,02	0,01		
	500	0,10	0,08	0,06	0,07			0,02	0,02	0,02	0,01		
	1000	0,08	0,06	0,05	0,06			0,01	0,02	0,02	0,01		
80	100	0,08	0,06	0,05	0,07			0,01	0,01	0,01	0,01		
	250	0,07	0,05	0,05	0,07			0,01	0,01	0,01	0,01		
	500	0,06	0,05	0,05	0,06			0,01	0,01	0,01	0,01		
	1000	0,07	0,05	0,05	0,06			0,01	0,01	0,01	0,01		
10	100	0,13	0,13	0,11	0,11	0,11	0,11	0,02	0,03	0,03	0,03	0,02	0,02
	250	0,11	0,11	0,09	0,10	0,10	0,11	0,02	0,03	0,02	0,02	0,02	0,02
	500	0,09	0,10	0,09	0,09	0,09	0,10	0,02	0,02	0,02	0,02	0,02	0,02
	1000	0,08	0,08	0,08	0,08	0,08	0,09	0,02	0,02	0,02	0,02	0,02	0,02
20	100	0,10	0,10	0,09	0,11	0,10	0,10	0,02	0,03	0,03	0,03	0,02	0,02
	250	0,09	0,09	0,09	0,10	0,09	0,09	0,02	0,02	0,02	0,02	0,02	0,02
	500	0,09	0,08	0,08	0,08	0,09	0,09	0,02	0,02	0,02	0,02	0,02	0,02
	1000	0,08	0,08	0,08	0,07	0,08	0,08	0,02	0,02	0,02	0,02	0,02	0,02
40	100	0,08	0,08	0,08	0,07	0,06	0,07	0,02	0,02	0,02	0,02	0,02	0,01
	250	0,07	0,07	0,08	0,06	0,05	0,06	0,02	0,02	0,02	0,02	0,01	0,01
	500	0,07	0,07	0,07	0,05	0,05	0,06	0,02	0,02	0,02	0,02	0,02	0,02
	1000	0,07	0,07	0,07	0,05	0,05	0,06	0,02	0,02	0,02	0,02	0,01	0,01
80	100	0,06	0,05	0,06	0,05	0,05	0,06	0,01	0,01	0,01	0,01	0,01	0,01
	250	0,05	0,05	0,05	0,05	0,05	0,05	0,01	0,01	0,01	0,01	0,01	0,01
	500	0,06	0,06	0,05	0,05	0,05	0,05	0,01	0,01	0,01	0,01	0,01	0,01
	1000	0,05	0,05	0,05	0,05	0,04	0,05	0,01	0,01	0,01	0,01	0,01	0,01

\*madde sayısı \*\*örneklem büyüklüğü

Tablo 4’te verilen üç kategorili maddeler için ATM ile kestirilen  $b_1$  parametresinin standart hataları 0,19 ile 0,09;  $b_2$ ’nin ise 0,18 ile 0,09 arasındadır. MHM’de kestirilen  $P_1$  ve  $P_2$  değerlerinin hatası 0,06 ile 0,02 arasında değişkenlik göstermektedir. En yüksek hatalar 10 maddeli koşulda kestirilmiş; madde sayısı arttıkça hatada düşüş tespit edilmiştir. Madde sayısı artışı hata üzerinde etkili olmuştur. Örneklem sayısı artışı da hatada küçük bir azalmaya işaret etse de bu değişim madde sayısındaki kadar belirgin değildir. İki modele ait bulgular karşılaştırıldığında, MHM’nin en yüksek hata değerinin ATM’nin en düşük hata değerinin altında olması MHM kestirimlerinin daha az hatalı kestirim yapıldığına işaret edebilir.

Tablo 4 beş kategorili maddeler için incelendiğinde, ATM ile kestirilen  $b_1$  parametrelerinin hatalarının 0,16 ile 0,06 arasında olduğu; diğer üç  $b$  parametresiyle hata düzeylerinin farklı koşullar için birbirine yakın olduğu görülmektedir. MHM’de hata değerleri 0,02 ile 0,01 arasında değer almıştır. MHM ile kestirilen parametrelerin hata düzeyleri örneklem bütüklüğü, madde sayısı ve maddenin yanıt kategorisi sayısının artışında ATM ile paralellik göstermekle birlikte görece daha düşüktür. Buradan yola çıkarak MHM kestirimlerinin daha az hatalı olduğuna işaret edilebilir.

Yedi yanıt kategorili koşullarda hata üç ve beş kategorili maddelere göre hata görece azalmıştır. En yüksek hata 10 madde ve 100 birey olduğunda kestirilmiştir. Hem birey sayısı hem de madde sayısı artışı hatada azalmaya işaret etmiştir. Yani hata değişimi büyük örneklemde ve uzun testlerde daha az gözlenmektedir. ATM’de kestirilen hatayla karşılaştırıldığında MHM kestirimlerindeki hatanın daha düşük olduğu görülmektedir.

