

Grup Başarısına Göre Madde ve Kişi Parametreleri Arasındaki İlişkinin İncelenmesi*

The Effect of Group Properties to Item and Person Parameters

Ayfer SAYIN¹, Hakan KOĞAR²

¹Gazi Üniversitesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme A.B.D.
ayfersayin@yahoo.com

²Akdeniz Üniversitesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme
A.B.D. hkogar@gmail.com

Makalenin Geliş Tarihi: 01.06.2018

Yayına Kabul Tarihi: 03.10.2018

ÖZ

Bu çalışmada KTK ve MTK'ye dayalı olarak madde parametrelerinin grubun özelliğine (başarılı ve başarısız olma durumuna) ne düzeyde bağlı olduğunu belirlemek amaçlanmıştır. Aynı zamanda KTK ve MTK'dan elde edilen bulgular arasındaki ilişkiler de incelenmiştir. Türkiye'de bir devlet üniversitesinin eğitim fakültesinin farklı 22 bölümünde öğrenim görmekte olan toplam 1916 öğrenci bu araştırmanın çalışma grubunu oluşturmaktadır. Farklı özelliklerdeki veri setlerinden elde edilen madde parametreleri arasındaki ilişkiler ve KTK ve MTK'den elde edilen madde parametreleri arasındaki ilişkiler ayrı ayrı incelenmiştir. Ayrıca, KTK için toplam puanlar ile MTK için yetenek parametreleri arasındaki ilişkiler de farklı veri setleri için ayrı ayrı incelenmiştir. Genel olarak KTK'dan elde edilen nokta çift serili korelasyon katsayısı ile 2PL modelden elde edilen a parametresi arasındaki ilişkiler orta düzeyde ve anlamlıdır. Diğer ilişki katsayılarının ise düşük düzeyde olduğu ya da ilişkinin olmadığı durumların ortaya çıktığı belirlenmiştir. Ölçme kuramlarının farklı dağılım özellikleri ve nitelikleri bulunan veri setlerindeki bireylerin yeteneklerini benzer şekilde kestirdiği tespit edilmiştir. Farklı veri setlerinden elde edilen güçlük parametreleri arasında genellikle yüksek ilişkiler belirlenmiştir. Güçlük indeksleri arasındaki ilişkiler KTK'ya dayalı gerçekleştirilen kestirimlerde MTK'ya dayalı kestirimlere göre daha yüksek hesaplanmıştır. Bu durum, MTK kestirimlerinin gruba bağlı değişim gösterdiği anlamına gelebilmektedir. Benzer şekilde madde ayırt edicilik indekslerinin hem KTK hem de MTK'ya dayalı modellerle hesaplamaları arasında anlamlı farklılıklar olduğu tespit edilmiştir.

* **Alıntılama:** Sayın, A. ve Koğar, H. (2019). Grup başarısına göre madde ve kişi parametreleri arasındaki ilişkinin incelenmesi. *Gazi Üniversitesi Eğitim Fakültesi Dergisi*, 39(1), 249-270

Anahtar Sözcükler: Klasik test kuramı, Madde tepki kuramı, Grup özellikleri, Madde parametreleri, Kişi parametreleri

ABSTRACT

The aim of this study is to determine the extent to which item parameters estimated via CTT and IRT depend on group properties (being mater or non-master). Besides, the relationship between findings obtained from CTT and IRT is evaluated. The study group of this study consists of 1916 undergraduates who study at different departments (22) in educational faculty in a state university of Turkey. The relationship between item parameters obtained from datasets having different properties and the relationships between item parameters obtained via CTT and IRT have been investigated separately. Besides, the relationships between total scores for CTT and ability parameters for IRT have been examined for different datasets. The relationship between point biserial correlation obtained from CTT and a parameter obtained from 2PL is moderate and significant. It has been determined that the other relationship coefficients are low or non-significant. It has been determined that measurement theories estimate the ability of individuals in datasets having different distribution and properties similar. The high relationships have generally been obtained among difficulty parameters estimated from different datasets. This situation shows that the estimates obtained for CTT are group-dependent and the estimates for IRT are group-independent. Similarly, it has been observed that there is a significant difference between item discrimination indexes estimated for CTT and IRT.

