

GEFAD / GUJGEF 38(2): 665-694 (2018)

Çok Kategorili Madde Tepki Kuramı Modellerinin Örneklem Büyüklüğü Açısından İncelenmesi

Investigation of Polytomous Item Response Theory Models in Terms of Sample Size

Serap BÜYÜKKIDİK¹, Hakan Yavuz ATAR²

¹Sinop Üniversitesi, Eğitim Bilimleri Bölümü. sbuyukkidik@gmail.com

²Gazi Üniversitesi, Eğitim Bilimleri Bölümü. hakanatar@gazi.edu.tr

Makalenin Geliş Tarihi: 22.09.2017

Yayına Kabul Tarihi: 21.05.2018

ÖZ

Bu çalışmada Rosenberg Benlik Saygısı ölçeğinin madde tepki kuramındaki farklı çok kategorili modellerinde model veri uyumu kontrol edilip, madde parametreleri kestirimleri arasındaki ilişkilerin seçilen modele göre farklılaşıp farklılaşmadığı incelenmiştir. Araştırmada 47974 bireyin verdiği yanıtlar arasından kayıp veriler temizlendikten sonra rastgele seçilen Amerika Birleşik Devletli 500, 1000 ve 2000 bireyin tek boyutlu dört kategorili 10 maddelik ölçeğe verdiği yanıtlar kullanılmıştır. Genelleştirilmiş kısmi puan, 1 parametrelili lojistik model gibi sınırlandırılmış genelleştirilmiş kısmi puan, kısmi puan ve aşamalı tepki modeliyle 500, 1000 ve 2000 kişilik örneklemelerden elde edilen verilerin analizinde -2log-olabilirlik, Akaike bilgi ölçütü ve Bayesian bilgi ölçütü model veri uyum katsayıları incelendiğinde en fazla uyumun her koşulda aşamalı tepki modeli ile gerçekleştiği bulunmuştur. Her modelden elde edilen madde parametreleri arasında manidar yüksek bir ilişkinin olduğu tespit edilmiştir. Farklı modellerden elde edilen bulgulara göre her üç örneklemden en yüksek ayırt ediciliğe sahip olan maddenin 6. madde, en az ayırt ediciliğe sahip olan maddenin ise 500 kişilik örneklem için 4. madde, 1000 ve 2000 kişilik örneklem için ise genelde 8. madde olduğu görülmüştür.

Anahtar Sözcükler: Genelleştirilmiş kısmi puan modeli, Kısmi puan modeli, Aşamalı tepki modeli

ABSTRACT

In this study, we investigated whether the relations between the item parameter estimates differed according to the selected models by controlling the model datafit in the different polytomous models in the item response theory of the Rosenberg Self-Esteem scale. Among the answers given by 47974 individuals in the study, responses of randomly selected US 500, 1000 and 2000 individuals in a one-dimensional, four-categorical 10-item scale were used after the missing data were cleared. When the 2log-likelihood, AIC and BIC model data fit coefficients obtained from the data obtained from 500, 1000 and 2000 individuals into the generalized partial credit, restricted generalized partial credit such as 1-parameter logistic model, partial credit and graded

response model were examined, it was found that the most adaptation was with graded response model in every condition. It was found that there was a high correlation between the item parameters obtained from each model. According to the findings obtained from different models, it was seen that item 6 had the highest discrimination from all three samples while item 4 had the least discrimination for 500 samples and item 8 for 1000 and 2000 samples in general.

Keywords: *Generalized partial credit model, Partial credit model, Graded response model*

GİRİŞ

Psikolojik yapıların ölçülmesinde evrensel olarak kabul görmüş tek bir yaklaşım mevcut değildir (Crocker ve Algina, 1986). Bu yaklaşımlar arasında, 1904'lerden günümüze gelen klasik test kuramı (KTK), KTK'nin bir uzantısı olan ve temeli varyans analizine dayalı olan genellenebilirlik kuramı ve madde tepki kuramı (MTK) bulunmaktadır. MTK modellerinin test sonuçlarının raporlanması ve hazırlanmasında KTK'ye göre birçok avantajları vardır (Hambleton ve Swaminathan, 1985). Alanyazında MTK'nin psikometriye getirdiği yenilikleri ve KTK'ye göre avantajlarını ortaya koyan birçok çalışma ortaya konmuştur (Hambleton ve Jones, 1993; Ndalichako ve Rogers, 1997; Fan, 1998; Macdonald ve Paunonen, 2002; Magno, 2009). Madde istatistiklerinin gruba bağımlılığı, şans başarısını göz ardı etmesi, test merkezli olması, bütün yetenek ranjı için tek hata kestirimi yapması KTK'nin MTK'ye göre bazı dezavantajları ve sınırlılıklarına örnek verilebilir (Hambleton ve Swaminathan Rogers, 1991). MTK'nin KTK'ye oranla madde güçlük parametresi kestirimlerinin örneklemeden örnekleme değişmediği, madde güçlük parametrelerinin MTK'de daha istikrarlı olduğu, benzer bir durumun iç tutarlılıklar için de söz konusu olduğu ve MTK'nin KTK'ye oranla manidar düzeyde daha az ölçme hatası ürettiği bulunmuştur (Magno, 2009). Bunun yanında MTK'nin madde parametrelerinin değişmezlik özelliğini sağlamayıp; KTK'nin MTK kadar kararlı sonuçlar verdiği araştırmalar da mevcuttur (Doğan ve Tezbaşaran, 2003).

MTK'de bireyin yetenek seviyesi ölçme araçlarına verilen yanıtın matematiksel bir fonksiyon yardımıyla kestirilir (Embretson ve Reise, 2000). MTK varsayımları sağlandığı takdirde, kullandığı matematiksel model aracılığıyla örneklemeden bağımsız madde ölçeklenmesi ve maddeden bağımsız yetenek kestirimine olanak sağlar (de Ayala,

2009). Madde tepki kuramının iki ana ve birbiriyle ilişkili varsayımı uygun boyutluluk ve yerel bağımsızlık varsayımlarıdır. Çoğu MTK modeli örtük özelliğin tek boyutla ile temsil edildiğini varsayar (de Ayala, 2009). Tek boyutluluk varsayımına göre madde yanıtları arasındaki ortak varyans bir gizil yetenek ile açıklanır (Embretson ve Reise, 2000).

Bir diğer varsayım ise yerel bağımsızlık varsayımdır. Bu varsayımın göre yetenek parametresi kontrol altına alındığında maddelere verilen yanıtlar istatistiksel olarak birbirinden bağımsızdır (Embretson ve Reise, 2000). Eğer MTK modelinde yetenek parametresi yalnızca tek bir boyutta kestiriliyor ise, yerel bağımsızlık varsayımının sağlanması tek boyutluluğun da sağlandığı anlamına gelir (Embretson ve Reise, 2000).

MTK'nin yukarıda belirtilen avantajlarının sağlanabilmesi için öncelikli olarak test verilerine en uygun MTK modeli seçilmesi gerekir (Hambleton ve Swaminathan, 1985). Test verilerine uygun model seçilmediğinde madde parametreleri hatalı kestirilecektir.

Tek boyutlu MTK modelleri doğru-yanlış (1-0) şeklinde iki kategorili puanlanan verilerin modellenmesinde kullanılır ve madde parametrelerinin sayısına göre bir, iki, üç ve dört parametrelili lojistik model olarak farklılaşır. Çok kategorili puanlanan verilerde ise kısmi puan modeli, genelleştirilmiş kısmi puan modeli ve aşamalı tepki modeli gibi farklı MTK modelleri kullanılmakta ve çok kategorili modeller de sıralı ya da sıralı olmama durumuna göre farklılaşmaktadır. Bu çalışmada kullanılan kısmi puan modeli, genelleştirilmiş kısmi puan modeli ve aşamalı tepki modelleri aşağıda kısaca açıklanmıştır.