İkinci alt problem kapsamında örneklem dağılımının -1,0 düzeyinde çarpık olduğu koşulda çeşitli örneklem büyüklüğü, test uzunluğu ve farklı yanıt kategorili maddeler olduğunda Aşamalı Tepki Modeli (PMTK) ve Monoton Homojenlik Modeliyle (PoMTK) yapılan kestirimlerde sırasıyla model veri uyumları, güvenilirlik değerleri, madde parametrelerinin hataları incelenmiştir.

#### Model veri uyumuna yönelik bulgular

Örneklem dağılımının -1,0 düzeyinde çarpık olduğu koşullar için model veri uyumlarının değerlendirilmesine yönelik bulgular Tablo 5’te sunulmuştur.

Tablo 5. Örneklem Dağılımı -1,0 düzeyinde Çarpık Olduğunda Model Veri Uyumuna Ait Bulgular

Madde	Aşamalı Tepki Modeli (PMTK)				Monoton Homojenlik Modeli (PoMTK)		
	Birey	-2*log benzerlik			H Katsayısı		
		3 kategori	5 kategori	7 kategori	3 kategori	5 kategori	7 kategori
10	100	293,4	295,14	665,54	0,17	0,18	0,19
	250	1171,4	291,09	704,32	0,18	0,19	0,18
	500	2641,6	201,52	2087,54	0,18	0,16	0,16
	1000	7037,7	1521,97	2090,21	0,18	0,16	0,16
20	100	854,4	1415,9	2232,6	0,21	0,21	0,21
	250	1817,6	2966,4	4902,8	0,19	0,17	0,19
	500	2640,0	5766,2	9771,3	0,19	0,17	0,14
	1000	58308,2	10272,5	18140,7	0,18	0,16	0,15
40	100	1762,0	3632,9	5075,0	0,23	0,19	0,16
	250	3716,5	8455,5	12222,9	0,18	0,16	0,16
	500	7649,6	17090,1	24923,3	0,17	0,15	0,15
	1000	14548	32905,7	48317,7	0,17	0,16	0,16
80	100	5519,1	7666,7	9964,7	0,22	0,17	0,21
	250	15129,5	19534,8	24397,7	0,18	0,14	0,16
	500	29933,1	37260,6	46093,8	0,18	0,16	0,16
	1000	60248,3	72733,9	89957,3	0,18	0,17	0,16

Tablo 5'te verilen parametrik modeller için model-veri uyumu için incelenen  $-2 \cdot \log$ -benzerlik değerlerinin örneklem büyüklüğü, madde ve maddenin yanıt kategori sayısı arttıkça arttığı tespit edilmiştir. Model veri uyumunun azalmasında değişkenlerin etkisine dayalı bir düşüş gerçekleşmiş olabilir; ancak  $-2 \cdot \log$ -benzerlik değerlerinin parametre değişiminden etkilenmesi tek başına yorumlanmasını güçleştirmektedir (Liu & Maydeu-Oliveres, 2014).

Parametrik olmayan MHM ile toplam puan ( $X_+$ ) üzerinden kestirilen H katsayılarının tamamı birbirine ve 0,2'ye yakın değer almıştır. Molenaar & Sijtsma (2002) kabul edilebilir uyum için alt ölçüt sınırı olarak H'nin 0,3 olmasını önermiştir. Bu bilgi doğrultusunda model uyumunun düşük olduğu görülmektedir. Böyle durumlar için MSP'nin otomatik madde seçimi ile maddelerin incelenmesine dayalı yöntemlerle model veri uyumu iyileştirilebilir (Molenaar & Sijtsma, 2002). Ancak bu çalışmada amaç parametrik modelle kıyaslama yapmak olduğundan doğrudan Mokken ölçekleme sonuçları verilmiştir. H uyum katsayısında örneklem, madde ve maddenin kategorili sayısı değişimi ATM'deki kadar etkili olmamıştır.

Bir diğer bulgu -1,0 düzeyinde çarpık dağılımdan elde edilen sonuçların normal dağılıma göre daha düşük olmasıdır. Diğer bağımsız değişkenlere göre model uyumu üzerinde dağılım şekli daha fazla etkili olmuştur. Bu bulgu, Mokken ölçeklemesinde H katsayısı hesaplanırken toplam puan ( $X_+$ ) kullanılması yani birey özelliklerinin sonuçları etkilemesine dayandırılabilir (Mokken, 1971).

#### Güvenirlilik değerlerine yönelik bulgular

Örneklem dağılımı -1,0 düzeyinde çarpık olduğu koşullar için ATM ve MHM'den hesaplanan güvenirliliklerin değerlendirilmesine yönelik bulgular Tablo 6'da sunulmuştur.

