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.

Geliş Tarihi: 04/08/2015 Kabul Tarihi: 20/11/2015

# AŞAMALI TEPKİ MODELİ VE KLASİK TEST KURAMI ALTINDA ELDE EDİLEN TEST VE MADDE PARAMETRELERİNİN KARŞILAŞTIRILMASI

İbrahim Alper KÖSE*

**ÖZ**

Bu çalışmanın temel amacı klasik test kuramı (KTK) ve aşamalı tepki modeli (ATM) altında Yapılandırmacı Öğrenme Ortamları Ölçeğinin kestirilen madde ve test parametrelerinin karşılaştırılarak, araştırmacılara hangi kuram altında kestirilen parametrelerin daha keskin ve güvenilir olduğunu sunmaktır. Geleneksel güvenirlik analizi ve özdeğer grafiği sonuçları veri grubunun tek boyutlu olduğunu ve ölçekten elde edilen sonuçların güvenilir olduğunu göstermektedir. Yapılandırmacı öğrenme ortamları ölçeği ortaokul formunun madde ortalaması 3,39 olarak hesaplanmış ve bu da orta düzeyde madde güçlüğüne karşılık gelmektedir. Madde ayırıcılık parametreleri incelendiğinde KTK altında 10 maddenin düşük ayırıcılığa sahip olduğu, ATM altında ise 3 maddenin düşük ayırıcılığa sahip olduğu ve ayırıcılık değerlerinin 0,61 ile 1,81 arasında değiştiği ortaya konmuştur. Her iki kuram altında madde ayırıcılık parametreleri arasında r=0,96 düzeyinde korelasyon bulunmuştur. Sonuç olarak her iki kuram altında kestirilen madde ve test parametrelerinin yüksek düzeyde ilişkili olduğu, model varsayımların karşılandığı ve sınırlılıkların göz önüne alındığı durumlarda her iki kuramında kullanılabileceği sonucuna ulaşılmıştır

**Anahtar Kelimeler:** Aşamalı tepki modeli, madde tepki kuramı, klasik test kuramı, Madde parametre kestirimi

# COMPARISON OF TEST AND ITEM PARAMETERS UNDER GRADED RESPONSE MODEL (IRT) AND CLASSICAL TEST THEORY

**ABSTRACT**

The primary objective of the present research was to investigate the test and item parameters of Constructivist Learnin Envirement Scale (CLES) under graded response model (GRM) and classical test theory (CTT) and submit the best model which was fitted to real data. The results from a traditionally reliability analysis and a eigenvalue plot indicated that CLES scale satisfied the unidimensionality assumption and reliable. The mean response of CLES scale items was 3,39 which reflected the fact that the items tended to have fairly moderate difficulty values overall. Regarding the item discrimination, there were 10 items with low discrimination values for CTT because their item dicsrimination values are less than 0,60 and .there was 3 items with a very low level for GRM analysis. Item discrimination parameters for each item ranged between 0,68 and 1,81. It should be noticed that item discriminations can be considered as generally adequate and efective at distinquishing among respondents based on their estimated trait levels. The discriminations from the graded response model and classical test theory analysis correlated highly (r=.96) with each other. As a result, it can be colcluded that item and test parameters were highly correlated and researchers could use two methods if their assumpions were met and limitations were considerd.

**Key Words:** Graded Response Model, Item Response Theory, Classical test Theory, Item Parameter Estimation.

* Yrd. Doç. Dr. Abant İzzet Baysal Üniversitesi, e-posta: i.alper.kose@gmail.com

## 1.GİRİŞ

Ölçme ve değerlendirme alanında çalışan araştırmacılar, değerlendirmelerin etkililiğini artırmak amacıyla sürekli olarak yeni yaklaşımlar geliştirmektedirler. Bu çalışmaların temel amacı bireylerin test maddelerine verdikleri tepkilerden en doğru ve kullanışlı bilgiler ortaya çıkarmaktır (Wu & Adams, 2006). Bu amaçları gerçekleştirmek için geliştirilen matematiksel modeller ve bilgisayar teknolojileri yardımıyla yeni kuramlar ortaya atılmaktadır.

Ölçme tarihinde, cevaplayıcının test performansı altında yatan örtük özelliği açıklamak üzere geliştirilmiş en temel kuram klasik test kuramıdır (KTK). KTK, testten alınan gözlenen puanı, gerçek puan ve ölçme hatası ile açıklayan basit bir kuramdır. Kuramın, birçok veri setleri tarafından karşılanabilen zayıf varsayımlarına rağmen, test geliştirme ve test puanlarının yorumlanmasını gerektiren geniş uygulamalarda kullanılmaktadır (Hambleton and Swaminathan, 1989). Lord ve Novick (1968)'in, daha sonra madde tepki kuramı (MTK) olarak bilinen, zihinsel test puanlarını açıklayan istatistiksel yaklaşımlarına kadar, KTK test puanlarının açıklanması ve yorumlanması ile ilgili başat kuram olmaya devam etmiştir (Sijtsma and Junker, 2006; Seungho-Yang, 2007).

Örtük özelikler kuramı olarak da bilinen MTK, geçtiğimiz yüzyılın en önemli gelişmelerinden birisidir. MTK, cevaplayıcının yetenek düzeyini, test maddelerine verdiği cevapları kullanarak, KTK'nın zayıf varsayımlarına karşı ortaya koyduğu güçlü varsayımlarla matematiksek modeller yardımıyla açıklayan modern test kuramıdır (Bobcock, 2009). Embretson and Reise (2000)'a göre MTK, bireylerin maddelere verdiği yanıtlarla ilgili önermeler veya olasılıklı matematiksel modeller yardımıyla bireyin örtük özellikteki yerini açıklayan modern test kuramıdır.

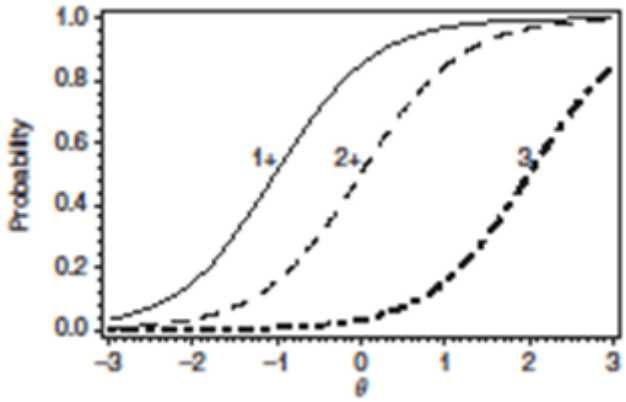
MTK'nın tarihçesini incelediğimizde Thurstone ve arkadaşlarının (1925; aktaran; Bock, Thissen & Zimowski, 1997) Binet zeka testindeki görevlerin ölçeklendirilmesi çalışmalarına dayandığı görülmektedir. Lord'un 1952 yılındaki MTK'nın 1-0 şeklinde puanlanan başarı testlerine uygulanmasıyla öğrenci başarısının ölçülmesinde kullanılması yaygınlaşmıştır (Lee, 1995). Bu tarihten sonra MTK'nın davranışlardaki bireysel farklılıkların altında yatan psikolojik yapıları açıklamak üzere kullanılmaya başlandığı görülmektedir (Bock, Thissen & Zimowski, 1997).