Kısmi Puan Modeli (KPM)

Kısmi puan modeli (Masters, 1982) bir parametrelili lojistik modelin (1PLM) özel bir çeşidi olan Rasch modelin genişletilmiş halidir. KPM orijinal olarak birden çok adım gerektiren test maddelerinin analizi için geliştirilmiştir. Bu yüzden KPM kısmi doğru yanıtların mümkün olduğu başarı testlerinde (örneğin, matematik problemleri) madde yanıtlarını tanımlamada katkıda bulunur (Embretson ve Reise, 2000). KPM'de madde eğimleri (ayırt edicilikleri (a)) maddeler arasında eşit kabul edilir. Yani a parametresi

her madde için aynı ve bire eşittir. Kesişim parametreleri, bir kategorideki bir yanıt önceki kategorideki bir yanıtta daha yüksek olasılığa sahip davranış düzeyini temsil eder. Kategori kesişim/eşik parametrelerinin (b) sayısı, yanıt seçeneklerinin sayısından bir eksiktir. Örneğin, dört kategorili bir ölçeğinin b parametre sayısı üçtür. Eşitlik (1)'de IPLM'de yetenek düzeyi θ olan bireyin i maddesini doğru cevaplama olasılığı, doğru cevaplama ve yanlış cevaplama olasılıklarına ayrıştırılarak farklı şekilde ifade edilmiştir. Modellerde ait denklemlerde aşağıdaki semboller kullanılmıştır.

$P_i(\theta)$ θ yetenek düzeyindeki rastgele seçilen bir bireyin madde i 'yi doğru cevaplama olasılığını,

a_i madde i 'nin ayırt edicilik parametresini,

b_i madde i 'nin eşik parametresini,

Z bireyin yetenek puanının standartlaştırılmış puanını göstermektedir.

$$P_i(\theta) = \frac{1}{1+e^{-D(\theta-b_i)}} = \frac{\exp(D(\theta-b_i))}{1+\exp(D(\theta-b_i))} = \frac{P_{i1}(\theta)}{P_{i0}(\theta)+P_{i1}(\theta)} \quad (1)$$

$P_{i1}(\theta)$ yetenek düzeyi θ olan rastgele seçilen bir bireyin madde i 'den 1 puan alma olasılığını, $P_{i0}(\theta)$ ise yetenek düzeyi θ olan rastgele seçilen bir bireyin madde i 'den 0 puan alma olasılığını belirtir. İki kategorili puanlama çok kategorili puanlamaya genelleştirilirse, yetenek düzeyi θ olan bireyin x kategorili puanlanan bir maddeden x puan alma olasılığı aşağıdaki gibi hesaplanır. (Han ve Hambleton, 2014, s. 15).

$$\frac{P_{ix}(\theta)}{P_{ix-1}(\theta)+P_{ix}(\theta)} = \frac{\exp(D(\theta-b_{ix}))}{1+\exp(D(\theta-b_{ix}))}, \quad x=1, 2, \dots, m_i, \quad (2)$$

Burada $P_{ix}(\theta)$ ve $P_{ix-1}(\theta)$ sırasıyla yetenek düzeyi θ olan bireyin x ve $x-1$ puan alma olasılıklarıdır. Eşitlik (2)'de m_i ise madde güçlük parametrelerinin sayısıdır. Çok kategorili puanlanan bir madde için, yetenek düzeyi θ olan rastgele seçilen bir bireyin i

maddesinden x puan alma olasılığı aşağıdaki gibi ifade edilebilir (Han ve Hambleton, 2014, s. 16).

$$P_{ix}(\theta) = \frac{\exp \sum_{k=0}^x (D(\theta - b_{ik}))}{\sum_h^{m_1} \exp \sum_{k=0}^h (D(\theta - b_{ik}))}, \quad x = 1, 2, \dots, m_i \quad (3)$$

Eşitlik (3)'un fonksiyonu puan kategorisi yanıt fonksiyonu olarak da adlandırılır (Han ve Hambleton, 2014, s. 16).

Genelleştirilmiş Kısmi Puan Modeli (GKPM)

Genelleştirilmiş kısmi puan modeli (Muraki, 1992) KPM'nin ayırt edicilik parametresinin modele dahil edilmesiyle genelleştirilmiş halidir.

$$P_{ix}(\theta) = \frac{\exp \sum_{k=0}^x (Z_{ik}(\theta))}{\sum_h^{m_1} \exp \sum_{k=0}^h (Z_{ik}(\theta))} \quad (4)$$

Burada

$$Z_{ik}(\theta) = Da_i (\theta - b_i + d_{ix}) \quad (5)$$

Burada d_{ix} madde i'nin puan kategorisi x'in göreceli güçlüğüdür.

$$b_{ix} = b_i - d_{ix} \quad (6)$$

Ve Eşitlik (5),

$$Z_{ik}(\theta) = Da_i (\theta - b_{ix}) \quad (7)$$

KPM ve GKPM arasındaki tek fark her bir madde için ayırt edicilik parametresinin (a_i) dahil edilmesidir (Han ve Hambleton, 2014). Masters (1982) GKPM formülünde eğim parametresi a , 1'e sabitlendiğinde ve bir eşik parametreleri (τ_{ig}) her bir madde için ayrı ayrı tahmin edildiğinde, bu parametrelerin θ üzerinde sıralanmasına gerek olmayan ek koşullarla birlikte, bu eşitliğin KPM için puanlama fonksiyonu formülüne dönüştüğünü belirtmiştir.

Aşamalı Tepki Modeli (ATM)

Aşamalı tepki modeli sıralı kategorik yanıtların olduğu durumda kullanılması uygun olan iki parametrelili lojistik modelin bir uzantısıdır (Embretson ve Reise, 2000). ATM’de her madde bir eğim parametresi (a) ve kategoriler arasında yanıt kategori sayısından bir eksik eşik parametresi (b) tarafından tanımlanmaktadır (Hays, Morales ve Reise, 2000).

Aşamalı tepki modeli Samejima (1969, 1972, 1995) tarafından genellikle likert tipi ölçeklerde ve öğrenci performansının değerlendirilmesinde kullanılan harf puanlama A, B, C, D, F gibi sıralı çoklu kategorileri (Samejima, 1996) ele almak için ortaya konulmuştur. Temelleri Thurstone’un birikimli sınır ölçümüne dayanmaktadır (Ostini ve Nering, 2006). Bu model aşağıdaki gibi ifade edilir.

$$P_{ix}^*(\theta) = \frac{\exp(Da_i(\theta - b_{ix}))}{1 + \exp(Da_i(\theta - b_{ik}))} \quad (8)$$

Eşitlik(8)’de, $P_{ix}^*(\theta)$, yetenek düzeyi θ olan rastgele seçilen bireyin i maddesinde x ya da daha fazla puan alma olasılığıdır. Bu fonksiyon birikimlikategori yanıt fonksiyonu (the cumulative category response function(BKYF)) olarak adlandırılır. Her bir puan kategorisi için olasılık

$$P_{ix}(\theta) = P_{ix}^*(\theta) - P_{ix+1}^*(\theta) \quad (9)$$

denklemleriyle hesaplanır. Bu yüzden, ATM’nin puan kategorisi yanıt fonksiyonu (score category response function (PKYF)) aşağıdaki eşitlik (10)’daki gibi de ifade edilebilir (Han ve Hambleton, 2014, s. 18).

$$P_{ix}(\theta) = \frac{\exp[-Da_i(\theta - b_{ix+1})] - \exp[-Da_i(\theta - b_{ix})]}{[1 + \exp[-Da_i(\theta - b_{ix})]][1 + \exp[-Da_i(\theta - b_{ix+1})]]} \quad (10)$$

ATM’nin madde parametrelerinin yorumlanması KPM ve GKPM’nin aksine, PKYF’ye değil, BKYF’ye dayalıdır. ATM’de bir eğim (a) parametresi ve puan kategorisinin bir eksiği kadar eşik parametresi(b_{ix}) kestirilir. Eşik değeri (b_{ix}) ise, %50 olasılıkla x eşik değerinin üzerinde puan almak için gerekli yetenek düzeyi olarak yorumlanır (Embretson ve Reise, 2000).