Keywords: Classical test theory, Item response theory, Group properties, Item parameters, Person parameters

GİRİŞ

Klasik Test Kuramı (KTK), 20. yüzyılın büyük bir bölümünde kullanılan en önemli ölçme kuramlarından biridir. Bu kuram, az sayıda varsayım gerektirmesinden dolayı, diğer bir deyişle matematiksel işlemlerinin sağladığı kolaylıklar neticesinde birçok test uygulamasında kullanılmıştır (Hambleton & Jones, 2012). KTK test düzeyinde bilgi elde etmek için geliştirilmiş olmakla birlikte madde istatistikleri de zamanla önemli bir uygulama olarak ön plana çıkmıştır.

Madde düzeyindeki KTK uygulamaları oldukça basit düzeydedir. KTK, belirli bir madde ile katılımcının yeteneği arasındaki ilişkiyi tanımlarken karmaşık kuramsal modeller kurmaz. Bir grup katılımcının bir madde üzerindeki başarı oranını belirlemekle yetinir. Bu başarı oranı, en bilinen madde istatistiği olan madde güçlüğü

(p_{ij}) olarak tanımlanır. Bir maddenin yüksek yetenekli katılımcılar ile düşük yetenekli katılımcıları ayırt etmesi, bir başka madde istatistiği olan madde ayırt ediciliği (r_{ij}) ile belirlenmektedir. Madde ayırt ediciliği madde puanları ile test puanları arasındaki ilişki olarak tanımlanmaktadır (Fan, 1998).

KTK'de madde parametreleri gruba bağlı kestirilmektedir. Madde ve birey parametrelerinin gruba bağımlı kestirilmesi ve matematiksel altyapısının zayıflığına rağmen KTK günümüzde halen birçok ölçme uygulamasında kullanılmaktadır. Örneğin test eşitlemede KTK temelli yöntemlerin kullanımı pratik sebeplerden dolayı daha sıklıkla tercih edilmektedir. KTK, test geliştirme ve test uygulamaları konusunda oldukça pratik bir kuramdır (Awopeju ve Afolabi, 2016).

Madde Tepki Kuramının (MTK) temel varsayımların biri, madde parametrelerinin gruptan bağımsız bir şekilde kestirilmesine olanak vermesidir. Bu varsayımı ile KTK'den uygulamalarından farklılık göstermekte ve test geliştiricilere avantaj sağlamaktadır. MTK, KTK'nın varsayımlarındaki sınırlılıkları aşmak için geliştirilen ve bilgisayar sistemlerine uyarlanabilir matematiksel çözümler de sunan modern bir test kuramıdır. MTK, güçlü bir matematiksel temele ve bilgisayarlar tarafından çözülebilen karmaşık algoritmalara sahiptir. Madde karakteristik eğrisi (MKE) adı verilen katılımcının test performansı ile yeteneği arasında matematiksel bir fonksiyon kurulmasını sağlamaktadır (Hambleton & Swaminathan, 1985; Harris, 1989). MTK, KTK'dan farklı olarak test düzeyinde bilgiye değil, madde düzeyindeki bilgilere odaklanmaktadır. Kullanılan MTK modeline göre bir ya da birkaç parametre kullanılarak kişinin yeteneği ile performansı arasındaki ilişki açığa çıkarılmaya çalışılır.

KTK'ye dayalı olarak ortalama, standart sapma, güvenilirlik gibi test istatistikleri; madde güçlük indeksi, madde ayırıcılık indeksi, madde varyansı gibi madde istatistikleri hesaplanmaktadır. Lord ve Novick (1968) KTK ile gerçekleştirilen hesaplamaların gruba bağlı olduğunu, dolayısıyla parametrelerin, verilerin toplandığı gruptan gruba değişiklik gösterdiğini ve bu durumun da farklı gruplarda test istatistiklerinin karşılaştırılmasını zorlaştırdığını ifade etmektedir (Hambleton, Swaminathan ve Rogers, 1991). Bu bağlamda incelendiğinde bir grup için orta güçlükte olduğu belirlenen bir test

ya da madde, başka bir grup için zor ya da kolay hesaplanabilir ya da bir grup için güvenilir olan madde/test, başka bir grup için güvenilir olmayabilir (Crocker ve Algina, 1986). KTK ile sözü edilen sınırlandırmaların önüne geçmek, madde ve test parametrelerini gruptan bağımsız bir şekilde hesaplayabilmek için MTK geliştirilmiştir. Crocker ve Algina, (1986), MTK ile bireylerin yetenek düzeylerinin testten bağımsız bir şekilde kestirileceğini çünkü MTK'nin farklı yetenek düzeylerindeki bireylerin teste yer alan her bir maddeye nasıl tepki verdiklerine dair matematiksel bir modellemeye sahip olduğunu belirtmektedir. Başka bir anlatımla MTK, madde ve test istatistiklerinin gruptan, yetenek düzeylerinin ise testten bağımsız bir şekilde kestirilmesine olanak verdiğini belirten bir kuramdır. Büyük örneklem grupları ile kestirimlerin gerçekleştirildiği MTK ile kestirilen parametreler gruptan bağımsız olduğu için değişmezlik özelliğine sahiptir ve bunun sonucunda sonuçlara bakılarak farklı grupların karşılaştırılması mümkün olmaktadır (Embretson ve Reise, 2000).