MTK'nın KTK üzerinde önemli avantajları bulunmaktadır. Bunlar; (a) madde parametreleri ve örtük özellik düzeyleri birbirinden bağımsızdır, (b) model, gözlenen test puanları düzeyinde değil, gözlenen madde tepkileri düzeyinde açıklanır, (c) her bir maddenin ölçeğe katkısı bilgi fonksiyonları yardımıyla belirlenir, (d) evrenin alt grupları arasında değişen madde fonksiyonu veya madde yanlılığı analizlerine imkan veren güçlü yöntemleri barındırır, (e) testlerin kısa veya paralel formları kolaylıkla oluşturulabilir, (f) cevaplayıcı puanları farklı sorulara cevap verseler bile eşitlenebilir (Uttaro & Lehman, 1999).

Tutum, kişilik gibi psikolojik özellikleri ölçen birçok ölçme aracı çoklu puanlanan maddelere sahiptir. Araştırmacılar bu maddeleri, daha güvenilir ve daha fazla bilgi vermeleri sebebiyle bu ölçeklerde tercih etmektedirler. Bu türde puanlanan maddeler için, çoklu puanlanan madde tepki modelleri geliştirilmiştir. Bu modeller cevaplayıcının yetenek düzeyi ile belli bir kategoride tepki verme arasında doğrusal olmayan ilişkiler kuran modellerdir (Emretson & Reise, 2000). Alanyazında Samejima (1969)'nın aşamalı tepki modeli olarak bilinen model, örtük özellik sürekliliğinde olan sıralı tepki modelleri

için geliştirilmiş olup, MTK'da 2 parametreli ikili tepki modellerinin karşılığıdır (Reise & Yu, 1990). Aşamalı tepki modelinin en büyük avantajı, ikili puanlanan modellere kıyasla bireylerin yeteneği hakkında daha fazla bilgi alınabilmesidir (Koch, 1983).

Çoklu puanlanan maddeler ikili puanlanan maddelere benzer olarak kategorik maddelerdir. Bir başka ifade ile, bu maddeler ikiden fazla tepki kategorisine sahiptir. Bu sıralı kategoriler, kategorileri ayıran sınır (boundary) veya eşik (treshold) değerleri ile ifade edilirler. Normal olarak, her zaman, kategori sayısının bir eksiği kadar eşik değeri vardır. Örneğin, 4'li likert tipi bir maddenin bu beş kategoriyi birbirinden ayıran 3 tane eşik değeri vardır (Ostini & Nering, 2006).




*Şekil 1.* Kategori eşik fonksiyonları

Madde karakteristik eğrisi çoklu puanlanan MTK modellerinde her bir kategori için ayrı ayrı oluşturulur. Bu durumun bir örneği 0-3 arasında puanlanan bir madde için şekil 1'de gösterilmiştir. Bu fonksiyonlardan, herhangi bir x kategorisinde tepki verme olasılığı, x veya daha üzeri kategoride seçme olasılığından yararlanarak oluşturulur (Demars, 2010).

ATM'de her bir madde iki parametre ile tanımlanır. Bunlardan birincisi madde ayırıcılık parametresidir (item discrimination parameter). Madde ayırıcılık parametresi, örtük özelliğin bir fonksiyonu olarak kategorilerdeki tepki olasılığının değişme gücü olarak tanımlanabilir (Rubio, Aguado, Hontangas and Hernandez, 2007). Baker (1985) madde ayırıcılık parametresinin yorumlanma aralıklarını;

| | |
|---|---|
| $a_i<0,20$ | Çok düşük ayırıcılık |
| $0,21<a_i<0,40$ | Düşük ayırıcılık |
| $0,41<a_i<0,80$ | Orta düzeyde ayırıcılık |
| $0,81<a_i<1.00$ | Yüksek ayırıcılık |
| $a_i>1.00$ | Çok yüksek ayırıcılık |

olarak belirtmiştir.

Maddeyi tanımlayan bir diğer parametre madde güçlük ya da lokasyon (yer) parametresidir. Madde güçlük parametresi, $b_i$, madde güçlüğünün ölçüsü, davranışın veya tutumun frekansı olarak bilgi veren parametredir (Rubio, Aguado, Hontangas and Hernandez, 2007). Aşamalı tepki modelinde ölçekteki her bir madde, (*i*), bir tane madde ayırıcılık parametresi, ($\alpha_i$) ve *j*=1,….$m_i$ tane kategori eşik parametresi ($\beta_{ij}$) ile tanımlanır. ATM cevaplayıcının belirli bir kategori ve daha üzerinde tepki verme olasılığı için m-1 tane eşik tepki fonksiyonu (boundary response function) tahmin eder. Örneğin, 5 kategori içeren bir madde için 4 tane eşik tepki fonksiyonu (ETF) oluşturulmaktadır. İlk ETF en düşük tepki seçeneğini seçme olasılığı için tanımlanır. İkinci ETF en düşük iki ve üzeri tepki verme olasılığını gösterir (Madera, 2003). ETF'nin formülü ikili tepki kategorisi için iki parametreli lojistik modele benzemektedir.

$$P_{ix}(\theta) = \frac{exp\left[\alpha_i\left(\theta-\beta_{ij}\right)\right]}{1+exp\left(\theta-\beta_{ij}\right)} \qquad (1)$$

Bu eşitlikte;

$\alpha$…..madde ayırıcılık parametresi,

$\beta_{ij}$…..kategori j için madde eşik parametresi,

$\theta$…..bireyin tahmin edilen yetenek düzeyini göstermektedir (LaHuis, Clarck & O'Brien, 2011).