ATM'de her bir madde için kategoriler arası eşik (b) parametresi sıralıdır. Bu sıralı b parametresi özelliği kısmi kredi ya da genelleştirilmiş kısmi kredi model için gereklilik değildir (Embretson ve Reise, 2000). Kısmi puan modelinin yanıt kategorilerin aralık sırası hakkında varsayımı yoktur (Hays, Morales ve Reise, 2000).

Alanyazında çok kategorili puanlanan verilerle yapılan MTK çalışmaları incelendiğinde çalışmaların daha çok örneklem büyüklüğü, madde sayısı ve puan kategori sayılarının farklılaştığı görülmektedir. Thissen, Pommerich, Billeaud, ve Williams, (1995) gerçek verilerde çok kategorili modellerden ATM ile madde ve yetenek parametresi kestirimleri yapmışlardır. Maydeu-Olivares, Drasgow ve Mead (1996) KPM ve ATM ile analizleri likert tipi simülatif verilerde gerçekleştirmişlerdir. Farklı örneklem büyüklüklerini ($N= 250, 500, 1000$ ve 3000) de ele aldıkları araştırmalarında 5 seçenekli 5 ile 25 madde arasında değişen ölçek uzunluğunda modeller çok benzer sonuçlar vermiştir ve her iki modelin de neredeyse eşit derecede uygun olduğu bulunmuştur.

Cook, Dodd ve Fitzpatrick (1999) araştırmalarında madde takımı puanı bağlamında KPM, GKPM ve ATM'yi karşılaştırmıştır. SAT I ($N= 2.548$) sonbahar 1994 uygulamasından elde edilen verilerden ve simülasyon verilerden yararlanarak. Yetenek kestirimleri, bilgi ve model uyumu analiz edilmiştir. Yetenek kestirimleri arasındaki korelasyonlar 0.9748 ile 0.9921 arasında bulunmuştur. KPM, GKPM ve ATM'nin bilgi fonksiyonları arasındaki ilişki, son iki model için ayırt edicilik parametresi kestirimlerini yansıtmıştır.

Kang ve Chen (2008) araştırmalarında GKPM, KPM ve derecelendirme ölçeği modeli de dahil olmak üzere $S-X^2$ 'nin çok kategorili MTK modellerine olan katkılarını ele almışlardır. Genelleştirilmiş $S-X^2$ 'nin madde modeli uyumu değerlendirmesindeki performansı ampirik Tip I hata oranları ve gücü açısından incelenmiş ve G^2 ile karşılaştırılmıştır. Sonuçlar, genelleştirilmiş $S-X^2$ 'nin eğitim ve psikolojik testlerde çok kategorili maddeler için kullanılabileceğini ortaya koymuştur.

Yurekli (2010) simülasyon çalışmasında GKPM ve ATM modeli kullanarak madde ayırt edicilik parametreleri arasındaki ilişkiyi araştırmıştır. Bunun için 20 maddelik, 5000 kişilik örneklemden ve farklı kategori sayılarından oluşan her iki modele uygun veriler üretmiştir. Elde edilen grafiklerden, GKPM verilerine dayanılarak üretilen 3, 4 ve 5 tepki kategorili ATM hariç, gerçek madde ayırt edicilik parametrelerinin kestirilen madde ayırt edicilik parametrelerinden daha yüksek olduğu görülmüştür. Gerçek ve kestirilen madde ayırt edicilik değerleri arasındaki korelasyonlar (r), 0.97'den yüksektir. GKPM'de bir maddenin daha fazla kategori sayısının olmasının, daha düşük madde ayırt edicilik parametresine yol açtığı, ATM'de ise için madde ayırt edicilik parametresinin kategori sayısından etkilediği bulunmuştur.

Uyar, Öztürk Gübeş ve Kelecioğlu (2013) Uluslararası Öğrenci Değerlendirme Programı (PISA) 2009 kapsamında uygulanan okumaya yönelik tutum ölçeğinden elde edilen madde puanları ATM ile incelemiştir. Araştırmalarında 11 maddelik ölçeğin madde parametrelerini kestirip, madde ve test bilgi fonksiyonlarına ilişkin bilgileri sunmuşlardır. Jiao ve Zhang (2014) madde takımları içeren çok düzeyli çok kategorili verilerde KPM ile analizleri gerçekleştirmiştir. Naumenko (2014) projesinde GKPM ile ATM arasındaki istatistiksel ve teorik model uyumunu, modeller arasındaki yetenek kestirimi sırasını ve çok kategorili ve ikili model kalibrasyonları arasındaki test bilgi fonksiyonlarının şeklini karşılaştırmayı amaçlamıştır. Araştırma sonucunda kullandıkları verilerde GKPM ve ATM eşdeğer model uyumu sergilemiştir. ATM ile GKPM'den elde edilen bilgi fonksiyonları karşılaştırıldığında ise, ATM'nin örtük yetenek kestirimleri hakkında daha geniş bilgi sağladığı bulunmuştur.

Koğar (2015) basit iki boyutlu yapılarda farklı örneklem büyüklüğü ($N=100, 500, 1000, 5000$) ve çeşitli koşullar altında ürettiği verilerde tek boyutlu parametrik olmayan MTK, tek boyutlu MTK ve çok boyutlu MTK'den elde edilen madde parametreleri, maddelere ait model veri uyumları ve teste ait model veri uyumlarını belirlemiş ve elde edilen sonuçları karşılaştırmıştır. Tek boyutlu MTK için madde ayırt ediciliğini belirlemek amacıyla kullanılan a (ayırt edicilik) parametresine ait standart hata değerleri incelendiğinde, örneklem büyüklüğündeki artış ile birlikte, standart hata değerlerinin

düştüğü belirlenmiştir. Tek boyutlu MTK'de b (güçlük) parametresine ait standart hata değerleri incelendiğinde, örneklem büyüklüğü koşulundaki değişim ile birlikte, bu parametreye ait standart hata değerlerindeki değişim bir örüntü ortaya koymamaktadır. Bıkmaz Bilgen ve Doğan (2017) gerçekleştirdikleri simülasyon çalışmasında çok kategorili maddeler için parametrik MTK kapsamındaki ATM ve parametrik olmayan MTK kapsamındaki Monoton Homojenlik Modeli (MHM) ile yapılan kestirimlerde örneklem büyüklüğü (N=100, 250, 500, 1000) ve çeşitli değişkenler açısından model veri uyumları, güvenilirlik değerleri, madde parametrelerinin standart hatalarını incelenmiştir. Normal dağılım sergileyen veri yapısında güvenilirlik kestiriminin her iki modelde de benzer olduğunu bulmuşlardır. MHM ile yapılan kestirimlerde parametrelerin standart hataları ATM'ye oranla daha düşük bulunmuştur. Örneklem büyüklüğü arttıkça ise standart hatalar azalmaktadır.

Alanyazındaki çok kategorili puanlanan verilerle yapılan MTK çalışmaları incelendiğinde yurt dışında birçok çalışmanın olduğu görülmektedir (Cook ve diğerleri, 1999; Jiao ve Zhang, 2014; Kang ve Chen, 2008; Maydeu-Olivares ve diğerleri, Naumenko, 2014; Thissen ve diğerleri, 1995; 1996; Yurekli, 2010). Türkiye'de ise çok kategorili MTK modellerinden ATM kullanılarak parametre kestirimi yapıldığı çalışmalar mevcuttur (Uyar ve diğerleri, 2013; Köse, 2015; Bıkmaz Bilgen ve Doğan, 2017). Tüm bu çalışmalar incelendiğinde farklı örneklem büyüklüklerinde, farklı çok kategorili MTK modellerinden elde edilen madde parametreleri arasındaki ilişkiyi ortaya koyan bir çalışmanın yurt içi ve dışında olmadığı görülmektedir. Bu çalışmanın çok kategorili verilerde farklı örneklem büyüklüğünü ve farklı çok kategorili MTK modellerini göz önünde bulundurması boyutuyla alanyazına katkı sağlayacağı düşünülmektedir.