Tablo 6. Örneklem Dağılımı -1,0 Düzeyinde Çarpık Olduğunda Güvenirlilik Değerlerine Ait Bulgular

Madde	ATM (PMTK)				MHM (PoMTK)		
	Birey	r			Rho Katsayısı		
		3 kategori	5 kategori	7 kategori	3 kategori	5 kategori	7 kategori
10	100	0,553	0,599	0,596	0,57	0,75	0,77
	250	0,556	0,595	0,600	0,55	0,75	0,76
	500	0,559	0,594	0,593	0,56	0,74	0,77
	1000	0,555	0,598	0,599	0,56	0,75	0,77
20	100	0,640	0,761	0,755	0,75	0,81	0,81
	250	0,643	0,727	0,701	0,75	0,77	0,78
	500	0,640	0,694	0,679	0,75	0,77	0,78
	1000	0,643	0,682	0,686	0,75	0,77	0,78
40	100	0,766	0,784	0,827	0,88	0,87	0,86
	250	0,766	0,798	0,827	0,84	0,85	0,86
	500	0,765	0,793	0,824	0,85	0,86	0,86
	1000	0,764	0,793	0,825	0,85	0,86	0,86
80	100	0,858	0,870	0,898	0,93	0,92	0,92
	250	0,857	0,871	0,898	0,91	0,92	0,92
	500	0,857	0,874	0,899	0,92	0,92	0,92
	1000	0,858	0,877	0,893	0,92	0,92	0,92

Tablo 6'da görüldüğü gibi ATM'den kestirilen güvenirlilik değerleri üç, beş ve yedi kategorili maddeler için 0,6 ile 0,9 arasında değişkenlik göstermiştir, ancak değerlerin ondalık kesir kısımlarında farklılık mevcuttur. Beş ve yedi kategorili maddelerin güvenirlilikleri birbirine daha yakındır. Yapılan Kruskal Wallis analizi sonucuna göre üç ayrı yanıt kategorisi için değerler birbirinden anlamlı düzeyde farklı değildir [ $\chi^2(2)=2.138, p>.05$ ]. ATM'de güvenirlilik madde sayısı artışıyla yükselmiştir. Örneklem büyüklüğünün artışı katsayıda belirgin bir artış ya da azalışa işaret etmemiş; değerler birbirine yakın olarak kestirilmiştir. Normal ve çarpık dağılıma kıyasla

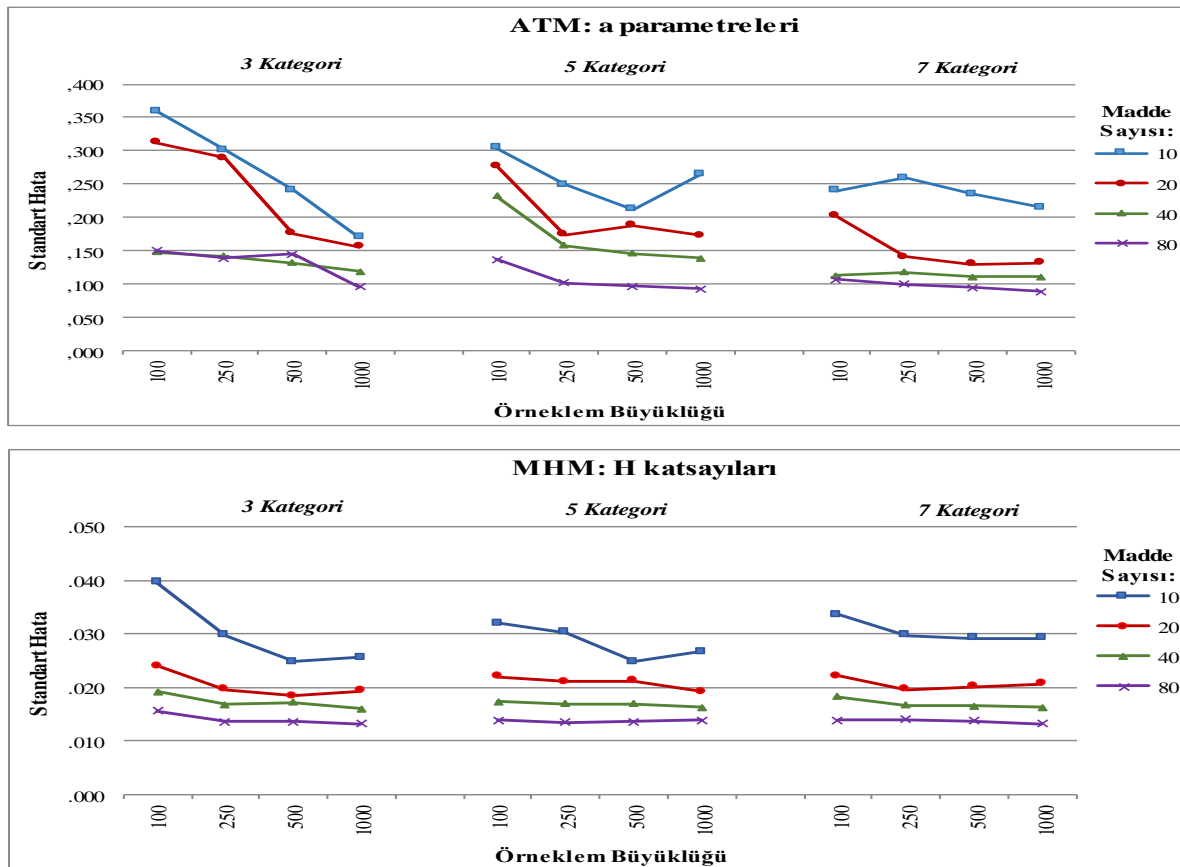
güvenirliklerin düşük olması kestirimler üzerinde dağılım şeklinin etkili olduğu şeklinde yorumlanabilir.