Aşamalı tepki modeli alanyazında dolaylı MTK modeli olarak betimlenir. Bunun nedeni herhangi bir kategoride tepki verme olasılığı formül 1'den doğrudan hesaplanamamasıdır (Embretson & Reise, 2000). Bunun yerine her bir kategorideki tepki verme olasılığı aşağıdaki eşitlikler yardımıyla hesaplanabilir.

$$P_{i_1} = 1 - P^*_{i_1}(\theta)$$

$$P_{i_2} = P^*_{i_1}(\theta) - P^*_{i_2}(\theta)$$
$$P_{i_3} = P^*_{i_2}(\theta) - P^*_{i_3}(\theta)$$
$$P_{i_4} = P^*_{i_3}(\theta) - P^*_{i_4}(\theta)$$
$$P_{i_5} = P^*_{i_4}(\theta) - 0$$

MTK güçlü bir test kuramı olmasına rağmen model veri uyumunun sağlanmasını gerektirir. Ancak alanyazında kabul görmüş bir model veri uyumu testi bulunmamaktadır. (Gray-Little et al., 1997). Parametre tahminlerindeki tekrar-döndürme (iteration) sayısı, kestirilen parametrelerin hata miktarı, parametre değişmezliği model veri uyumu hakkında bilgi verebilecek yöntemlerdir (Rubio et al., 2007).

Ölçme araçları geliştirilirken soru bankalarının oluşturulmasında en önemli göstergelerden birisi de test-madde bilgi fonksiyonudur (Uttaro & Lehman, 1999). Test bilgi fonksiyonu, ölçekte bulunan maddelerin bilgi fonksiyonları toplamına eşit olan değerdir. Matematiksel olarak;

$$I(\theta) = \sum_{i}^{n} I_i(\theta)$$

eşitliği ile ifade edilir. Eşitlikte $I(\theta)$ test bilgi fonksiyonunu, $I_i(\theta)$ madde bilgi fonksiyonunu temsil etmektedir (Ostini & Nering, 2006). Madde ne kadar büyük ayırıcılık değerine sahip olursa, daha küçük hatalı yetenek kestirimi ve daha fazla madde bilgi düzeyine neden olacaktır. Bilgi, testte bulunan diğer maddelerden bağımsızdır ve yetenek sürekliliğinde, her bir yetenek düzeyinde ayrı ayrı kestirilir. Ölçekte bulunan maddelerin, bilgi fonksiyonları toplamalı özellik gösterdiğinden, test geliştirme sürecinde test bilgi fonksiyonuna en fazla katkı yapacak maddelerin seçimine olanak sağlamaktadır. Klaisk kuramdaki güvenirlik ile MTK'daki bilgi fonksiyonu arasındaki en önemli fark, güvenirlik örnekleme bağlı iken, bilgi fonksiyonları için bu durum örneklemden bağımsızdır. Bu özellik MTK'nın KTK üzerindeki en önemli avantajlarından birisidir (Uttaro & Lehman, 1999).

Alanyazın incelendiğinde ATM'nin uygulamasına yönelik birçok araştırma bulunmaktadır. ATM'nin tutumların ölçülmesine yönelik uygulaması (Madera, 2003; Marie, 1997; Chow & Winzer, 1992; Tezbaşaran & Kelecioğlu, 2004; Koch, 1983), Likert tipi ölçeklere uygulanması (Lee, 1995; Uttaro & Lehman, 1999; Keith, 1983; Rubio, at al, 2007; Sukirno & Siengthai 2010; Mielenz ve diğerleri 2010), yapay veriler üzerine uygulanması (Reise & Yu, 1990; LaHuis, et al, 2011; Muraki, 1990; Roberts & Laughlin, 1996), performansın ölçülmesine yönelik (De Ayala et al, 1989; Stoen et al, 1995) uygulamalar örnek olarak verilebilir.

Öğrenci tutumlarının değerlendirilmesi ölçeği maddeleri ATM altında Madera (2003) tarafından incelenmiş ve KTK ile ATM altında elde edilen parametreler arasında yüksek bir ilişki bulunmuştur. Koch (1983) bu bulguyu farklı bir tutum ölçeğinde, Uttaro ve Lehman (1999) Yaşam Kalitesi ölçeğinde, Sukirno & Siengthai 2010 meslek doyumu ölçeği üzerinde desteklemiştir. Mielenz ve diğerleri (2010) özyeterlik ölçeği verilerini ATM altında incelemiş, ölçeğin özellikle uç yetenek düzeyindeki bireylere ait bilgi fonksiyonunun daha az olduğunu veya daha hatalı kestirimlerde bulunduğunu ortaya koymuştur.

Klasik test kuramı test ve madde analizinde uzun yıllar kullanılan başat kuram olmasına rağmen, gerektirdiği zayıf varsayımların yanında test ve madde parametrelerinin uygulandığı örnekleme ve testte bulunan maddelere bağlı olması önemli sınırlılıklarındandır. Madde tepki kuramı ve bu kuramın dereceli puanlanan maddeler için uzantısı olan Aşamalı tepki modeli ile KTK'nın bu sınırlılıklarını gidererek daha kesin, testin uygulandığı örneklemden ve testte bulunan maddelerden bağımsız kestirimler yapıldığı iddia edilmektedir. Bu çalışma ile her iki kuram altında elde edilen test ve madde parametreleri karşılaştırılacak, daha az hatalı, daha geçerli ve güvenilir kestirimlerin hangi kuramdan geldiği ortaya konmaya çalışılacaktır. Elde edilen bilgiler ışığında hangi kuramın tercih edilebileceği araştırmacılara sunulacaktır.

## 2. YÖNTEM

### 2.1.Araştırma Grubu

Yapılandırmacı Öğrenme Ortamları Ölçeği Milli Eğitim Bakanlığına bağlı ortaokullarda okuyan 325 öğrenciye gönüllülük esasına göre uygulanmıştır. Araştırmaya katılan öğrencilerin 205'i kız (%63), 120'si erkektir (%37).

### 2.2.Ölçme Aracı

Bu araştırmada veri toplama aracı olarak Yapılandırmacı Öğrenme Ortamları Ölçeği kullanılmıştır. Ölçek (Arkün& Aşkar, 2010) tarafından geliştirilmiş olup 28 maddeden oluşmaktadır. Ölçme aracı derslerdeki yapılandırmacı öğrenme ortamlarına ilişkin algıları ölçmek amacıyla geliştirilmiştir. Ölçekteki maddeler 5 noktada puanlanmakta olup 1, öğrenme ortamını değerlendiren madde için en olumsuz algıyı, 5 ise en olumlu algıyı oluşturmaktadır. Ölçek 6 alt boyuttan oluşmakta olup, ölçekten en düşük 28, en yüksek 140 puan alınmaktadır. Ölçekten alınan yüksek puanlar öğrenme ortamlarının yapılandırmacı yaklaşıma uygun olduğunu göstermektedir.

### 2.3.İşlem

Ölçek araştırmacı tarafından gerekli izinler alınarak uygulanmıştır. Uygulama öncesinde araştırmanın amacı öğretmenlere ve öğrencilere açıklanmış konu ile ilgili sorular cevaplanmıştır. Ölçeğin uygulaması 10-15 dakika almaktadır. Uygulama sonucunda toplanan veriler SPSS 15.0 programına aktarılmış ve veriler kontrol edilmiştir. Kontrol işlemi sonucunda, 10 öğrencinin ölçeği eksik doldurmasından dolayı, 8 öğrencinin ise örüntülü doldurma nedeniyle cevapları veri grubundan çıkarılmıştır.