Bu araştırmada Rosenberg benlik saygısı ölçeğinden elde edilen verilerde aynı veri havuzundan rastgele seçilen 500, 1000 ve 2000 kişilik farklı örneklem büyüklüklerinde MTK'ye dayalı GKPM, 1PLM uzantısı GKPM, KPM ve ATM'den hangisine daha fazla model-veri uyumu gösterdiğinin, bahsedilen her bir model için kestirilen madde parametreleri arasında nasıl bir ilişkinin olduğunun ve en fazla model veri uyumuna

sahip modelde madde parametrelerinin ne olduğunun tespit edilmesi amaçlanmıştır. Bu amaçla aşağıdaki sorulara yanıtlar aranmıştır.

1. 500, 1000 ve 2000 kişilik verilerde GKPM, 1 parametrelili lojistik model uzantısı GKPM, KPM ve ATM'den elde edilen model uyum indeksleri nasıldır?
 - 1.1. Farklı örneklem büyüklüğü koşullarında hangi model daha fazla model- veri uyumu sergilemektedir?
 - 1.2. Örneklem büyüklüğü değiştikçe çok kategorili MTK modellerinin sergilemiş oldukları model- veri uyumları nasıl değişmektedir?
2. 500, 1000 ve 2000 örneklem büyüklüğü koşullarında GKPM, 1 parametrelili lojistik model uzantısı GKPM, KPM ve ATM ile kestirilen madde parametreleri arasında manidar bir ilişki var mıdır?
 - 2.1. Örneklem büyüklüğü değiştikçe çok kategorili MTK modellerinden elde edilen madde parametreleri arasındaki ilişkiler nasıl değişmektedir?
 - 2.2. Kullanılan çok kategorili MTK modellerinden elde edilen parametreler arasındaki ilişkiler incelendiğinde hangi modeller arasında daha yüksek manidar ilişkiler bulunmaktadır?
3. En fazla model veri uyumu sergileyen modele ait Rosenberg Benlik Saygısı ölçeğinin madde parametreleri nasıl kestirilmiştir?

YÖNTEM

Araştırmanın Türü

Bu araştırma farklı örneklem büyüklüklerinde (500, 1000, 2000), farklı çok kategorili MTK modellerinden (GKPM, 1PLM uzantısı GKPM, KPM ve ATM) elde edilen model veri uyumlarını karşılaştırması ve madde parametreleri arasındaki ilişkileri ele alması nedeniyle var olan durumu ortaya koymayı amaçlayan betimsel bir araştırmadır.

Çalışma Grubu

Araştırmanın evreni farklı yaş ve cinsiyetten 47974 bireyin İngilizce Rosenberg Benlik Saygısı Ölçeğine vermiş olduğu yanıtlardan oluşmaktadır. Araştırmanın çalışma grubunu Amerika Birleşik Devletleri'nden ölçekteki maddelere cevap veren kayıp değerler temizlendikten sonra rastgele örnekleme yoluyla seçilen 500, 1000 ve 2000 kişilik bir grup oluşturmaktadır. Araştırmada bu örneklem büyüklüklerinin seçilmesinin nedeni alanyazında genelde bu değerlerle yapılan çalışmaların var olmasıdır.

Veri Toplama Aracı

Araştırma verileri tek boyutlu 10 maddelik Rosenberg Benlik Saygısı Ölçeği ile toplanmıştır. Rosenberg Benlik Saygısı Envanterini 1965 yılında Morris Rosenberg 12 alt ölçekte toplam 63 maddeden oluşacak şekilde Amerika'da 5024 lise öğrencisi üzerinde geçerlik ve güvenirlik çalışmalarını yürüterek geliştirilmiştir. Rosenberg alt ölçeklerin ayrı ayrı kullanılabilceğini belirtmiştir. Envanterin ilk on maddesi Rosenberg Benlik Saygısı Ölçeğini oluşturmaktadır. Ölçeğin marjinal güvenirlik katsayısı incelendiğinde 500 kişilik örneklem için 0.92; 1000 kişilik örneklem için 0.93; 2000 kişilik örneklem için 0.93 bulunmuştur. Araştırmada kullanılan ölçek 1=kesinlikle katılmıyorum, 2=katılmıyorum, 3=katılıyorum ve 4=kesinlikle katılıyorum olmak üzere dört kategoriden oluşmaktadır. Ölçekte beş madde (3,5,8,9,10) ters kodlanmıştır. Araştırma verilerine <https://openpsychometrics.org/> adlı yabancı bir veri havuzundan erişilmiştir. Bu veri havuzunda farklı ölçeklere ve testlere farklı demografik özelliklere sahip bireylerin verdikleri gönüllülük esasına dayalı yanıtlar bulunmaktadır. Toplam 32 veri toplama aracına ait veriler araştırmacıların kullanımına açıktır.

Verilerin Analizi

Verilerin analizi dört aşamada gerçekleştirilmiştir. İlk aşamada her örneklem büyüklüğü için MTK varsayımları kontrol edilmiştir. Ardından her birçok kategorili MTK modeli için uyum iyiliği istatistikleri hesaplanmış ve bu modellerin model veri uyumları karşılaştırılmıştır. Üçüncü aşamada her modele ait madde parametreleri kestirilmiştir. Dördüncü ve son aşamada ise 500, 1000 ve 2000 kişilik örneklem büyüklüklerinden ve

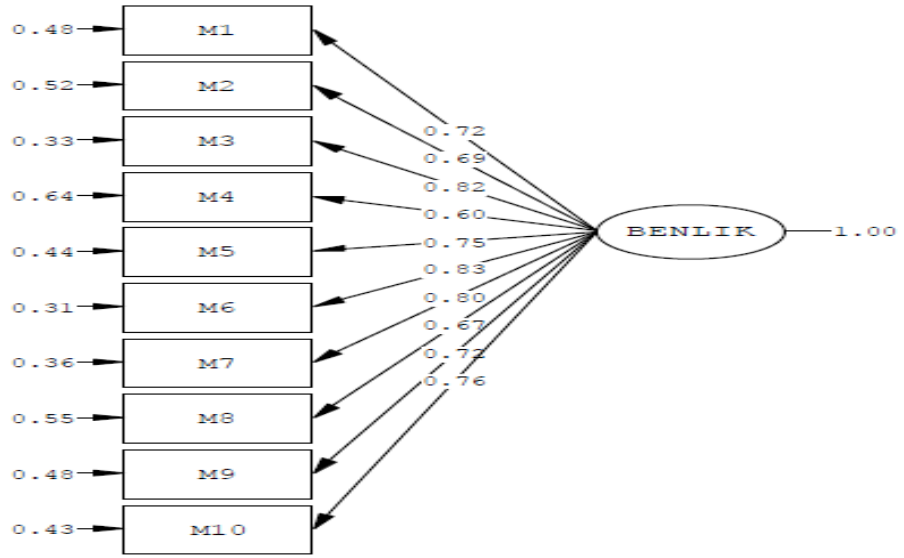
GKPM, 1PLM uzantısı GKPM, KPM ve ATM modellerinden kestirilen madde parametreleri arasında manidar bir ilişki olup-olmadığı incelenmiştir. Örneklemin rastgele olarak seçilmesinde, normallik varsayımının incelenmesinde ve kestirilen parametreler arasındaki ilişkilerin hesaplanmasında SPSS programından yararlanılmıştır. Tek boyutluluk varsayımının incelenmesinde LISREL yazılımından yararlanılmıştır. Marjinal güvenilirlik katsayılarının hesaplanması ve her bir madde çifti arasında kestirilen yerel bağımlılık X^2 (LD X^2) değeri ile yerel bağımsızlığın incelenmesinde IRTPRO (Item Response Theory for Patient-Reported Outcomes) yazılımı kullanılmıştır. Parametrelerin kestiriminde R Studio “lrm” paketinden yararlanılmıştır.

MTK Varsayımlarının Kontrol Edilmesi

MTK varsayımlarının sağlanıp-sağlanmadığına her örneklem büyüklüğünden elde edilen veriler için ayrı ayrı bakılmıştır. Tek boyutluluk, yerel bağımsızlık ve normallik varsayımların her örneklem büyüklüğü için sağlandığı görülmüştür.