Tablo 6 MHM açısından incelendiğinde üç kategorili maddeler için güvenilirlik değerleri 0,6 ile 0,9; beş ve yedi kategorili maddeler için 0,7 ile 0,9 arasında elde edilmiştir. ATM’de olduğu gibi beş ve yedi kategorili maddelerin güvenilirlik bulguları birbirine daha yakındır. Yapılan Kruskal Wallis analizi sonucuna göre üç ayrı kategori değerleri birbirinden anlamlı düzeyde farklı değildir [ $\chi^2(2)=.902, p>.05$ ]. Buradan yola çıkarak güvenilirlik değerleri üzerine kategori artışının büyük bir etkiye sahip olmadığı söylenebilir. Madde sayısı değişiminin ise değerler üzerinde etkili olduğu görülmektedir. Örneklem artışı açısından değerlendirildiğinde güvenilirlik üzerinde belirgin bir etki tespit edilmemiştir. Değerlerin normal dağılıma nazaran az olması çarpık dağılımda daha hatalı kestirim yapıldığı şeklinde ifade edilebilir.

İki model bulguları birlikte incelendiğinde MHM güvenilirlik değerlerinin ATM’ye göre daha yüksek kestirildiği görülmektedir. Ancak yapılan Mann Whitney U testi sonuçlarına göre fark anlamlı değildir (üç kategorili maddeler için ( $U=108; p>.05$ ); beş kategorili maddeler için ( $U=79,5; p>.05$ ); yedi kategorili maddeler için ( $U=80; p>.05$ ). İki model arasında istatistiksel olarak anlamlı olmasa da fark olmasının nedeni ATM’nin daha sıkı kurullarla kestirim yapması; MHM’nin daha esnek olması nedeniyle hataları daha fazla tolere etmesine dayalı olabilir (Sijtsma vd, 2008).

#### Madde hatalarına ilişkin bulgular

ATM’den elde edilen  $a$  parametrelerinin ve MHM’den kestirilen  $H$  değerlerinin standart hatalarının araştırmanın bağımsız değişken koşullarına göre değişimini veren grafik Şekil 3’te sunulmuştur.



Şekil 3. Örneklem Dağılımı -1,0 Düzeyinde Çarpık Olduğunda Ayırıcılık Parametrelerinin Standart Hata Değerleri

Şekil 3'te görselleştirilen üç kategorili maddelerde  $a$  parametrelerinin standart hata değerlerine ilişkin bulgular incelendiğinde, değerlerin 0,36 ile 0,10 arasında olduğu görülmektedir. MHM'den kestirilen  $H$  katsayılarının hataları ise 0,035 ile 0,009 arasındadır. Genel olarak örneklem büyüklüğü arttığında beklentiye uygun olarak hata azalmıştır. Bu azalma düşük sayıda maddeli koşullarda daha belirgindir. Madde sayısı değişiminin etkisi incelendiğinde en yüksek değerler 10 maddeli koşula aittir. Hatadaki azalma en belirgin olarak ATM'de 20 maddeden 40 maddeye; MHM'de ise 10 maddeden 20 maddeye çıkıldığında gözlenmiştir. 40 madde ve üzerinde hem hatalar birbirine yakın hem de değerleri daha düşüktür. Bu bulgudan hareketle madde sayısı artışı hata üzerinde örneklem büyüklüğü artışından daha fazla etkili olmuştur yorumu yapılabilir. Beş ve yedi kategorili maddelerin bulguları üç maddeli durumla paraleldir. ATM ve MHM'deki üç kategorili maddelerin  $b$  parametrelerinin standart hata bulguları Tablo 7'de sunulmuştur.

Tablo 7.Örneklem Dağılımı -1,0 Düzeyinde Çarpık Olduğunda Standart Hata Değerlerine Ait Bulgular