### 2.4.Verilerin Analizi

Yapılandırmacı öğrenme ortamları ölçeğinden elde edilen verilerin analizinde KTK altında yapılan analizler SPSS 15.0 programı, ATM altında yapılan analizler MULTILOG programı ile analiz edilmiştir. KTK altında, madde ayırıcılık parametresi için, madde toplam korelasyonlarından, madde güçlük parametresi için madde ortalamalarından yararlanılmıştır. Ölçeğin iç tutarlık anlamında güvenirliği için alfa katsayısı esas alınmıştır. ATM altında madde ve test parametrelerinin kestirimi için Smejima (1969)'ın aşamalı tepki modeli kullanılmıştır. Bu seçimin nedenleri, (1) sıralı tepki kategorileri için ATM ilk ve en çok kullanılan MTK modelidir, (2) Farklı ayırıcılık parametrelerine sahip maddeleri barındıran ölçekler için uygundur, (Rubio, et al., 2007). Güvenirlik kanıtları için ise marjinal güvenirlik kestirimi kullanılmıştır.

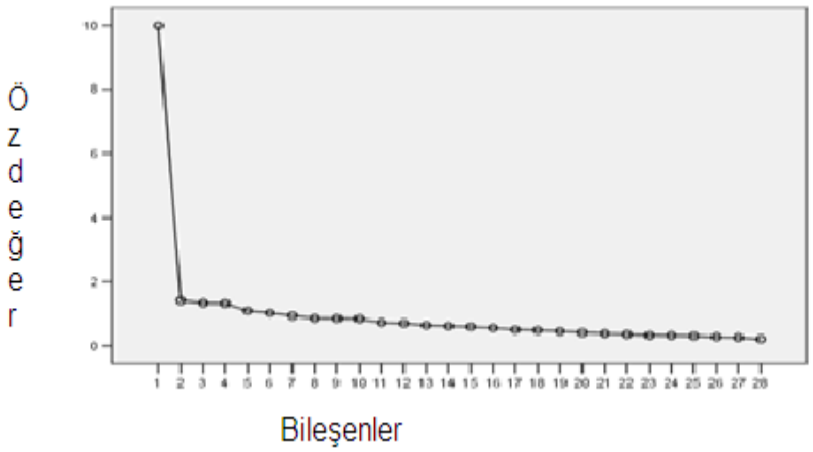
## 3.BULGULAR

Araştırmanın amacına uygun olarak öncelikle KTK'nda ölçekteki maddelerin madde güçlük ve ayırt edicilik değerleri hesaplanmıştır. Dereceli puanlanan maddelerde madde güçlüğü tepkilerin ortalaması olarak kabul edilir. Örneğin 5 noktada tepki verilebilen bir madde için madde güçlük değeri 5 ise maddenin çok kolay olduğunu, 1 ise çok kolay olduğunu ifade etmektedir. KTK'da madde ayırt edicilik değerleri ise, madde toplam korelasyonlarına karşılık gelmektedir. Yüksek ayırt ediciliğe sahip maddeler, ölçeğin ölçtüğü özelliği temsil ettiğini göstermektedir. Ölçeğin iç tutarlık katsayısı alfa katsayısı ile hesaplanmıştır.

KTK altında analiz edilen 28 maddenin madde güçlük ve madde ayırt edicilik değerleri Tablo 1'de özetlenmiştir. Tablo 1 incelendiğinde 10 maddenin (1-3-7-8-9-10-20-21-27-28) madde ayırt edicilik değerleri 0,60'dan düşük çıkmıştır. KTK'da Koch (1983)'a göre madde ayırt edicilik değeri 0.60'ın altında olan maddeler ölçekten çıkarılmalıdır. Bunun nedeni bu maddelerin ölçeğin ölçtüğü özelliği temsil etmede yetersiz olarak nitelendirilmesidir. Ölçekteki maddelerin ortalama güçlüğü 3,39 olarak hesaplanmıştır.

Bu değer ortalama güçlüğe karşılık gelmektedir. Ölçeğin iç tutarlık anlamında hesaplanan alfa katsayısı ise 0,93 olarak belirlenmiştir.

MTK, KTK üzerinde önemli avantajları olmasına rağmen, MTK'nın gücü tekboyutluluk varsayımının karşılanmasına dayanmaktadır. Veri grubunun yapısı veya boyutluluğu özdeğer grafiği ile ortaya konabilir. Özdeğer grafiği özellikle başat faktörün varlığını ortaya çıkarmada etkili bir yöntemdir (Hambleton & Swaminathan, 1985). Veri grubundaki başat boyut olduğuna, ilk faktörün açıkladığı varyansın %20 ve üzerinde olduğu durumda karar verilebilir (Lee, 1995). Şekil 2 özdeğer grafiği ile veri grubunun faktör yapısını göstermektedir. İlk faktör, ikinci faktörden 5-6 kat daha büyük özdeğere sahiptir ve bu başat faktör toplam değişkenliğin % 35,76'sını açıklamaktadır. Bu bulgular veri grubunun tek boyutlu olduğunu desteklemektedir.




*Şekil 2.* Özdeğer grafiği

Yapılandırmacı öğrenme ortamları ölçeğinden elde edilen veriler ATM altında madde ve test parametrelerinin kestirilmesi için MULTILOG (Thissen, 1988) programıyla analiz edilmiştir. 28 madde için kestirilen bir madde ayırıcılık parametresi $a_i$, ve dört madde güçlük/eşik parametreleri Tablo 1'de özetlenmiştir. Madde ayırıcılık parametresi maddenin örtük özellik sürekliliğinde cevaplayıcıları yeteneklerine göre ayırabilme gücü olarak ifade edilir. Yüksek ayırıcılığa sahip maddeler bireyleri yeteneklerine göre daha iyi ayıran maddelerdir, düşük ayırıcılığa sahip maddeler ise bireyleri yetenek düzeyine göre ayırmada yetersiz maddelerdir. Bunun yanında yüksek ayrıcılığa sahip maddeler test bilgi fonksiyonuna daha fazla katkı yapan maddelerdir (Hambleton at al., 1991).