Alan yazında KTK ve MTK'dan elde edilen madde parametrelerinin karşılaştırıldığı çalışmalar bulunmaktadır. Kuramlar arasındaki farklılığa rağmen iki kuramdan elde edilen madde istatistiklerinin pozitif bir ilişki gösterdiğine dair çeşitli araştırmalar bulunmaktadır (Adedoyin, Nenty & Chilisa, 2008; Çıkrıkçı-Demirtaşlı, 2002; Fan, 1998; Gelbal, 1994; MacDonald ve Paunonen, 2002; Royce, 2009). Örneğin Royce (2009) sözel ve sözel olmayan testler üzerinde yaptığı araştırmada, 2 parametrelili lojistik model (2PL) ile KTK'den elde edilen madde istatistiklerinin birbirine çok yakın olduğunu belirlemiştir. Bununla birlikte MacDonald ve Paunonen (2002) bu iki ölçme kuramını karşılaştırdığı araştırmasında daha önce elde edilen KTK ve MTK'nin madde ve birey bazındaki istatistiklerinin benzerliğini ifade eden araştırmaların tüm eğitim ve psikolojik testler için genellenemeyeceğini ifade etmişlerdir. Fan (1998), MacDonald ve Paunonen (2002) gibi araştırmacılar MTK'nin kuramsal üstünlüğüne rağmen KTK ile karşılaştırıldığında kayda değer bir farkın olmadığını; Lawson (1991), Skaggs ve Lissitz (1988), ve Stage (1999) gibi araştırmacılar ise MTK ve KTK'dan elde edilen parametre tahminleri arasında önemli farkların olduğunu belirlemişlerdir.

MTK ve KTK'dan elde edilen madde ayırt edicilik katsayıları arasında yüksek bir ilişki belirleyen araştırma bulguları, bu ilişkiyi yalnızca ayırt ediciliğin geniş bir dağılıma sahip olduğu ve madde güçlüğünün küçük değerler aldığı durumlarda ortaya koyabilmektedir (Progar, Sočan ve Peč, 2008). Bu durum, farklı özelliklerdeki gruplarda da KTK ve MTK'dan elde edilen madde istatistiklerinin değişmezliği sağlanabilir mi? sorusunu akla getirmektedir. Bu sorunun cevabına yönelik yapılan araştırmalar (Fan, 1998; MacDonald ve Paunonen, 2002; Progar, Sočan ve Peč, 2008) farklı dağılımlara ve farklı özelliklere sahip katılımcılara odaklanmaktadır. Bu araştırmaların elde edilen ortak bulgusu KTK ve MTK'dan elde edilen madde parametrelerinin değişmezliğinin birçok açıdan sağlandığı; özellikle madde güçlüğü parametresinin madde ayırt ediciliğine göre daha kararlı olarak elde edildiği şeklindedir. Ayrıca, MTK parametrelerinin değişmezliğinin ancak MTK'dan elde edilen model veri uyumunun yüksek olması durumunda gerçekleştiği de belirtilmiştir. Tüm bu bilgiler doğrultusunda bu araştırma, alan yazındaki araştırmaların eksiklerini giderme amacıyla daha geniş bir örneklem büyüklüğü, gerçek veriye dayalı analizler ve madde ve birey parametrelerinin birlikte kullanılması ile gerçekleştirilmiştir.

Bu çalışmada KTK ve MTK'ye dayalı olarak madde parametrelerinin grubun özelliğine (başarılı ve başarısız olma durumuna) ne düzeyde bağlı olduğunu belirlemek amaçlanmıştır. Aynı zamanda KTK ve MTK'dan elde edilen bulgular arasındaki ilişkiler de incelenmiştir.