Tek Boyutluluk

Her örneklem büyüklüğü koşulunda tek boyutluluk varsayımı sağlanmaktadır. Bunun için 500, 1000 ve 2000 kişilik örneklemelerden elde edilen verilere alanyazında Rosenberg Benlik Saygısı Ölçeği'nin tek boyutlu yapıyı ölçtüğü belirtildiğinden doğrulayıcı faktör analizi ile tek boyutlu yapı ortaya konmuştur. Aşağıda Rosenberg Benlik Saygısı Ölçeği 500 kişilik örneklem verisi için faktör yapısına yer verilmiştir.



Şekil 1. Rosenberg Benlik Saygısı Ölçeği 500 Kişilik Örneklem Verileri Faktör Yapısı

Tablo 1. Rosenberg Benlik Saygısı Ölçeği 500 Kişilik Örneklem Verileri Model Uyum İndeksleri

Uyum Ölçütleri	İyi Uyum	Kabul Edilebilir Uyum	Ölçek İçin Elde Edilen Değerler
RMSEA	$0 \leq \text{RMSEA} \leq .05$	$.05 < \text{RMSEA} \leq .08$	0.08
p değeri (RMSEA < .05)	$10 < p \leq 1.00$	$.05 \leq p \leq .10$	1.00
SRMR	$0 \leq \text{SRMR} \leq .05$	$.05 < \text{SRMR} \leq .10$	0.06
NFI	$95 \leq \text{NFI} \leq 1.00$	$.90 \leq \text{NFI} < .95$	0.94
NNFI	$97 \leq \text{NNFI} \leq 1.00$	$.95 \leq \text{NNFI} < .97$	0.95
CFI	$.97 \leq \text{CFI} \leq 1.00$	$.95 \leq \text{CFI} < .97$	0.96

Kaynak: Schermelleh-Engel, Moosbrugger ve Müller (2003)

500 kişilik örneklemden elde edilen verilerde uyum iyiliği istatistikleri (Non-Normed Fit Index (NNFI), Normed Fit Index (NFI), Goodness of Fit Index (GFI), Comparative Fit Index (CFI)) incelendiğinde ise 0.90'nın üstünde kabul edilebilir değerler elde edilmiştir. Aynı şekilde RMSEA (Root Mean Square Error of Approximation) değeri de kabul edilebilir değer aralığında bulunmaktadır. Diğer örneklem büyüklükleri için de (1000, 2000) faktör yükleri ve uyum iyiliği istatistikleri benzer bulunmuştur.

Yerel Bağımsızlık

Her bir madde çifti arasında kestirilen yerel bağımlılık X^2 (LD X^2) değerlerinin 10'un altında olması yerel bağımsızlığın bir göstergesi olarak ele alınmıştır (Cai, du Toit ve Thissen, 2011). Bu aşamada her örneklem büyüklüğü koşulunda GPCM, 1PLM uzantısı GKPM, KPM ve ATM için ayrı ayrı kestirilen LD X^2 değerleri incelenmiş ve genelde 10'un altında bulunmuştur. Marjinal ki-kare değerleri aynı şekilde her model için incelenmiş ve 0'a çok yakın değerler elde edilmiştir. Bu durumda her örneklem büyüklüğü koşulunda GKPM, 1PLM uzantısı GKPM, KPM ve ATM için yerel bağımsızlık varsayımının ihlal edilmediği söylenebilir.

BULGULAR

İlk aşamada çok kategorili madde tepki kuramı modellerinin model veri uyumu için -2logolabilirlik (-2loglikelihood), Akaike bilgi ölçütü (Akaike information criterion (ABÖ)) ve Bayesian bilgi ölçütü (Bayesian information criterion (BBÖ)) değerleri incelenmiştir. Elde edilen model veri uyumu indeksleri Tablo 2’de verilmiştir.

Tablo 2. Çok Kategorili MTK Modelleri Model Uyum İndeksleri

Koşul	Model	-2logolabilirlik	ABÖ	BBÖ
500	GKPM	9721.63	9801.63	9970.21
	GKPM(1PL M Uzantısı)	9812.25	9874.25	10004.90
	KPM	10026.16	10086.16	10212.60
	ATM	9665.31	9745.31	9913.90
1000	GKPM	19863.94	19943.94	20140.25
	GKPM(1PL M Uzantısı)	20105.64	20167.64	20319.78
	KPM	20518.90	20578.90	20726.13
	ATM	19700.37	19780.37	19976.68
2000	GKPM	39463.98	39543.98	39768.02
	GKPM(1PL M Uzantısı)	39929.34	39991.33	40164.96
	KPM	40797.68	40857.67	41025.70
	ATM	39140.92	39220.92	39444.96

ABÖ ve BBÖ değerleri küçülüp sifira yaklaştıkça model veri uyumunun daha iyi olduğunu göstermektedir (de Ayala, 2009, s. 142). -2 logolabilirlik değeri de sifira yaklaştıkça (düşükçe) uyum artar (Kline, 2005). Tablo 2 incelendiğinde her örneklem koşulu için (N=500, 1000, 2000) aşamalı tepki modelinin sifira en yakın en iyi uyum indekslerine sahip olduğu görülmektedir. Örneklem büyüklüğü değiştikçe çok kategorili MTK modellerinin sergilemiş oldukları model veri uyumları değişmemektedir, her örneklem büyüklüğünde sırasıyla en yüksek uyumun ATM, GKPM, 1PLM uzantısı GKPM ve KPM'de olduğu görülmektedir.

Farklı modellerden kestirilen madde parametreleri arasındaki ilişkiler Spearman sıra farkları korelasyon katsayısıyla hesaplanmıştır. Tablo 3'te 500, 1000 ve 2000 örneklem büyüklüğü için GKPM ve ATM ile kestirilen ayırt edicilik (a) parametreleri arasındaki korelasyon katsayıları verilmiştir.

Tablo 3. Farklı Çok Kategorili Modellerde a Parametresi Arasındaki İlişkiler

Koşul	Modeller	Korelasyon Değerleri
500	GKPM-ATM	0.84*
1000	GKPM-ATM	0.95*
2000	GKPM-ATM	0.94*

* $p < .01$

GKPM ve ATM ile kestirilen ayırt edicilik parametreleri arasındaki ilişkiler incelendiğinde tüm örneklem büyüklüklerinde manidar yüksek bir ilişkinin olduğu görülmektedir ($p < 0.01$). İlişki katsayıları incelendiğinde a parametresi için en yüksek ilişkinin 0.952 ile 1000 kişilik örneklemden elde edildiği görülmektedir. KPM ve 1-PLM uzantısı GKPM modelinde a parametreleri sabit bir değer olduğundan Tablo 3'te yer verilmemiştir.

Tablo 4. Farklı Çok Kategorili Modellerde b_1 Parametresi Arasındaki İlişkiler

Koşul	Model	GKPM	GKPM(1PLM Uzantısı)	KPM
500	GKPM			
	GKPM(1PLM Uzantısı)	0.95*		
	KPM	0.95*	1*	
	ATM	0.99*	0.94*	0.94*
1000	GKPM			
	GKPM(1PLM Uzantısı)	0.95*		
	KPM	0.95*	1*	
	ATM	1*	0.95*	0.95*
2000	GKPM			
	GKPM(1PLM Uzantısı)	0.99*		
	KPM	0.99*	1*	
	ATM	0.99*	0.96*	0.96*

* $p < 0.01$

Tablo 4 incelendiğinde farklı örneklem büyüklükleri için kestirilen b_1 eşik parametreleri arasındaki ilişkinin genellikle örneklem büyüklüğü arttıkça arttığı görülmektedir. Her üç örneklem için her modelde kestirilen b_1 parametrelerinin KPM ile 1PLM uzantısı GKPM arasında mükemmel ilişki gösterdiği görülmektedir. Daha sonra 1000 örneklem büyüklüğü için ATM ile GKPM arasında mükemmel ilişki vardır. Her örneklem büyüklüğünde her modelde b_1 parametreleri için kestirilen değerler arasında yüksek manidar bir ilişki bulunmuştur ($p < 0.01$).