Tablo 1'deki 28 maddenin ayırıcılık değerleri incelendiğinde, düşük ayırıcılığa sahip madde bulunmadığı, ayırıcılık değerlerinin 0,68 ve 1,81 arasında değiştiği gözlenmiştir. Ölçekte bulunan maddelerden 7, 21 ve 28. maddeler hariç, diğer maddelerin çok yüksek ayırıcılılığa sahip olduğu ifade edilebilir. Klasik test kuramı ve aşamalı tepki modeli altında elde edilen madde ayırıcılık parametreleri arasındaki ilişki test edilmiş ve r=0,96 (p<.01)'lık yüksek bir ilişki bulunmuştur. Her iki kuram altında hesaplanan bu ayırıcılık değerlerinin yüksek olması beklenen bir durumdur. Bunun nedeni olarak, ölçekteki maddelerin başat bir boyut altında toplanması gösterilebilir.

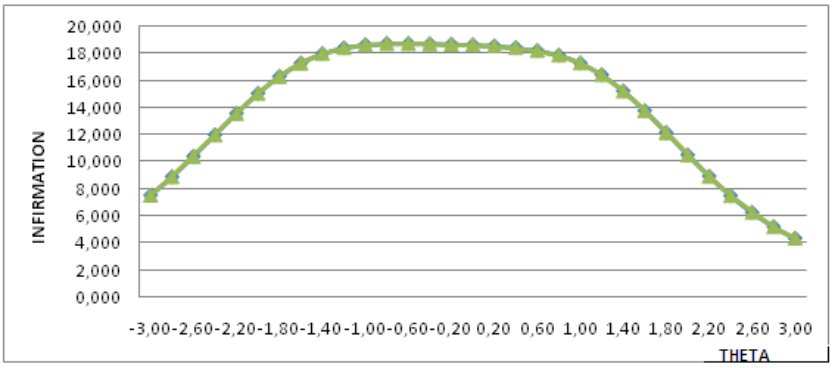
**Tablo 1.**
*KTK ve ATM altında kestirilen madde parametreleri*

| | KTK | | ATM | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Madde | α | b | α | SE | $\beta_1$ | SE | $\beta_2$ | SE | $\beta_3$ | SE | $\beta_4$ | SE |
| 1 | **0,456** | 3,27 | **1,07** | 0,14 | **-2,23** | 0,36 | **-1,33** | 0,24 | **0,57** | 0,18 | **1,75** | 0,30 |
| 2 | 0,602 | 3,59 | **1,46** | 0,17 | **-2,06** | 0,30 | **-1,37** | 0,20 | **-0,09** | 0,12 | **0,99** | 0,16 |
| 3 | **0,562** | 3,44 | **1,16** | 0,16 | **-1,80** | 0,30 | **-1,25** | 0,23 | **0,05** | 0,15 | **1,13** | 0,20 |
| 4 | 0,695 | 3,38 | **1,81** | 0,21 | **-1,45** | 0,20 | **-0,79** | 0,13 | **0,09** | 0,10 | **1,02** | 0,14 |
| 5 | 0,645 | 3,38 | **1,76** | 0,18 | **-1,24** | 0,18 | **-0,70** | 0,13 | **0,09** | 0,11 | **0,82** | 0,13 |
| 6 | 0,701 | 3,49 | **1,86** | 0,22 | **-1,43** | 0,19 | **-0,96** | 0,16 | **-0,12** | 0,10 | **1,04** | 0,13 |
| 7 | **0,425** | 3,39 | **0,87** | 0,14 | **-2,39** | 0,48 | **-1,41** | 0,32 | **0,12** | 0,20 | **1,39** | 0,30 |
| 8 | **0,568** | 3,47 | **1,18** | 0,17 | **-1,93** | 0,30 | **-1,04** | 0,21 | **-0,10** | 0,15 | **1,08** | 0,21 |
| 9 | **0,565** | 3,43 | **1,45** | 0,18 | **-1,49** | 0,24 | **-0,81** | 0,17 | **-0,06** | 0,13 | **0,95** | 0,18 |
| 10 | **0,588** | 3,31 | **1,43** | 0,17 | **-1,79** | 0,26 | **-0,94** | 0,16 | **0,32** | 0,14 | **1,41** | 0,20 |
| 11 | 0,629 | 3,22 | **1,54** | 0,18 | **-1,36** | 0,21 | **-0,72** | 0,15 | **0,28** | 0,12 | **1,26** | 0,17 |
| 12 | 0,619 | 3,29 | **1,48** | 0,19 | **-1,55** | 0,23 | **-0,71** | 0,16 | **0,17** | 0,12 | **1,19** | 0,18 |
| 13 | 0,620 | 3,24 | **1,51** | 0,17 | **-1,46** | 0,21 | **-0,74** | 0,14 | **0,29** | 0,12 | **1,19** | 0,18 |
| 14 | 0,605 | 3,19 | **1,33** | 0,19 | **-1,69** | 0,26 | **-0,74** | 0,17 | **0,41** | 0,15 | **1,36** | 0,20 |
| 15 | 0,678 | 3,40 | **1,72** | 0,20 | **-1,65** | 0,22 | **-0,80** | 0,14 | **0,05** | 0,11 | **1,08** | 0,15 |
| 16 | 0,624 | 3,49 | **1,53** | 0,18 | **-2,10** | 0,29 | **-0,89** | 0,15 | **0,05** | 0,13 | **0,92** | 0,16 |
| 17 | 0,612 | 3,32 | **1,42** | 0,18 | **-1,81** | 0,27 | **-0,81** | 0,17 | **0,30** | 0,14 | **1,19** | 0,18 |
| 18 | 0,600 | 3,43 | **1,43** | 0,17 | **-1,77** | 0,27 | **-0,86** | 0,16 | **0,00** | 0,12 | **1,07** | 0,17 |
| 19 | 0,605 | 3,52 | **1,39** | 0,17 | **-1,92** | 0,27 | **-1,09** | 0,19 | **-0,04** | 0,13 | **0,91** | 0,16 |
| 20 | **0,563** | 3,51 | **1,18** | 0,17 | **-2,16** | 0,33 | **-1,33** | 0,25 | **0,00** | 0,15 | **1,19** | 0,22 |
| 21 | **0,390** | 3,35 | **0,68** | 0,14 | **-2,95** | 0,66 | **-1,66** | 0,44 | **0,19** | 0,24 | **2,02** | 0,45 |
| 22 | 0,602 | 3,37 | **1,38** | 0,18 | **-2,07** | 0,28 | **-1,02** | 0,19 | **0,22** | 0,13 | **1,36** | 0,20 |
| 23 | 0,649 | 3,32 | **1,61** | 0,21 | **-1,72** | 0,25 | **-0,95** | 0,17 | **0,34** | 0,11 | **1,30** | 0,18 |
| 24 | 0,669 | 3,44 | **1,68** | 0,21 | **-1,66** | 0,22 | **-0,80** | 0,15 | **0,07** | 0,12 | **0,85** | 0,13 |
| 25 | 0,664 | 3,40 | **1,72** | 0,20 | **-1,65** | 0,23 | **-0,81** | 0,14 | **0,05** | 0,11 | **1,13** | 0,15 |
| 26 | 0,618 | 3,32 | **1,41** | 0,19 | **-1,66** | 0,26 | **-0,82** | 0,16 | **0,12** | 0,13 | **1,24** | 0,19 |
| 27 | **0,568** | 3,43 | **1,33** | 0,19 | **-1,73** | 0,28 | **-0,99** | 0,20 | **0,03** | 0,14 | **0,91** | 0,17 |
| 28 | **0,450** | 3,52 | **0,87** | 0,14 | **-2,18** | 0,44 | **-1,34** | 0,31 | **-0,30** | 0,21 | **1,02** | 0,26 |