*	**	<i>Parametrik Model</i>				<i>Parametrik Olmayan Model</i>			
		SH <sub>b1</sub>	SH <sub>b2</sub>	SH <sub>b3</sub>	SH <sub>b4</sub>	SH <sub>P1</sub>	SH <sub>P2</sub>	SH <sub>P3</sub>	SH <sub>P4</sub>
10	100	0,90	0,50			0,06	0,06		
	250	0,55	0,54			0,06	0,05		
	500	0,61	0,46			0,06	0,05		
	1000	0,36	0,29			0,06	0,05		
20	100	0,82	0,34			0,05	0,04		
	250	0,40	0,37			0,04	0,04		
	500	0,40	0,30			0,04	0,04		
	1000	0,34	0,26			0,04	0,04		
40	100	0,60	0,25			0,03	0,03		
	250	0,37	0,29			0,03	0,03		
	500	0,32	0,22			0,03	0,03		
	1000	0,21	0,18			0,03	0,03		
80	100	0,35	0,24			0,02	0,02		
	250	0,33	0,20			0,02	0,02		
	500	0,32	0,25			0,02	0,02		
	1000	0,14	0,12			0,02	0,02		
10	100	0,24	0,20	0,12	0,24	0,03	0,03	0,03	0,03
	250	0,19	0,16	0,10	0,18	0,02	0,02	0,03	0,03
	500	0,11	0,13	0,10	0,15	0,03	0,03	0,03	0,03
	1000	0,08	0,10	0,10	0,10	0,03	0,03	0,03	0,03
20	100	0,23	0,12	0,10	0,23	0,02	0,03	0,02	0,02
	250	0,13	0,12	0,06	0,17	0,02	0,03	0,02	0,02
	500	0,08	0,11	0,05	0,09	0,02	0,02	0,02	0,02
	1000	0,08	0,08	0,06	0,08	0,02	0,03	0,02	0,02
40	100	0,12	0,14	0,12	0,17	0,02	0,03	0,02	0,02
	250	0,11	0,11	0,05	0,15	0,02	0,02	0,02	0,02
	500	0,10	0,09	0,06	0,07	0,02	0,02	0,02	0,02
	1000	0,08	0,08	0,06	0,08	0,02	0,02	0,02	0,02
80	100	0,08	0,11	0,10	0,15	0,01	0,02	0,02	0,02
	250	0,07	0,10	0,10	0,12	0,01	0,02	0,02	0,02
	500	0,06	0,10	0,09	0,09	0,01	0,02	0,02	0,02
	1000	0,07	0,08	0,10	0,07	0,01	0,02	0,02	0,02

\*madde sayısı \*\*örneklem büyüklüğü

Tablo 7'de verilen üç kategorili maddeler için ATM ile kestirilen  $b_1$  parametresinin standart hataları 0,90 ile 0,14;  $b_2$ 'nin 0,50 ile 0,12 arasındadır. MHM'de hem  $P_1$  hatası 0,06 ile 0,02; hem de  $P_2$  hatası 0,06 ile 0,02 arasında değişkenlik göstermektedir. Madde sayısı artışı hata üzerinde etkili olmuştur. En yüksek hatalar 10 maddeli koşulda kestirilmiş; madde sayısı arttıkça hatada düşüş tespit edilmiştir. Örneklem büyüklüğü artışı hatada küçük bir azalmaya işaret etse de bu değişim madde sayısındaki kadar belirgin değildir. İki modele ait bulgular karşılaştırıldığında, MHM'nin en yüksek

hata değerinin ATM'nin en düşük hata değerinin altında olması MHM kestirimlerinin daha az hatalı kestirim yaptığına işaret edebilir.

Tablo 7'de beş kategorili maddeler için ATM ile kestirilen  $b_1$  parametrelerine hatalarının 0,24 ile 0,06 arasında olduğu; diğer üç  $b$  parametresiyle hata düzeylerinin farklı koşullar için birbirine yakın olduğu görülmektedir. MHM'de hata değerleri 0,03 ile 0,01 arasında değer almıştır. MHM ile kestirilen parametrelerin hata düzeyleri örneklem bütüklüğü, madde sayısı ve maddenin yanıt kategorisi sayısının artışında ATM ile paralellik göstermekle birlikte görece daha düşüktür. Buradan yol çıkarak MHM kestirimlerinin daha az hatalı olduğu söylenebilir.

Tablo 8. Örneklem Dağılımı -1,0 Düzeyinde Çarpık Olduğunda Yedi Kategorili Maddelerin Standart Hata Değerlerine Ait Bulgular

		<i>Parametrik Model</i>						<i>Parametrik Olmayan Model</i>					
*	**	SH <sub>b1</sub>	SH <sub>b2</sub>	SH <sub>b3</sub>	SH <sub>b4</sub>	SH <sub>b5</sub>	SH <sub>b6</sub>	SH <sub>P1</sub>	SH <sub>P2</sub>	SH <sub>P3</sub>	SH <sub>P4</sub>	SH <sub>P5</sub>	SH <sub>P6</sub>
10	100	0,20	0,14	0,17	0,18	0,12	0,19	0,03	0,03	0,03	0,03	0,03	0,03
	250	0,18	0,13	0,16	0,16	0,10	0,18	0,02	0,03	0,02	0,03	0,03	0,03
	500	0,15	0,11	0,13	0,11	0,09	0,15	0,02	0,02	0,02	0,03	0,03	0,02
	1000	0,10	0,08	0,10	0,08	0,09	0,12	0,02	0,02	0,02	0,03	0,03	0,02
20	100	0,20	0,09	0,12	0,15	0,10	0,19	0,02	0,03	0,03	0,03	0,03	0,02
	250	0,17	0,08	0,12	0,13	0,06	0,15	0,02	0,02	0,03	0,03	0,02	0,02
	500	0,09	0,08	0,11	0,08	0,05	0,12	0,02	0,02	0,02	0,03	0,02	0,02
	1000	0,08	0,08	0,08	0,08	0,06	0,11	0,02	0,02	0,02	0,02	0,02	0,02
40	100	0,17	0,08	0,14	0,13	0,12	0,18	0,02	0,02	0,02	0,02	0,02	0,01
	250	0,16	0,07	0,11	0,08	0,05	0,13	0,02	0,02	0,02	0,02	0,02	0,01
	500	0,07	0,08	0,09	0,07	0,06	0,06	0,02	0,02	0,02	0,02	0,02	0,02
	1000	0,08	0,08	0,08	0,08	0,06	0,05	0,02	0,02	0,02	0,02	0,02	0,02
80	100	0,15	0,06	0,08	0,10	0,10	0,07	0,01	0,01	0,02	0,02	0,02	0,01
	250	0,12	0,05	0,11	0,10	0,10	0,07	0,01	0,01	0,02	0,02	0,02	0,01
	500	0,09	0,03	0,09	0,09	0,09	0,06	0,01	0,01	0,02	0,02	0,02	0,01
	1000	0,08	0,02	0,08	0,10	0,10	0,06	0,01	0,01	0,02	0,02	0,02	0,01