Tablo 5. Farklı Çok Kategorili Modellerde b_2 Parametresi Arasındaki İlişkiler

Koşul	Model	GKPM	GKPM(1PLM Uzantısı)	KPM
500	GKPM			
	GKPM(1PLM Uzantısı)	0.82*		
	KPM	1*	0.82*	
	ATM	0.99*	0.81*	0.99*
1000	GKPM			
	GKPM(1PLM Uzantısı)	0.99*		
	KPM	0.99*	1*	
	ATM	0.99*	0.98*	0.98*
2000	GKPM			
	GKPM(1PLM Uzantısı)	1*		
	KPM	0.99*	1*	
	ATM	0.99*	0.99*	0.99*

* $p < 0.01$

Tablo 5 incelendiğinde örneklem büyüklüğü arttıkça genelde farklı modellerden kestirilen b_2 parametreleri arasındaki ilişkinin arttığı görülmektedir. İlişki katsayıları incelendiğinde örneklem büyüklüğü farklılaştıkça ilişkinin düzeyinin de farklılaştığı görülmektedir. Örneğin; 500 kişilik örneklemde GKPM ile KPM arasında mükemmel ilişki varken, 1000 ve 2000 kişilik örneklemde bu değer 0.988 olduğu görülmektedir. Benzer şekilde 500 kişilik örneklemde 1PLM uzantısı GKPM ve KPM arasında 0.818 değerinde bir ilişki varken, 1000 ve 2000 kişilik örneklemde bu değer yükselerek mükemmel ilişki var olmuştur. Buradan örneklem büyüklüğü ile farklı modellerden kestirilen parametreler arasında düzenli bir örüntünün olmadığı, örneklem büyüklüğü

artıkça veya azaldıkça kestirilen parametreler arasındaki ilişkinin düzeyinin artacağı veya azalacağını söyleyemeyeceği sonucu çıkmaktadır. Farklı örneklem büyüklüğü koşullarında farklı modellerden kestirilen b_2 parametreleri arasında manidar yüksek bir ilişki bulunmaktadır ($p < 0.01$).

Tablo 6. Farklı Çok Kategorili Modellerde b_3 Parametresi Arasındaki İlişkiler

Koşul	Model	GKPM	GKPM(1PLM Uzantısı)	KPM
500	GKPM			
	GKPM(1PLM Uzantısı)	0.99*		
	KPM	0.99*	1*	
	ATM	0.95*	0.93*	0.93*
1000	GKPM			
	GKPM(1PLM Uzantısı)	0.99*		
	KPM	0.99*	0.99*	
	ATM	0.97*	0.95*	0.93*
2000	GKPM			
	GKPM(1PLM Uzantısı)	0.97*		
	KPM	0.95*	0.99*	
	ATM	0.95*	0.90*	0.88*

* $p < 0.01$

Tablo 6 incelendiğinde farklı örneklem büyüklüğü koşullarında farklı modellerden kestirilen b_3 parametreleri arasında manidar yüksek bir ilişki bulunmaktadır ($p < 0.01$). Tablo 6'dan da görüleceği üzere örneklem büyüklüğü 2000 olduğunda genelde farklı modellerden kestirilen b_3 parametreleri arasındaki ilişkinin azaldığı görülmektedir.

Ayrıca ilişki katsayılarının değerlerinin örneklem büyüklüğüne göre düzenli bir örüntü sergilemediği de Tablo 6'dan anlaşılmaktadır.

Tablo 3, Tablo 4, Tablo 5 ve Tablo 6 incelendiğinde 500, 1000 ve 2000 örneklem büyüklüğü koşullarında GKPM, 1PLM uzantısı GKPM, KPM ve ATM ile kestirilen madde parametreleri arasında manidar bir ilişkinin olduğu görülmektedir ($p < 0.01$). Örneklem büyüklüğü değiştikçe çok kategorili MTK modellerinden elde edilen madde parametreleri arasındaki ilişkilerin değişmesinde her hangi bir örüntü bulunmamaktadır. Buradan örneklem büyüklüğü arttıkça farklı modellerden elde edilen madde parametreleri arasındaki ilişkilerin derecesinin her zaman artmayacağı sonucu çıkarılabilir. Kullanılan çok kategorili MTK modellerinden elde edilen parametreler arasındaki ilişkiler incelendiğinde genellikle GKPM, 1PLM uzantısı GPM ile KPM arasında daha yüksek manidar ilişkiler bulunmaktadır.

Tablo 7. Aşamalı Tepki Modeli ile Kestirilen Madde Parametreleri

Koşul	500				1000				2000			
	a	b1	b2	b3	a	b1	b2	b3	a	b1	b2	b3
/ Madde												
M1	2.26	-2.45	-1.32	0.23	2.66	-1.91	-0.99	0.44	2.84	-1.92	-0.92	0.52
M2	2.18	-2.87	-1.70	0.03	2.27	-2.44	-1.49	0.27	2.37	-2.33	-1.42	0.43
M3	3.13	-1.56	-0.62	0.46	3.06	-1.42	-0.43	0.69	2.94	-1.41	-0.40	0.79
M4	1.51	-3.07	-1.27	0.66	1.60	-2.55	-1.02	0.83	1.62	-2.55	-1.02	0.93
M5	2.41	-1.54	-0.51	0.55	2.61	-1.32	-0.32	0.77	2.48	-1.39	-0.31	0.88
M6	3.37	-1.49	-0.39	0.86	3.43	-1.22	-0.22	1.01	3.30	-1.22	-0.15	1.10
M7	2.84	-1.36	-0.17	0.98	2.66	-1.13	0.02	1.24	2.70	-1.11	0.03	1.28
M8	1.83	-1.16	0.12	1.33	1.61	-1.15	0.29	1.45	1.54	-1.12	0.35	1.58
M9	2.19	-1.12	0.14	1.05	2.13	-0.94	0.39	1.23	2.24	-0.90	0.44	1.30
M10	2.37	-1.27	-0.22	0.59	2.46	-1.06	-0.03	0.76	2.52	-0.99	0.08	0.83

En fazla model veri uyumunun sergilendiği ATM ile kestirilen madde parametreleri Tablo 7’de verilmiştir. 500 kişilik örneklem için ATM altında en yüksek a parametresinin 3.37 ile madde 6 olduğu; en düşük a parametresinin ise 1.83 ile madde 8 olduğu görülmektedir. 1000 kişilik örneklem için ATM ile kestirilen ayırt edicilik parametreleri incelendiğinde 3.431 (6. Madde) ile 1.601 (4. Madde) arasında değiştiği görülmektedir. 2000 kişilik örneklem için ATM ile kestirilen ayırt edicilik parametreleri incelendiğinde yine 3.302 ile 6. maddenin en yüksek ayırt ediciliğe sahip olduğu, 1.545 değeri ile madde 8’in ise en düşük ayırt ediciliğe sahip madde olduğu bulunmuştur. Baker (2001, s.35)’e göre genelde bu maddeler 1.70’ten yüksek olduğundan çok yüksek ayırt ediciliğe, ayırt ediciliği 1.70’in altında maddeler ise 1.35-1.69 aralığında yer aldığından yüksek ayırt ediciliğe sahip maddeler olarak nitelendirilebilir. Ayırt edicilik parametreleri GKPM ile de kestirildiğinde benzer bulgular bulunmuştur. Farklı bir bulgu olarak 1000 örneklemde elde edilen verilerde en düşük ayırt ediciliğe sahip madde 8’dir. ATM ile kestirilen a parametreleri GKPM ile kestirilen a parametrelerinden göreceli olarak yüksektir.