Madde güçlük veya eşik parametre değerleri ATM altında incelendiğinde beklendiği gibi eşik parametreleri an düşükten en yükseğe doğru sıralanmıştır. Yani, her madde için $b_{i1}$ en düşük eşik parametresine, $b_{i4}$ en yüksek eşik parametresine sahiptir. Örnek olarak, 1. maddenin, $b_{11}$=-2,23 ve $b_{14}$=*1,7* olarak hesaplanmıştır. Bu maddenin 1. kategorisinde %50 olasılıkla tepki vermek için gerekli yetenek düzeyi θ=-2,23 olarak, 5. kategoride (en yüksek) %50 olasılıkla tepki vermek için gerekli yetenek düzeyi ise θ=1,75 olarak hesaplanmıştır. 21. maddede ise, 5. kategoride %50 olasılıkla tepki verme olasılığı ise en yüksektir (θ=2,02).

Madde parametrelerinin değişmezlik özelliğinin test edilmesi için, örneklem tesadüfi olarak ikiye ayrılmış ve madde parametreleri bu gruplar üzerinden kestirilmiştir. Madde ayırıcılık parametreleri arasındaki korelasyon r=0,82 (p<.01) olarak hesaplanmıştır. Madde eşik parametreleri arasındaki korelasyon ise, $b_{i1}$ için r= 0,84, $b_{i2}$ için r= 0,87, $b_{i3}$ için r= 0,79 ve $b_{i4}$ için r=0,76 (p<.01)'dir. Bu bulgular ölçekte bulunan maddelerin değişmezlik özelliğinin bulunduğunu ve ATM'nin veri grubuna uygunluğunu desteklemektedir.

Ölçekte bulunan maddelerin test bilgi fonksiyonuna yaptıkları katkılar hesaplanmış ve test bilgi fonksiyonu Şekil 3'de gösterilmiştir. Şekil 3 incelendiğinde örtük özellik skalasında ölçeğin orta yetenek düzeyine hitap ettiği, en yüksek ayırıcılığa, en yüksek bilgiye ve en düşük hataya θ=-1,40 ve θ=+1,40 aralığında ulaştığı görülmektedir. Buna karşın, uç yetenek düzeyindeki bireylerin ölçekteki maddelere verdikleri tepkilerin hata miktarlarının arttığı görülmektedir. ATM altında hesaplanan marjinal güvenirlik değeri 0,94 olarak hesaplanmıştır.




*Şekil 3*. Test bilgi fonksiyonu

## 4.TARTIŞMA ve SONUÇ

Bu çalışmanın temel amacı KTK ve ATM altında kestirilen madde ve test parametrelerinin karşılaştırılarak, araştırmacılara hangi kuram altında kestirilen parametrelerin daha keskin ve güvenilir olduğunu sunmaktır. Çalışma iki kuram altında bir dizi karşılaştırma esasına dayanmaktadır. Madde tepki kuramı altında test ve madde parametrelerinin kestirilebilmesi için veri grubunun tek boyutluluk varsayımını karşılaması gerekmektedir. Bu amaçla veri grubuna açımlayıcı faktör analizi yapılmış ve bulgular özdeğer grafiği ile özetlenmiştir. Özdeğer grafiği ve birinci faktörün açıkladığı varyans (% 35,76) veri grubunun tek boyutluluk varsayımını karşıladığını desteklemektedir. Bu bulgular Arkün & Aşkar (2010)'ın bulguları ile farklılık göstermektedir. Söz konusu çalışmada ölçek üniversite öğrencilerine uygulanmış olup 6 alt boyut açımlayıcı ve doğrulayıcı faktör analizi ile desteklenmiştir. Bu çalışmada ölçek ortaokul öğrencilerine uygulanmış olup farklı bir faktör yapısına ulaşılmıştır.

Her iki kuram altında kestirilen madde ayırıcılık parametreleri incelendiğinde KTK altında kestirilen 10 maddenin Koch (1983)'un 0,60 kriterini karşılamadığı ve ölçekten çıkarılması gerektiği sonucuna varılmıştır. ATM altındaki madde ayırıcılık parametreleri incelendiğinde ise 3 madde hariç (7-21-28), maddelerin yüksek ayırıcılığa sahip olduğu, test bilgi fonksiyonuna yaptıkları katkının yüksek olduğu sonucuna varılmıştır. Bunun yanında madde ayırıcılık parametrelerinin 0,68 ve 1,81 arasında değişim gösterdiği, bir başka ifade ile maddelerin farklı yetenek düzeyindeki bireylere hitap edebildiği sonucuna ulaşılmıştır. İki kuram altında kestirilen madde ayırıcılık parametreleri arasındaki korelasyonun 0,96 çıkması, kuramlar arasında kestirilen parametreler arasındaki uyuma işaret etmektedir. Bu ulaşılan yüksek ilişki Koch (1983), Sukirno & Sununta (2010) ve Madera (2003)'un çalışmasını destekler niteliktedir.

Madde eşik parametreleri incelendiğine KTK altında 1, ATM altında 4 eşik parametresi kestirilmiştir. ATM altında her bir kategori için kestirilen eşik parametrelerine hitap eden yetenek düzeyleri ifade edilebilmektedir. Her iki kuram altında elde edilen güvenirlik değerleri karşılaştırıldığında, KTK altında 0,93, ATM altında 0,94'lük değerler elde edilmiştir. Bu benzer güvenirlik değerleri Uttaro ve Lehman (1999)'un çalışmasıyla paralellik göstermektedir.

Test ve madde bilgi fonksiyonlarında ATM araştırmacılara daha büyük bir bakış açısı sağlamaktadır. Grafiksel gösterim olanağı ATM'nin bir diğer avantajlarındandır. Bu çalışmada ATM altında yapılan analizlerde ölçeğin -1,40 ile +1,40 arasında geniş bir yetenek düzeyine hitap ettiği görülmüştür. Buna karşın uç yetenek düzeylerindeki bireylerde ölçeğin sağladığı bilgi düşmektedir. Mielenz ve diğerleri (2010) yapmış oldukları çalışmada özellikle $\theta>2,0$ olduğu durumlarda test bilgi fonksiyonunun düştüğünü ortaya koymuşlardır. Bu çalışmada elde edilen bulgu ile Mielenz ve diğerleri (2010)'in bulguları paralellik göstermektedir.