\*madde sayısı \*\*örneklem büyüklüğü

Tablo 8'de görüldüğü gibi yedi yanıt kategorili koşullarda standart hata üç ve beş kategorili maddelere göre görece azalmıştır. En yüksek hata 10 madde ve 100 birey olduğunda kestirilmiştir. Hem birey sayısı hem de madde sayısı artışı hatada azalmaya işaret etmiştir. Yani hata değişimi büyük örnekleme ve uzun testlerde daha az gözlenmektedir.

## SONUÇLAR ve TARTIŞMA

Parametrik ve parametrik olmayan MTK modellerinin karşılaştırılması kapsamında parametrik olmayan modellerin PMTK modellerinin daha iyi anlaşılmasını sağladığı, parametrik modellerin zayıf model-veri uyumu sağladığı durumlarda daha esnek bir çerçeve sunduğu, daha az sayıda madde ve birey ile çalışıldığı durumlarda model veri uyumu, güvenilirlik, geçerlik konularında daha iyi sonuç sağladığı gerekçeleriyle daha kullanışlı olduğunu öne sürülmüştür (Junker ve Sijtsma, 2001). Bu araştırmada alt problemler parametrik (ATM) ve parametrik olmayan (MHM) MTK modellerinden kestirilen ait model veri uyumları, güvenilirlik değerleri, madde parametrelerinin hataları birinci alt problem örneklem dağılımı normal, ikinci alt problem örneklem dağılımı -1,0 düzeyinde çarpık olduğunda bulguların verilmesi şeklinde düzenlenmiştir. Araştırmada dağılımın çarpıklık özelliği de bir bağımsız değişken olarak ele alındığı için bu bölümde alt problemlere ilişkin sonuçlar birlikte değerlendirilerek verilmiştir.

ATM'de normal ve çarpık dağılım kendi içinde değerlendirildiğinde örneklem büyüklüğü, testin uzunluğu ve testteki maddelerin kategori sayısı artışı değişkenlerinin, teste ait model veri uyumu değerlerini artırdığı görülmektedir. Bu artış normal dağılım koşullarında çarpık dağılımlara göre daha düzenlidir. Parametrik model veri uyumu için hesaplanan  $-2 \cdot \log$ -benzerlik değerinin örneklem



ve parametre sayısına dayalı hesaplanması sonuçların tek başına yorumlanmasını güçleştirmektedir (Pampel, 2000). Bu bağlamda araştırmada ele alınan bağımsız değişkenlerin parametrik model uyumuna etkisine ilişkin genellemeye gidilmemiştir.

Teste ait model veri uyumunun değerlendirilmesine ilişkin sonuçlar incelendiğinde PoMTK'da model veri uyumu için parametrik modelden farklı olarak her kestirim için tek başına değerlendirilebilecek H katsayısının hesaplanması bir üstünlük olarak düşünülebilir (Sjitsma & Molenaar, 2002). Bu yönüyle PoMTK, bağımsız değişkenlerin etkisinin model veri uyumu üzerinde doğrudan gözlenmesine fırsat tanıyarak farklı koşullar için karşılaştırma yapma imkanı sağlamaktadır. PoMTK'da her üç dağılım koşulunda da örneklem büyüklüğü arttığında, testin uzunluğu ve testteki maddelerin kategori sayısı artışıyla birlikte, model uyumu için kullanılan ölçeklenebilirlik (H) katsayısı önemli değişim göstermemiş, tüm koşullarda bu katsayının görece birbirine yakın olduğu tespit edilmiştir. Bu durum, belirtilen değişkenlerin H katsayısının kestirim üzerinde etkili olmadığını işaret etmektedir. Dağılım normal olduğunda elde edilen H katsayılarının birbirine yakın olması ve dağılım çarpıklığı arttığında katsayıda görülen ani düşüşe dayanarak MHM'de model veri uyumunu en çok etkileyen değişkenin dağılımın çarpıklığı olduğu sonucuna ulaşılmıştır.