Tablo 7’de yer alan b parametreleri incelendiğinde ise farklı örneklemde elde edilmesine karşın en yüksek (M8) ve en düşük (M4) değere sahip maddelerin parametrelerinin sıralamasının değişmediği görülmektedir. Eşik parametreleri 0.50 olasılıkla j eşik kategorisinin üzerinde cevap vermek için gerekli özellik düzeyini temsil eder. Tablo 7 incelendiğinde bir kişinin ilk kategoriyi cevaplaması için düşük benlik saygısına sahip olması yeterli olmakta iken, bir üst kategori için daha yüksek benlik saygısı gerekmektedir. 500 örneklem için eşik parametreleri, -3.074 ile 1.326 değerleri arasında, 1000 örneklem için -2.548 ile 1.447 arasında, 2000 kişilik örneklem için -2.555 ile 1.579 arasında değerler almaktadır. Tablo 7 de görüldüğü gibi %50 olasılıkla eşik değerinin altında puan almak için gerekli en düşük yetenek düzeyi 500 kişilik örneklemde 1000 ve 2000 örnekleme göre daha düşüktür, 1000 ve 2000 örneklem için oldukça yakındır. Her örneklem büyüklüğü için eşik parametreleri incelendiğinde b_1 ve b_2 değerinin genelde negatif b_3 değerlerin pozitif olduğu görülmektedir. b_1 parametreleri incelendiğinde maddelerin daha çok düşük benlik saygısı düzeyindeki bireyler hakkında

bilgi verdiğini görülmektedir. Bu maddelerin b_2 parametreleri incelendiğinde orta ve düşük benlik saygısı düzeyinde daha fazla bilgi vermektedir. b_3 parametreleri incelendiğinde ise genelde orta ve yüksek benlik saygısına sahip bireyler hakkında daha fazla bilgi edinilmektedir.

TARTIŞMA ve SONUÇ

500, 1000 ve 2000 kişilik verilerde geliştirilmiş kısmi puan, 1 parametrelili lojistik model gibi sınırlandırılmış geliştirilmiş kısmi puan, kısmi puan ve aşamalı tepki modelinden elde edilen model uyum indeksleri incelendiğinde her örneklem büyüklüğü koşulunda aşamalı tepki modelinin en fazla model veri uyumunu sağladığı belirlenmiştir. Örneklem büyüklüğü değişse bile ATM'nin en yüksek uyumu sağladığı, ardından GKPM, 1PLM uzantısı GKPM ve KPM'nin sırasıyla uyum gösterdiği bulunmuştur. Buradan örneklem büyüklüğünün değişmesinin bu veri setinde farklı modellerin veri uyumu sırasını değiştirmeyeceği sonucu çıkarılabilir. Cook, Dodd ve Fitzpatrick (1999) 1994 Akademik Durum Belirleme Testi (Scholastic Assessment Test I (SAT I)) verileri ve simülasyon çalışmasından elde ettikleri verilerle GKPM, ATM ve KPM'yi karşılaştırmış ve KPM'nin SAT I sözel verileri için GKPM ve ATM'den, simülasyon verilerde ise ATM'nin ve GKPM'nin ise KPM'den göreceli olarak daha iyi uyum gösterdiklerini belirtmişlerdir. Bu bulgular veri setinin yapısına göre kullanılan MTK modellerinin model veri uyumunun değişeceğinin, bir modelin diğer bir modelden bir veri setine dayanarak üstün olduğunu söylemenin güç olduğunu göstermektedir. Bunun yanında herhangi bir kategori sayısı için a parametrelerinin tutarlılığından dolayı ATM'nin sıralı yanıt verileri için daha uygun olduğunu belirten araştırmalara da rastlanmıştır (Yurekli, 2010). Naumenko (2014) GKPM, KPM ve 1PLM ATM, ATM ile ilişkili $S-X^2$ istatistiklerinin tutarlı bir şekilde üstün uyum gösteremediğini, genel olarak, veri uyumunun, GKPM ve ATM arasında eşdeğer görüldüğünü belirtmiştir. Bu bulgu farklı örneklem büyüklüklerinde model veri uyumunda öncelikle ATM, ardından GKPM'nin daha iyi uyum indeksleri üretmesiyle kısmen tutarlılık göstermektedir. Koğar (2014) tek boyutlu MTK'de örneklem büyüklüğündeki artış ile birlikte, teste ait

model veri uyumu indekslerinin de arttığına değinmiştir. Bu durum model veri uyumu indekslerinin hesaplanma formülünden dolayı beklenen bir durum olarak düşünülebilir. Koğar (2014)'ün bulgularıyla tutarlı olarak bu araştırmada da örneklem büyüklüğü arttıkça model veri uyumu indeksleri değerlerinin de arttığı bulunmuştur.

500, 1000 ve 2000 kişilik örneklem büyüklüğü koşullarında GKPM, 1 parametrelili lojistik model gibi sınırlandırılmış GKPM, KPM ve ATM ile kestirilen madde parametreleri arasında manidar bir ilişki bulunmuştur. Bu ilişkiler örneklem büyüklüğüne bağlı olarak değişmemektedir. Daha büyük örnekleme (2000), orta dereceli örneklemeden (1000) elde edilen verilerin farklı modellerle analizinden kestirilen parametreler arasındaki ilişki daha düşük olabilmektedir. Elde edilen bulgulardan yola çıkarak kullanılan çok kategorili MTK modellerinden elde edilen parametreler arasındaki ilişkiler incelendiğinde belirli modeller arasında daha yüksek manidar ilişkiler bulunmaktadır denilememektedir; çünkü farklı parametrelerde, farklı koşullarda farklı sonuçlar elde edilmiştir. En fazla model veri uyumu sergileyen ATM ile kestirilen Rosenberg Benlik Saygısı ölçeğinin madde parametreleri incelendiğinde ise; madde ayırt edicilik parametresinin yüksek; madde eşik parametreleri incelendiğinde ise b_1 parametresinin daha çok düşük benlik saygısına sahip bireyler hakkında bilgi verirken, b_2 parametresinin düşük ve orta benlik saygısına sahip bireyler hakkında, b_3 parametresinin ise orta ve yüksek benlik saygısına sahip bireyler hakkında bilgi verdiği bulunmuştur. Hays, Morales ve Reise (2000, s. 4) genelde yüksek eğitim/ ayırt edicilik parametresine sahip maddelerin daha fazla bilgi sağladığını belirtmişlerdir. Bu durumda madde bilgi fonksiyonları da incelendiğinde ATM altında 500 ve 1000 kişilik örneklem büyüklükleri için en fazla bilginin madde 6'dan, en az bilginin ise madde 4'den sağlandığı söylenebilir. 2000 kişilik örneklem için en fazla bilgi yine madde 6'dan; en az bilgi ise madde 8'den sağlanmıştır. Madde bilgi fonksiyonları incelendiğinde en yüksek ayırt ediciliğe sahip maddelerin en fazla bilgiyi ve en düşük ayırt ediciliğe sahip maddenin en az bilgiyi sağladığı bulunmuştur. Elde edilen sonuçlar Hays, Morales ve Reise (2000, s. 4)'in yüksek ayırt edicilik parametresine sahip maddelerin daha fazla bilgi sağladığı bulgularıyla örtüşmektedir. Çok kategorili MTK

modellerinin özelliklerine değinen alanyazında birçok araştırma vardır (Samejima, 1995; Hemker, Sijtsma, Molenaar ve Junker, 1996, 1997; Sijtsma ve Hemker, 1998, 2000; Akkermans, 1999; Van Der Ark, 2001). Tüm bu çalışmalar incelendiğinde her ölçüte göre ve her durumda bir modelin bir modelden üstün olduğunu söylemek güçtür.

De Ayala (2009) çok kategorili MTK modellerinde varsayımlar karşılandığı taktirde en az 500 örneklem büyüklüğü ile çalışılmasını önermiştir. Benzer şekilde ATM ile parametre kestirimlerinin daha güvenilir ve daha az hatalı olması için en az 500 örneklem büyüklüğüne ihtiyaç duyulmaktadır (Bıkmaz Bilgen ve Doğan, 2017). Bu çalışmada ise örneklem büyüklüğünün farklı modellerden elde edilen parametre kestirimleri arasındaki ilişkilerin derecesinde ve model veri uyumunda önemli değişikliklere yol açmadığı bulunmuştur. Bundan sonraki çalışmalarda farklı örneklem büyüklükleri ve farklı yazılımlar ile çok kategorili MTK modelleri çalışılabilir.