Ölçme ve değerlendirmenin temel amaçlarından birisi de bireylerin test maddelerine verdikleri tepkilerden en doğru ve kullanışlı bilgiler ortaya çıkarmaktır (Wu & Adams, 2006). Gelişen teknoloji ve artan bilgi birikimi araştırmacılara bu bilgileri ortaya çıkarmak için yeni olanaklar sağlamaktadır. Madde tepki kuramı ve çoklu puanlanan maddeler için uzantısı olan aşamalı tepki modeli de bu olanaklardan bir tanesidir. ATM araştırmacılara yeni bir perspektif sağlamasına rağmen sağlanması güç olan tek boyutluluk varsayımı yeniliklerin yanında sınırlılıkları da beraberinde getirmektedir. Bu çalışmada elde edilen sonuçlar incelendiğinde ATM altında yapılan kestirimler sayesinde madde performansı ve test parametreleri hakkında daha fazla bilgi edinilmesine rağmen, KTK altında ulaşılan sonuçlar arasındaki uyumun fazla olduğu görülmüştür. Ancak , ölçek geliştirme çalışmalarında madde hakkında daha fazla bilgi veren, test bilgi fonksiyonuna en fazla katkı yapan maddelerin seçiminde ATM ön plana çıkmaktadır. Tezbaşaran ve Kelecioğlu (2004)'nun çalışması ölçek geliştirmede ATM'nin önemini ön plana çıkarmaktadır. Bu katkının görsel olarak sunulabilmesi araştırmacılar için önemli bir seçenektir. MTK'nın büyük örneklemler gerektirmesi, karşılanması mutlak varsayımlara sahip olması araştırmacıları düşündürmelidir.

Bu çalışmada ATM ve KTK altında madde ve test parametreleri karşılaştırılmıştır. Gelecekteki araştırmalarda bu karşılaştırmanın farklı uzunluktaki ölçeklerde ve farklı örneklem büyüklüklerinde yapılması önerilmektedir. Bu çalışmada kullanılan ölçme aracı 5 noktada puanlanan bir ölçme aracıdır. 5'den az veya fazla noktada puanlanan ölçme araçlarında kuram karşılaştırmaların tekrarlanması araştırmacılara bir diğer öneridir.

**KAYNAKÇA**

Arkün, S. & Aşkar, P. (2010). The development of a scale on assessing constructivist learning environments. *HU Journal of Education, 39*, 32-43

Baker, F.B. (1985). *The basics of item response theory*. Portsmouth, NH:Heineman

Bobcock, B.G.E.(2009).*Estimating a Noncompensatory IRT Model Using a modified Metropolis algorithm.* Unpublished Doctoral Dissertation.The University of Minesota.

Bock, R. D.,Thissen, D. & Zimowski, M.F. (1997). IRT estimation of domain scores. *Journal of Educational Measurement, 34*, 197-211

Chow, P. & Winzer, M.M. (1992). Reliability and validity of a scale measuring attitudes toward mainstreaming. *Educational and Psychological Measurement, 52*, 223-228.

De Ayala, R.J., Dodd, B.G. & Koch, W.R. (1989). Acomparison of the graded response and partial credit models for assessing writing ability. Paper Presented at the Anuual Meeting of the National Council on Measurement in Education, San Francisco, CA.

Demars, C. (2010). *Item response theory. Understanding statistics measurement*. Oxford University Press.

Embretson, S.E. & Reise, S.P. (2000). *Item Response Theory for Psychologists.* Mahwah, NJ:Erlbaum.

Gray-Little, B., Williams, V.S.L. & Hancock, T.D. (1997). An item response theory analysis of the Rosenberg Self Esteem Scale. *Personality and Social Psychology Bulletin, 23*, 443-451.

Hambleton, R.K. & Swaminathan, H. (1985). *Item Response Theory.* Kluwer-Nijhoff Publishing. Boston-USA.

Hambleton, R.K. & Swaminathan, H. (1989). *Item Response Theory*. *Principles And Applications.* Kluwer-Nijhoff Publishing. Boston-USA.

Hambleton, R.K., Swaminathan, H.& Rogers, H.J. (1991). *Fundamentals of Item Response Theory.* Sage Publications, London.

LaHuis, D.M., Clark, P. & O'Brien, E. (2011). An examination of item response theory item fit indices fort he graded response model. *Organizational Reseach Methods, 14*(1), 10-23.

Lane, S., Stone, C.A., Ankenmann, R.D. & Liu, M. (1995). Examination of the assumptions and properties of the graded item response model: An example using a mathematics perfromance assessment. *Applied Measurement in Education, 8*(4), 313-340.

Lee, K.H. (1995). Application of the graded response model to the revised Tennessee self-concept scale: Unidimensionality, parameter invariance, and differential item functioning. Unpublished Doctoral Dissertation. University of Southern California.

Keith, P.R. (1983). Application of a graded response model to the assessment of job satisfaction. Unpublished Doctoral Dissertation. University of Illinois.

Koch, W.R. (1983). Likert scaling using the graded response latent trait model. *Applied Psychological Measurement, 7*(1), 15-32.

Madera, E.K. (2003). Application of the graded response model to the assessment of student attitudes. Unpublished Doctoral Dissertation. University of Toronto.

Marie, L.A. (1997). The application of item response theory to employee attitude survey. Unpublished Doctoral Dissertation. University of Connecticut.

Mielenz, T.J., Edwards, M.C. & Callahan, L.F. (2010). Item response theory analysis of two questionnaire measures of arthritis-related self efficacy beliefs from community based US samples. *Hindawi Publishing Corporation Arthritis*.

Muraki, E. (1990). Fitting a polytomous item response model to Likert type data. *Applied Psychological Measurement, 14*(1), 59-71.

Ostini, R. & Nering, M.L. (2006). *Polytomous Item Response Theory Models*. Sage Publications, Inc. California, USA.

Reise, S.P. & Yu, J. (1990). Parameter recovery in the graded response model using MULTİLOG. *Journal of Educational Measurement, 27* (2), 133-144.

Roberts, J.S. & Laughlin, J.E. (1996). The graded unfolding model: A unidimensional item response model for unfolding graded responses. Research Report, Educational Testing Service, Princeton.