Testlere ait güvenilirliklerin değerlendirilmesine ilişkin sonuçlar incelendiğinde MHM ile kestirimde genel olarak madde sayısı ve maddenin yanıt kategori sayısı arttıkça güvenilirlik artmaktadır. Normal dağılım koşullarında ATM ile MHM'den paralel bulgular elde edilirken, dağılım çarpıklaştıkça MHM'nin ATM'den daha yüksek ya da daha düşük değerler verdiği görülmekte, bu bağlamda güvenilirliğin dağılım şeklinden diğer bağımsız değişkenlere göre daha fazla etkilendiği söylenebilir. Örneklem büyüklüğü değişiminin ise etkili olmadığı görülmektedir.

PMTK ile yapılan kestirimlerde parametrelerin standart hatalarının örneklem büyüklüğü, madde sayısı ve maddelerin kategori sayısı arttıkça azaldığı görülmektedir. Dağılım çarpıklaştıkça hata düzeyi de artmaktadır. Standart hata değeri ayırıcılık parametrelerinde, eşik parametrelerine göre görece daha düşüktür. Eşik parametresi kestirimlerinin a parametrelerine göre daha hatalı olduğu görülmektedir. Bu durum b parametresinin her madde için ayırıcılık değerinden daha fazla sayıda olmasından kaynaklı olabilir. PoMTK ile yapılan kestirimlerde parametrelerin standart hatalarının parametrik kestirimlere oranla görece düşük olduğu; örneklem büyüklüğü, madde sayısı ve maddelerin kategori sayısı arttıkça azaldığı görülmektedir. Dağılım çarpıklaştıkça hatalarda parametrik kestirimdeki kadar büyük artış gözlenmemektedir.

Sonuç olarak, ATM ile parametre kestirimlerinin daha güvenilir ve daha az hatalı olması dağılımın normallik özelliği göstermesi ve en az 500 örneklem büyüklüğünün sağlanmasıyla gerçekleşmektedir. Madde sayısının 20, kategori sayısının en az 5 olmasının kestirimlerde parametre değişmezliğinin sağlanmasında etkili olan faktörler olduğu sonucuna ulaşılmıştır. Dolayısıyla araştırma koşulları örneklem büyüklüğü, madde sayısı ya da madde kategori sayısı değişimine imkân vermediği durumlarda tüm koşullardan daha az hatalı ve daha kararlı yapıda kestirim sunan MHM tercih edilmesi önerilebilir.

## KAYNAKÇA

- Ankenmann, R. D., & Stone, C. A. (1992, April). *A monte carlo study of marginal maximum likelihood parameter estimates for the graded model*. Paper presented at the Annual Meeting of the Council on Measurement in Education, San Francisco, CA.
- Davey, T., Nering, M. L., & Thompson, T. (1997). *Realistic simulation of item response data* (ACT Research Report Series 97-4, July). Retrieved from: <http://files.eric.ed.gov/fulltext/ED414297.pdf>
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. NY: Guilford.
- DeMars, C. (2010). *Item response theory*. New York: Oxford University.
- Dyehouse, M. A. (2009). *A comparison of model-data fit for parametric and nonparametric item response theory models using ordinal level ratings*. (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3379330).
- Doğan, N. (2002). *Klasik test kuramı ve örtük özellikler kuramının örneklem bağlamında karşılaştırılması*. (Doktora tezi, Hacettepe Üniversitesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Ankara).

- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey, NJ: Lawrence Erlbaum Associates.
- Emons, W. H. M. (2008). Nonparametric person-fit analysis of polytomous item scores. *Applied Psychological Measurement*, 32(3), 224–247. <http://dx.doi.org/10.1177/0146621607302479>.
- Junker, B. W., & Sijtsma, K. (2001). Nonparametric item response theory in action: An overview of the special issue. *Applied Psychological Measurement*, 25, 211-220.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Academic Publishers Group.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement*, 12, 38-47. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1745-3992.1993.tb00542.x/pdf>
- Han, K. T. (2007). WinGen: Windows software that generates IRT parameters and item responses. *Applied Psychological Measurement*, 31(5), 457-459.
- Han, K. T., & Hambleton, R. K. (2007). *User's manual: WinGen* (Center for Educational Assessment Report No. 642). Amherst, MA: University of Massachusetts, School of Education.
- Hemker, B. T., Sijtsma, K., Molenaar, I. W., & Junker, B. W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psikometrika*, 62, 331-347.
- Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two and three parameter logistic item characteristic curves: A monte carlo study. *Applied Psychological Measurement*, 6, 249-260.
- Kogar, H. (2015). Madde tepki kuramına ait parametrelerin ve model uyumlarının karşılaştırılması: Bir Monte Carlo çalışması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1), 142–157.
- Liu, Y., & Maydeu-Olivares, A. (2014). Identifying the source of misfit in item response theory models. *Multivariate Behavioral Research*, 49, 354-371.
- Maydeu-Olivares, A., & Joe, H. (2005). Further empirical results on parametric vs. nonparametric IRT modeling of Likert type personality data. *Multivariate Behavioral Research*, 40, 275-293.
- Mokken, R. J. (1971). *A theory and procedure of scale analysis with applications in political research*. Berlin: Walter de Gruyter, Mouton.
- Molenaar, I. W. (2001). Thirty years of nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 295-299. Retrieved from <http://dx.doi.org/10.1177/01466210122032091>
- Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. Thousand Oaks: Sage.
- Pampel, F. C. (2000). *Logistic Regression*. Aprimer. Thousand Oaks, CA: Sage Publications.
- Sijtsma, K., Emons, W. H. M., Bouwmeester, S., Nyklicek, I., Roorda, L. D. (2008). Nonparametric IRT analysis of Quality-of-Life Scales and its application to the World Health Organization Quality-of-Life Scale (WHOQOL-Bref). *Quality of Life Research*, 17(2), 275-290.
- Sijtsma, K. (2005). Nonparametric item response theory models. *Encyclopedia of Social Measurement*, Volume 2.
- Sijtsma, K., & Molenaar, I. W. (2002). *Nonparametric item response theory and related topics*. London: Sage.
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52, 589-617. Retrieved from <http://dx.doi.org/10.1016/j.paid.2010.02.011>
- Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, 51, 567-577.
- Zenisky, R. K., Hambleton, S. G. Sireci. (2002). Identification and evaluation of local item dependencies in the Medical College Admissions Test. *Journal of Educational Measurement*, 39(4), 291-309. Retrieved from <http://dx.doi.org/10.1111/j.1745-3984.2002.tb01144.x>