YÖNTEM

Çalışma Grubu

Türkiye'de bir devlet üniversitesinin eğitim fakültesinin farklı 22 bölümünde öğrenim görmekte olan toplam 1916 öğrenci bu araştırmanın çalışma grubunu oluşturmaktadır. Araştırmaya İngilizce öğretmenliği (%16; n=308), sınıf öğretmenliği (%12; n=222), fen bilgisi öğretmenliği (%9; n=174), sosyal bilgiler öğretmenliği (%9; n=172), ilköğretim matematik öğretmenliği (%9; n=169), okul öncesi öğretmenliği (%8; n=151), resim-iş

öğretmenliği (%6; n=124), rehberlik ve psikolojik danışmanlık öğretmenliği (%3; n=59), bilgisayar ve teknolojileri öğretmenliği (%3; n=54), müzik öğretmenliği (%3; n=51), Almanca öğretmenliği (%3; n=51), görme engelliler öğretmenliği (%2; n=46), zihinsel engelliler öğretmenliği (%2; n=42), Fransızca öğretmenliği (%2; n=41), Arapça öğretmenliği (%2; n=36), matematik öğretmenliği (%2; n=34), biyoloji öğretmenliği (%2; n=33), kimya öğretmenliği (%2; n=33), fizik öğretmenliği (%2; n=30), tarih öğretmenliği (%2; n=30), felsefe grubu öğretmenliği (%2; n=29) ve coğrafya öğretmenliği (%1; n=27) bölümlerinde öğrenim görmekte olan öğrenciler katılmıştır.

Verilerin Toplanması

Araştırma kapsamında 2011-2012 eğitim öğretim döneminde üniversitenin eğitim fakültesinde öğrenim görmekte olan ve ortak derslerden biri olan Türkçe I: Yazılı Anlatım dersinin final sınavına katılan toplam 1940 öğrencinin cevaplarına ulaşılmıştır. Ancak 24 öğrencinin yabancı uyruklu olduğu ve sonuçları da uç değer gösterdiği için söz konusu öğrenciler veri setinden çıkarılmıştır. Ders, tüm bölümlerin birinci sınıf güz dönemine ait bir derstir. Fakültede toplam 22 bölümde örgün eğitim şeklinde gerçekleştirilen dersin yürütülmesinde, 10 farklı öğretim üyesi sorumlu olmuştur.

Final sınavının oluşturulmasında fakültede kurulan bir merkez aracılığıyla öğretim üyelerinden dersin kazanımlarını ve ağırlık düzeyini belirledikleri bir belirtke tablosu oluşturulmuştur. Belirtke tablosu doğrultusunda her bir kazanıma yönelik ikişer soru oluşturularak bir madde havuzu elde edilmiştir. Madde havuzundan rastgele sorular seçilmiş ve belirlenen sorular için öncelikle uzman görüşleri alınmış, ardından ölçme ve değerlendirme uzmanlarının görüşleri de alınarak gerekli düzeltmeler yapılmıştır. Yönergelerin oluşturulmasının ardından dört farklı kitapçık türü oluşturulmuştur. Öğrencilerin sınava girecekleri sınıfların belirlenmesi de merkezi bir şekilde gerçekleştirilmiş, öğrenciler kendi adlarına oluşturulan optik formlar ile sınava katılmışlardır. Uygulama için fakültenin 62 dersliği kullanılmıştır.

Türkçe I: Yazılı Anlatım dersinin final sınavında toplam 50 soru yer almaktadır. Sorulardan ikisi final sonrasında iptal edilmiştir. Soruların puanlaması 100 üzerinden

gerçekleştirilmiş ve sınavda yanlış cevapların doğru cevapları etkilemeyeceği öğrencilere belirtilmiştir.

Öğrencilerin cevapları üniversitenin öğrenci işleri daire başkanlığından alınmıştır. Toplam 48 sorudan oluşan test içinde araştırmanın amacı doğrultusunda veri setleri oluşturulmuştur.

Verilerin Analizi

Araştırmanın amacı doğrultusunda geliştirilen veri analizi deseni Tablo 1’de belirtilmiştir.

Tablo 1. Veri analizi deseni

Koşul	Tüm grup	Örneklem	KTK	MTK	İlişki
		N=1916	Güçlük-ayırıcılık	Güçlük-ayırıcılık	Korelasyon
Başarılı ve başarısız gruplar	Koşul-1: Kümeleme analizine göre	N ₁ =1198 N ₂ =718	Güçlük-ayırıcılık	Güçlük-ayırıcılık	Kendi içinde ve dışında ilişki
	Koşul-2: ortalama üstüne göre	N ₁ =1099 N ₂ =817	Güçlük-ayırıcılık	Güçlük-ayırıcılık	Kendi içinde ve dışında ilişki
	Koşul-3: mutlak başarıya göre (%75)	N ₁ =1376 N ₂ =540	Güçlük-ayırıcılık	Güçlük-ayırıcılık	Kendi içinde ve dışında ilişki