Rubio, V.J., Aguado, D., Hontangas, P.M. & Hernandez, J.M. (2007). Psychometric properties of an emotional adjustment measure. *European Journal of Psychological Assessment, 23* (1), 39-46.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores (Psychometric monograph No. 17). Richmond, VA: Psychometric society. Retrieved from http://www.psychometrika.org/journal/online/MN17.pdf

Seungho Yang, M. A. (2007*). A Comparison of Unidimensional and Multidimensional RaschModels Using Parametrer Estimates and Fit İndices When Assumption of Unidimensionality is Violated.* Unpublished Doctoral Dissertation. The Ohio State University

Sijstma, K. &Hemker, B.T. (2000). A taxonomyof IRT models for ordering persons anditems using simple sum scores. *Journal of Educational and Behavioral Statistics, 25* (4), 391-415.

Sijstma, K. & Junker, B.W. (2006). Item response theory: Past performance, present developments and future expectatitons. *Behaviormetrika, 33* (1), 75-102.

Stone, C.A., Akenmann, R.D. & Liu, M. (1995). Examination of the assumptions and properties of the graded item response model: An example using a mathematics perfromance assessment. *Applied Measurement in Education, 8*(4), 313-340.

Sukirno, A. & Sununta, S. (2010). The comparison of graded response model and classical test theory in human resource research: A model fitness test. *Research and Practice in Human Resource Management, 18*(2), 77-86.

Tezbaşaran, A. & Kelecioğlu, H. (2004). Madde-ölçek korelasyonlarına, alt-üst grup ortalamalarına ve aşamalı tepki modeline göre geliştirilen sigaraya ilişkin

tutum ölçeğinin madde ve ölçek özelliklerinin incelenmesi. XIII. Ulusal Eğitim Bilimleri Kurultayı, İnönü Üniversitesi, Malatya.

Thissen, D. (1988). MULTILOG (Computer program). Mooresville, IN: Scientific Software.

Uttaro, T. & Lehman, A. (1999). Graded response modeling of the Quality of Life Interview. *Evaluation and Program Planning, 22*(1999), 41-52

Wu, M. & Adams, R. (2006). Modelling mathematics problem solving item responses using a multidimensional IRT model. *Mathematics Education Research Journal, 18*(2), 93-113.

## EXTENDED ABSTRACT

### 1.Introduction

Researchers in the field of educational assessment are continually developing new approaches to improve the efficiency of assessments. They are often concerned with methodologies that can extract the most useful and accurate information from students' responses to test items (Wu & Adams, 2006). With the help of improved mathematical models and computer technologies, new theories have been developing in the field of educational and psychological assessment.

In the measurement history, the leading theory to explain latent trait underlying exaimee's test performance is Classical Test Theory (CTT). CTT is a simple model which states that the oserved score on a test is the sum of the ture score and measurement error. CTT is based on weak assumptions, that is, the assumptions can be met easily by most data sets, and therefore, the models can and have been applied to a wide variety of test development and test score analysis problems (Hambleton and Swaminathan, 1989). One of the most imortant improvements the last century is IRT, also known as latent trait theory, in psychological measurement. IRT is a modern test theory which explains examinee's ability level by using responses to test items with strong assumptions against CTT's weak assumptions with mathematical models (Bobcock, 2009). Numerous measurement instruments, especially in the attitude and personality assessment domains, include items with multiple ordered-response categories are used. Researchers employ these formats for a variety of reasons but mainly because they are more informative and reliable than dichotomous scored items. For these multiple-category types of item-response data, polytomous IRT models are needed to represent the nonlinear relation between examinee trait level and the probability of responding in a particular category (Emretson & Reise, 2000).

Researchers in item response theory have concentrated on the implementation of binary statistical models for achievement and ability measurement. There has been minimal interest in exploring the potential of polytomous models. The graded model (Samejima, 1969), in contrast to the well established one,- two-, and three- parameter binary models, is appropriate when item responses can be ordered into more than two categories along an agree-disagree or high-low trait continuum (Reise & Yu, 1990). The graded response model is a generalization of the two-parameter model. A major feature of the graded response model was that more information about a person's ability or attitude level could be obtained for graded responses than for binary responses (Koch, 1983).

Although CTT has been used for item and test parameter analysis for a long time, sample dependencey for item and test parameter analiysis to sample is among CTT's crucial limitations. IRT and it's polytomous extension graded response model aims to overcome CTT's limitations and estiamte item and test parameters more precisely. With this study, item and test parameters were compared and submit the best model for researchers.

**2.Method**

The data used for the present research consisted of responses from a total sample of 307 secondary school students to a 28-item Likert scale. The instrument, Constructivist Learning Environment Scale (Arkün& Aşkar, 2010), was designed to measure the perceptions of students on constructivist learning environments for their courses. All of the items were scored on a 5-poit scale, with a score of 1 indicating an unfavorable response for constructivist environments and a score of 5 indicating a favorable response for constructivist environments.

**3.Results**

The results of the traditional item analysis, the difficulties and discrimination values of all 28 items from the CLES were estimated. 10 items (1-3-7-8-9-10-20-21-27-28) should be removed from the scale because their item dicsrimination values are less than 0,60. The criterion for item removal is item discriminations less than or equal to 0,60 (Koch, 1983). The mean response of CLES items was 3,39 which reflected the fact that the items tended to have fairly moderate difficulty values overall. The coefficient alpha reliability of the scale was 0,93.

Regarding the item discrimination in ATM, there were 3 items with a low level. Item discrimination parameters for each item ranged between 0,68 and 1,81. It should be noticed that item discriminations can be considered as generally adequate and efective at distinquishing among respondents based on their estimated trait levels. The discriminations from the graded response model and classical test theory analysis correlated highly (r=.96) with each other. The high correlation between the traditional and the GRM discriminations was expected because both reprsent the loadings of the items on the first factor of the data.

The results of the estimated treshold paarmeters showed that, as expected, these estimated tresholds were ordered from the lowest to highest, with $b_{i1}$ indicating lower and $b_{i4}$indicating higher attitude levels. For instance, in item 1,$b_{11}$=-2,23 and $b_{14}$*=1,7.* A very low trait level (-2,23) is required to have a probability of 0,5 in order to respond 1 (which is the lowest trait level) whereas a very high trait level (1,75) is needed to have a probability of 0,5 in order to respond on category 5 (the highest trait level). As a consequence, subjects with a very high trat level tend to show the maximum level respond in this item. Item 21 showed that a very high trait level ($b_{21,4}$=2,02) is required to have a probability of 0,5 in order to choose category 5 while a very low trait level ($b_{21,1}$=-2,95) is needed to have a probability of 0,5 to respond on category 1.

**4. Discussion**

As a result of findings of the present study, it can be concluded that item and test parameters were highly correlated and researchers could use two methods if their assumpions were met and limitations were considered.