KAYNAKLAR

- Akkermans, W. (1999). Polytomous item scores and Guttman dependence. *British Journal of Mathematical and Statistical Psychology*, 52(1), 39–62.
- Baker, F. B. (2001). *The basics of item response theory* (Second edition). Washington DC.:ERIC.
- Bıkmaz Bilgen, Ö., & Doğan, N. (2017). Çok Kategorili Parametrik ve Parametrik Olmayan Madde Tepki Kuramı Modellerinin Karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(4), 354-372.
- Cai, L., du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: User guide*. Lincolnwood, IL: Scientific Software International.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*(First edition). California: Thomson Learning.
- Cook K.F., Dodd B.G.,& Fitzpatrick S.J. (1999). A comparison of three polytomous item response theory models in the context of testlet scoring.*Journal of outcome measurement*, 3(1), 1-20.
- De Ayala, R. J. (2009). *The theory and practice of item response theory (Methodology in the social sciences)*.New York: Guildford Press.
- Doğan, N. ve Tezbaşaran, A.A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklem bağlamında karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 25(25), 58-67.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and psychological measurement*, 58(3), 357-381.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Nijhoff.
- Han, K. T., & Hambleton, R. K. (2014). *User's manual for wingen 3: windows software that generates IRT model parameters and item responses*(Center for Educational Assessment Report No. 642). Amherst, MA: University of Massachusetts.

- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage.
- Hambleton, R. K. , & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
- Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical care*, 38(9 Suppl), II28. *National Institutes of Health Public Access Author Manuscript*. Retrieved 17/03/2017 from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1815384/pdf/nihms14476.pdf>
- Hemker, B. T., Sijtsma, K., Molenaar, I. W., & Junker, B. W. (1996). Polytomous IRT models and monotone likelihood ratio of the total score. *Psychometrika*, 61(4), 679–693.
- 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. *Psychometrika*, 62(3), 331–347.
- Jiao, H., & Zhang, Y. (2014). Polytomous multilevel testlet models for testlet-based assessments with complex sampling designs. *British Journal of Mathematical and Statistical Psychology*, 68(1), 65-83.
- Kang, T., & Chen, T. T. (2008). Performance of the generalized S-X2 item fit index for polytomous IRT models. *Journal of Educational Measurement*, 45(4), 391-406.
- Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*, Thousand Oaks, CA: Sage.
- Koğar, 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
- Köse, A. (2015). Aşamalı tepki modeli ve klasik test kuramı altında elde edilen test ve madde parametrelerinin karşılaştırılması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 15(2), 184-197.
- Macdonald, P., & Paunonen, S. V. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and psychological measurement*, 62(6), 921-943.

- Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment*, 1(1), 1-11.
- Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.
- Maydeu-Olivares, A., Drasgow, F., & Mead, A. D. (1996). Distinguishing among parametric item response models for polytomous ordered data. *Applied Psychological Measurement*, 18(3), 245-256.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159-176.
- Naumenko, O. (2014). Comparison of various polytomous item response theory modeling approaches for task based simulation cpa exam data. *AICPA 2014 Summer Internship Project*. The University of North Carolina, Greensboro.
- Ndalichako, J. L., & Rogers, W. T. (1997). Comparison of finite state score theory, classical test theory, and item response theory in scoring multiple-choice items. *Educational and psychological measurement*, 57(4), 580-589.
- Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*(No. 144). Thousand Oaks: Sage.
- Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society. Retrieved 20.03.2017 from <http://www.psychometrika.org/journal/online/MN17.pdf>
- Samejima, F. (1972). *A general model for free-response data* (Psychometric Monograph No. 18). Richmond, VA: Psychometric Society. Retrieved 18.03.2017 from <http://www.psychometrika.org/journal/online/MN18.pdf>
- Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika*, 60(4), 549-572.
- Samejima, F. (1996). *The graded response model*. In: van der Linden, W. J., & Hambleton, R., (editors). *Handbook of modern item response theory* (p. 85-100). New York, NY: Springer.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit. *Measures Of Psychological Research Online*, 8(2), 23-74.

- Sijtsma, K., & Hemker, B. T. (1998). Nonparametric polytomous IRT models for invariant item ordering, with results for parametric models. *Psychometrika*, 63(2), 183–200.
- Sijtsma, K., & Hemker, B. T. (2000). A taxonomy of IRT models for ordering persons and items using simple sum scores. *Journal of Educational and Behavioral Statistics*, 25(4), 391–415.
- Thissen, D., Pommerich, M., Billeaud, K., & Williams, V. S. (1995). Item response theory for scores on tests including polytomous items with ordered responses. *Applied Psychological Measurement*, 19(1), 39-49.
- Uyar,Ş., Öztürk Gübeş, N. ve Kelecioğlu, H. (2013). PISA 2009 tutum anketi madde puanlarının aşamalı madde tepki modeli ile incelenmesi. *Eğitim ve Öğretim Araştırmaları Dergisi*, 2(4),125-134.
- Van Der Ark, L. A. (2001). Relationships and properties of polytomous item response theory models. *Applied Psychological Measurement*, 25(3), 273-282.
- Yurekli, H. (2010). *The relationship between parameters from some polytomous item response theory models*. (Unpublished Master's thesis). The Florida State University, College Of Education, Florida.

SUMMARY

There are different polytomous models for the application of item response theory. In this study; in the randomly selected 500,1000 and 2000 people sample sizes in the data obtained from Rosenberg self-esteem scale; it is aimed to determine to which model model-data fit will be shown from the generalized partial credit (GPCM) based on item response theory, restricted generalized partial credit such as 1 parameter logistic model, partial credit (PCM) and graded response model (GRM), how relationship will be between estimated item parameters and what item parameters will be in the model having the most model–data fit. Descriptive study comprised of the answers of 47974 individuals from different age and sex to English Rosenberg self-esteem scale. The sample of the research was comprised of randomly selected 500,1000, and 2000 people samples after missing and extreme values were cleared. As a result of the research, it was seen that the graded response model provided maximum model data fit. Even if the size of the sample changed, it was found that GRM provided the highest fit, then GPCM, restricted GPCM with 1PLM and PCM respectively. The result here was that the change in sample size did not change the model data fit order of different models in this data set. Cook, Dodd and Fitzpatrick (1999) compared GPCM, GRM and PCM with the data they obtained from SAT 1 and simulations and remarked PCM showed relatively better fit than GPCM and GRM for the SAT 1 verbal data and GRM and GPCM showed relatively better fit than PCM in simulation data. These findings showed that model data fit of IRT models used changed according to structure of the data set, it is difficult to say a model was superior to a model based on a data set. According to Yurekli (2010) there are also studies that indicate that for any number of categories, a parameters was more appropriate for GRM sequential acknowledged data because of its consistency. Naumenko (2014) remarked GPCM, PCM and 1PL GRM, the $S-X^2$ statistics associated with GRM show that they could not comply consistently, in general, data fit was equivalent between GPCM and GRM. This finding is partly consistent with the fact that GRM, followed by GPCM, produces better fit indices in model data fit at different sample sizes. In the case of 500, 1000 and 2000 sample size, a meaningful relationship was found between the estimated item parameters by GPCM, Generalized Partial Credit restricted like 1 parameter logistic model, PCM and GRM. These relationships did not change depending on the size of the sample. In the higher sample (2000), the relationship between the estimated parameters from different model analyzes of the data obtained from the intermediate sample (1000) may be lower. When the relations between the parameters obtained from the polytomous IRT models used from the obtained findings are examined, it can not be said that there are higher significant relations among the certain models; because different parameters have different results in different conditions. When the item parameters of the Rosenberg Self-Esteem estimated by GRM, which has the highest model fit, are examined; it is found that the item discrimination parameter is high; When item threshold parameters are examined, b_1 parameter gives more information about individuals with low self-esteem, b_2 parameter gives information about individuals with low and medium self-esteem, b_3 parameter gives information about middle and high self esteem. Hays, Morales and Reise (2000) generally indicated that items with high slope/discrimination parameters provide more information. In this case, it can be said that the maximum amount of information for sample size of 500 and 1000

people under GRM is provided from item 6 and the minimum information from item 4. For the sample of 2000 people, the most information was obtained again from item 6; the least information was provided in item 8. De Ayala (2009) has suggested working with at least 500 sample sizes if assumptions are made in polytomous IRT models. In this study, it was found that the sample size did not lead to significant changes in the degree of relations between the parameter estimates obtained from different models and the model data fit.