Bu araştırmanın amacı doğrultusunda 1940 kişilik çalışma grubundan elde edilen veri seti 4 farklı koşul altında incelenmiştir. Birinci koşul, tüm çalışma grubunun analize dâhil edildiği durumdur. İkinci koşulda çalışma grubu kümeleme analizi aracılığıyla ikiye bölünmüştür. Üçüncü koşulda çalışma grubu; ortalamanın üstünde ve altında olacak şekilde iki farklı gruba ayrılmıştır. Dördüncü koşulda ise çalışma grubu mutlak başarıya göre (%75’lik bir başarı oranını elde edenler ve elde edemeyenler) olarak ikiye

ayrılmıştır. İkinci, üçüncü ve dördüncü koşullarda her bir gruptan elde edilen sonuçlar kendi içinde ve birbirleri ile karşılaştırılarak incelenmiştir.

Veri setleri oluşturulduktan sonra her bir koşulda KTK ile madde ayırıcılık katsayısı için üst-alt ayırıcılık, nokta çift ve çift serili korelasyon olmak üzere üç farklı; madde güçlük katsayısı için ise üst-alt grup ve tüm grup olmak üzere 2 farklı güçlük parametresi elde edilmiştir. MTK'da ise 2PL, 3PL ve nominal response model için ayrı ayrı madde güçlük (b parametresi) ve madde ayırıcılık (a parametresi) elde edilmiştir.

Birinci koşulda KTK ve MTK'ye dayalı madde parametreleri arasındaki ilişkiler Pearson Momentler Çarpımı Korelasyon Katsayısı ile incelenmiştir. Ayrıca, KTK için toplam puanlar ile MTK için yetenek parametreleri arasındaki ilişkiler de incelenmiştir.

Birinci koşul için ayırt ediciliği yüksek ve düşük olan maddeler kendi içlerinde iki farklı grup oluşturularak, KTK ve MTK'den elde edilen bulguların ayırıcılık gücünden etkilenip etkilenmediği ayrıca incelenmiştir. Bu amaçla, ayırıcılığı düşük olan madde grubundan KTK ve MTK'dan elde edilen madde parametreleri ile ayırıcılığı yüksek olan madde grubundan KTK ve MTK'dan elde edilen madde parametreleri arasındaki ilişkiler incelenmiştir.

İkinci, üçüncü ve dördüncü koşulda ise farklı ölçütlere göre ayrılmış veri setlerinde, KTK ve MTK'ye dayalı madde parametreleri elde edilmiştir. Farklı özelliklerdeki veri setlerinden elde edilen madde parametreleri arasındaki ilişkiler ve KTK ve MTK'den elde edilen madde parametreleri arasındaki ilişkiler ayrı ayrı incelenmiştir. Ayrıca, KTK için toplam puanlar ile MTK için yetenek parametreleri arasındaki ilişkiler de farklı veri setleri için ayrı ayrı incelenmiştir.

Varsayımların İncelenmesi

Öncelikle MTK'nin normallik, tek boyutluluk, yerel bağımsızlık ve değişmezlik varsayımları incelenmiştir.

Normallik

Veri setinin normal dağılım gösterip göstermediğinin belirlenmesi amacıyla öncelikle çarpıklık ve basıklık katsayısı hesaplanmıştır, ardından histogram grafikleri incelenmiştir. Sonuçlar Tablo 1’de gösterilmiştir.

Tablo 1. Araştırma kapsamında oluşturulan veri setlerinin çarpıklık ve basıklık katsayıları

Veri setleri	N	Çarpıklık	SE	Basıklık	SE
Tüm grup	1916	-0,694	0,056	0,170	0,112
Koşul 1	1198	0,102	0,071	-0,523	0,141
	718	-0,759	0,091	0,034	0,182
Koşul 2	1099	0,421	0,074	-0,628	0,147
	817	-0,981	0,086	0,305	0,171
Koşul 3	1376	0,254	0,066	-0,774	0,132
	540	-0,930	0,105	0,159	0,210

Tablo 1’de yer alan bilgiler incelendiğinde verilerin evrene ait uzayda normal dağılım gösterdiği belirlenmiştir.