## EXTENDED ABSTRACT

### Introduction

Item response theory (IRT) models are becoming more popular. Parametric item response theory (PIRT) models have been used more than nonparametric IRT models, especially in educational measurement. A reason for this may be that nonparametric IRT models were developed later than parametric IRT models. Nonparametric IRT models have several advantages over more-restrictive *parametric* IRT models. NIRT models are based on less-restrictive assumptions, thus they allow more items into the measurement tool. NIRT models offer diverse tools for item analysis that give information about the dimensionality of the data and the properties of the items; and provide item location and discrimination values, which have an easy interpretation for users. Several families of PIRT models for polytomous item scores have been proposed. The family of Graded Response Models (GRM) is suitable for analyzing ordered item scores collected by means of polytomous

response scales. And Monotone Homogeneity Model (MHM) is nonparametric extension of GRM. All known parametric IRT models for polytomous items are special cases of the nonparametric MHM. This means that any item set satisfying the requirements of a parametric IRT model for polytomous items also satisfies the requirements of the nonparametric MHM. In literature related to educational measurement, there is a need for comparison of both models in various conditions. To achieve this purpose, item response theory applications are developed.

The purpose of this study was to compare the standart errors of item parameter estimates, reliability values, and model-data fit for test under different sample size, sample distribution, number of items, and number of categories of items conditions.

### **Method**

The purpose of this study was to identify the effects of sample size, sample distribution, test length, the number of categories in items on the estimates of Graded Reponse Model and Monotone Homogeneity Model for polytomously scored items. To achieve this aim, 192 simulation conditions which is composed of sample size, sample distribution, number of items, number of category items variables were designed. Estimates for GRM and Monotone Homogeneity Model were investigated through sample size (N = 100, 250, 500, 1000), sample distribution (normal, skewed), item number (10, 20, 40, 80), item category number (3, 5, 7) by calculating model-data fit, reliability values, item parameters, standard error of parameters values. The findings had been obtained through 25 replications.

With the purpose of estimating item parameters, a and b values for GRM; H and P values for MHM had been estimated. In order to evaluate the model-data fit for the test, -2log-likelihood value for GRM; scalability coefficients for MHM had been obtained. Reliability estimates of tests according to GRM and MHM were estimated. In addition, the standart erros of item parameter estimates had been calculated.

### **Results and Discussion**

As a result of the research, it was found model-data fit values were affected by increase of variable number in Graded Response Models and since these values cannot be interpreted alone, it is difficult to make comparisons and generalization of those values. The practical calculation of model-data fit, and interpretation without need for another value in MHM provided superiority over GRM. Due to similar results in small samples and fewer items conditions and in larger samples and multiple items conditions, MHM had wider range of implementation.

It was also found that the reliability values gave similar results for both models. The increase in sample size had little effect on reliability estimates. In both models, values increased with the number of items and item response category. Reliability estimates decreased as the skewness of distribution increased.

Standard errors of item parameter estimates decreased as the number of items, number of categories of items, and sample size increased in Graded Response Model. Standart error of parameters were estimated higher at small samples and at conditions with fewer items. In Monotone Homogeneity Model, the standart error of parameters were lower than GRM at small sample and short test conditions and took close values to each other at all the conditions.

In conclusion, standart error of parameter estimates were smaller with normal distribution and larger sample sizes (with at least 500 examinees). And it was concluded that the number of items which was at least 20 and the number of categories which was at least 5 were effective factors in providing the parameter goodness in the estimates. Thus, in case the research conditions do not allow changes in sample size, the number of items, or number of categories, it can be suggested that MHM can be preferred since it provides less faulty and more stable estimations across all the conditions.