Tek boyutluluk

Türkçe I: Yazılı Anlatım final testi bir başarı testi olduğu için öncelikle kapsam geçerliği ile tek boyutluluk varsayımı incelenmiştir. Bu doğrultuda testte yer alan 48 soruya yönelik Türkçe eğitiminde doktorasını tamamlamış üç konu alan uzmanının görüşü alınmış, testte bulunan tüm maddelerin aynı özelliğini ölçtüğü uzmanlarca belirlenmiştir.

Uzman görüşlerinin ardından her bir veri setine yönelik tetrakorik korelasyona dayalı açıklayıcı faktör analizi hesaplanmıştır. Factor 10.4 programı ile analizler yapılmıştır. Ağırlıklandırılmamış en küçük kareler yöntemi ile yapılan AFA bulgularına göre KMO değeri 0.82, Bartlett küresellik testi sonuçları anlamlı olarak bulunmuştur. AFA sonucunda test maddelerinin iki boyutta toplandığı; birinci faktörün öz değeri 8.16, ikinci faktörün öz değeri ise 2.17 olarak elde edilmiştir. Tek boyutlu yapıya ait GFI

değeri 0.91'dir. Tüm faktör yükleri 0.30'un üstündedir. Elde edilen sonuçlar doğrultusunda testin tek boyutlu olduğu kabul edilmiştir.

Yerel bağımsızlık

Yerel bağımsızlık tek boyutlulukla ilişkili bir özelliktir. Eğer bir test tek boyutluluk özelliğini gösteriyorsa, bu testte yer alan maddelerin yerel bağımsızlık özelliğine sahip olduğu söylenebilir (Hambleton ve Swaminathan, 1985). Uzman görüşlerine göre ve açımlayıcı faktör analizi sonuçlarına göre testin tek boyutluluk özelliğine sahip olması, yerel bağımsızlığa sahip olduğunu da göstermektedir. Ayrıca test içerisinde ortak köklü sorular olmadığı gibi aynı kazanımı ölçmeye yönelik sorular olmadığı da uzmanlarca belirtilmiştir.

Model-veri uyumu

Model-veri uyumu varsayımının incelenmesi amacıyla -2likelihood değeri, AIC ve BIC teste ait model veri uyumları ile birlikte alfa ve omega güvenilirlik katsayıları hesaplanmış bulgular Tablo 2'de gösterilmiştir.

Tablo 2. Araştırma kapsamında oluşturulan veri setlerinin -2likelihood değeri, AIC, BIC ve güvenilirlik katsayıları

		Veri seti-1	Veri seti- 2	Veri seti-3	Veri seti-4	Veri seti-5	Veri seti-6	Veri seti-7
-2 LL	1PL	75653.36	3532.78	32162.24	28741.42	36839.74	40030.10	25720.46
	2PL	75006.04	3198.78	34630.96	32968.10	39688.60	45312.72	27661.92
	3PL	75005.94	3398.56	34630.88	32968.54	39688.52	45312.72	27661.90
AIC	1PL	75751.36	3814.77	32260.24	28843.41	36937.74	40128.11	25818.47
	2PL	75198.03	3758.78	34822.96	33168.09	39880.60	45504.73	27853.92
	3PL	75199.95	3720.56	35824.88	33168.54	39882.52	45506.71	27855.89
BIC	1PL	76023.70	4182.10	32484.49	29098.52	37168.32	40384.23	26028.75
	2PL	75731.60	4488.22	35262.30	33668.31	40332.34	46006.52	28265.91
	3PL	75739.07	4139.99	35268.80	33668.75	40338.97	46013.72	28272.18
Alfa		0.75	0.92	0.52	0.30	0.50	0.28	0.50
Omega		0.75	0.92	0.53	0.31	0.51	0.29	0.51

Tablo 2 incelendiğinde model veri uyumları incelendiğinde, veri seti-1 ve veri seti-2'nin 1PL'ye daha iyi uyum gösterdiği; Diğer tüm veri setlerinin ise 3PL'ye daha iyi uyum gösterdiği belirlenmiştir. Veri seti-4 ve veri seti-6 için düşük güvenilirlik katsayısı değerleri elde edilmiştir. Güvenirlikleri düşük hesaplanan veri setlerinin kendi içinde normal dağılım özelliği gösteren sağa çarpık dağılımlar olduğu belirlenmiştir. Diğer bir deyişle başarısız grup olarak nitelendirilen gruplara ilişkin güvenilirlik değerleri düşük hesaplanmıştır. Bununla birlikte diğer veri setlerinin güvenilirliği orta ve yüksek düzeylerde hesaplanmıştır.

BULGULAR

Araştırma kapsamında öncelikle her bir veri seti için KTK ve MTK'ye dayalı madde parametreler hesaplanmış ve parametreler arasındaki ilişkiler Pearson Momentler Çarpımı Korelasyon Katsayısı ile incelenmiştir. Bulgulara Tablo 3'te yer verilmiştir.

Tablo 3. KTK ve MTK'ye dayalı madde parametreleri arasındaki ilişkiler

	Veri seti-1	Veri seti-2	Veri seti-3	Veri seti-4	Veri seti-5	Veri seti-6	Veri seti-7
Güçlük parametresi	KTK-MTK(1PL)	.963**	.918**	.984**	.449**	.983**	.927**
	KTK-MTK(2PL)	.678**	.729**	.924**	.727**	.894**	.802**
	KTK-MTK(3PL)	.683**	.851**	.919**	.837**	.409**	.870**
Ayırt edicilik parametresi	KTK(alt-üst)-MTK(2PL)	-.036	-.551**	.383**	-.385**	.171	-.321*
	KTK(alt-üst)-MTK(3PL)	-.025	.012	-.016	.070	.338*	.243
	KTK(nokta çift)-MTK(2PL)	.482**	-.375**	.769**	-.299*	.525	-.202
	KTK(nokta çift)-MTK(3PL)	.489**	.120	.193	.118	-.117	.301*

* p < .05 ** p < .01

KTK ve MTK'dan elde edilen güçlük parametreleri incelendiğinde, genel olarak yüksek düzeyde ilişkiler gözlenmiştir. Veri seti-3 ve veri seti-7'de bu ilişkiler en yüksek düzeyde elde edilmiştir. Bu veri setlerindeki örneklem büyüklüklerinin diğer veri setlerinden daha düşük düzeyde olması dikkate çekicidir. Tüm örneklemin dâhil edildiği analiz bulgularına göre KTK'dan elde edilen klasik güçlük parametresi ile 2PL ve 3PL modellerinden elde edilen güçlük parametresi arasında orta düzeyde bir ilişki olduğu belirlenmiştir. Diğer tüm ilişki katsayıları yüksek düzeyde ve anlamlıdır. Ayırt edicilik katsayıları arasındaki ilişkilerde ise bir örüntü belirlenmemiştir. En yüksek ilişki katsayısı veri seti-3 için KTK'dan elde edilen nokta çift serili korelasyon katsayısı ile 2PL modelden elde edilen a parametresi arasında belirlenmiştir. Genel olarak KTK'dan elde edilen nokta çift serili korelasyon katsayısı ile 2PL modelden elde edilen a parametresi arasındaki ilişkiler orta düzeyde ve anlamlıdır. Diğer ilişki katsayılarının ise düşük düzeyde olduğu ya da ilişkinin olmadığı durumların ortaya çıktığı belirlenmiştir. KTK ve MTK'dan elde edilen güçlük parametreleri arasında güçlü bir ilişki gözlenirken, ayırt edicilik katsayıları için aynı şeyleri söylemek mümkün değildir. Bu durum farklı varsayımları bulunan KTK ve MTK'ye dayalı hesaplanan ayırt edicilik parametresinin de farklılaştığına işaret etmektedir.

İkinci araştırma sorusuna cevap olarak tüm veri setinden ve oluşturulan diğer yedi veri setinden elde edilen madde güçlük ve madde ayırt edicilik parametreleri arasındaki ilişkiler incelenmiştir. Bu amaçla elde edilen ilişkilere ait ortalama değerlere yer verilmiştir. İlişkilere ait ortalama değerler Tablo 4'te yer almaktadır.

Tablo 4. Her bir veri setinden elde edilen madde parametreleri arasındaki ilişkilerin ortalaması

	Güçlük parametresi				Ayırt edicilik parametresi			
	KTK	1PL	2PL	3PL	KTK(alt-üst)	KTK(nokta çift)	2PL	3PL
VS-1 – VS-2	.971	.970	.461	.532	.669	.219	.695	.482
VS-1 – VS-3	.971	.978	.879	.879	.224	.269	.762	.455
VS-1 – VS-4	.979	.496	.681	.672	.814	.479	.631	.459
VS-1 – VS-5	.986	.990	.762	.104	.628	.654	.731	-
								.354
VS-1 – VS-6	.990	.984	.791	.729	.876	.634	.789	.426
VS-1 – VS-7	.972	.979	.773	.775	.453	.552	.492	.419

Tablo 4 incelendiğinde madde güçlük indeksleri arasındaki ilişkilere yönelik; KTK'ya dayalı olarak farklı veri setlerinde hesaplanan maddelerin güçlük indeksleri arasındaki ilişkilerin 0,887 ile 0,930 arasında değişiklik gösterdiği tespit edilmiştir. Diğer bir anlatımla oluşturulan veri setlerinden (başarılı-başarısız) hesaplanan madde güçlük indekslerinin büyük ölçüde tüm veri setinden elde edilen sonuçlarla benzerlik gösterdiği belirlenmiştir. MTK'ya dayalı 1PL model ile kestirilen madde güçlük indeksleri arasındaki ilişkilerin 0,931 ile 0,502 arasında olduğu tespit edilmiştir. Tüm veri seti ile üçüncü ve beşinci veri setlerinden elde edilen güçlük indekslerinin orta; diğer veri setlerinin ise yüksek düzeyde ilişkili olduğu belirlenmiştir. Öğrencilerin kümeleme ve bağlı değerlendirme sonucunda başarılı olarak gruplandırıldıkları veri setlerindeki ilişkilerin 1pl modelde, tüm veriden farklılıklar gösterdiği tespit edilmiştir. 2PL model ile kestirilen madde güçlük indeksleri arasındaki ilişkilerin 0,840 ile 0,544 arasında değişiklik gösterdiği belirlenmiştir. Kümeleme analizi ve bağlı değerlendirme sonucunda başarısız olarak gruplandırılan veri setleri ile tüm veri setinden hesaplanan madde güçlük indeksleri arasındaki ilişkilerin orta düzeyde olduğu saptanmıştır. 3PL ile hesaplanan güçlük indeksleri arasındaki ilişkilerin ise 0,857 ile -0,462 arasında farklılık gösterdiği belirlenmiştir. Kümeleme analizine göre başarısız olarak belirlenen gruptan elde edilen güçlük indeksleri ile tüm veri setinden hesaplanan güçlük indeksleri arasında negatif yönde bir ilişki olduğu tespit edilmiştir.

Tablo 4 madde ayırt edicilikleri bakımından incelendiğinde; KTK alt-üst gruba dayalı gerçekleştirilen hesaplamalarda anlamlı ilişkiler olmadığı belirlenmiştir. Nokta çift serili

korelasyon katsayısında ise tüm veri setinden elde edilen değerlerle diğer veri setlerinden elde edilen değerler arasında negatif yönde ilişkiler olduğu tespit edilmiştir. Diğer bir anlatımla bir veri setinde ayırıcılığı yüksek çıkan maddenin diğer bir veri setinde ayırt ediciliği düşük ya da orta hesaplanmıştır. Bu durum; veri setinin başarılı-başarısız olarak sınıflandırılmasından kaynaklı olabilir.

MTK'nin 2PL modeline dayalı hesaplanan madde ayırıcılık indeksleri arasındaki ilişkilerin 0,436 ile 0,743 arasında değişiklik gösterdiği belirlenmiştir. Tüm veri setinden elde edilen değerle en yüksek ilişkinin 3PL'ye dayalı hesaplanan madde ayırıcılık indeksi arasındaki ilişkilerin de -0,108 ile 0,628 arasında hesaplandığı saptanmıştır. Tüm veri setinden hesaplanan madde ayırıcılık indeksi ile en yüksek ilişkinin mutlak başarıya göre (%75) doğru cevap veren gruptan elde edilen değerlerle hesaplandığı tespit edilmiştir.

Madde güçlük ve madde ayırt edicilik parametrelerinin veri setlerine göre farklılaşp farklılaşmadığı tek yönlü ANOVA ile incelenmiştir. Anlamlı farklılıkların belirlenmesinde Bonferroni testi uygulanmış, sonuçlar Tablo 5'te gösterilmiştir.

Tablo 5. Madde güçlük ve ayırt edicilik parametrelerinin değişimine ilişkin hesaplanan ANOVA sonuçları

	Güçlük parametresi				Ayırt edicilik parametresi			
	KTK	1PL	2PL	3PL	KTK(alt-üst)	KTK(nokta çift)	2PL	3PL
VS-1 – VS-2	•	•	+	•	+	+	+	+
VS-1 – VS-3	+	•	•	•	+	+	+	+
VS-1 – VS-4	•	+	+	+	+	+	+	+
VS-1 – VS-5	•	•	•	-	+	+	+	+
VS-1 – VS-6	•	•	+	•	+	+	+	+
VS-1 – VS-7	+	•	•	•	+	+	+	+

• = anlamlı fark yok ($p > 0.05$), + = anlamlı fark vardır, soldaki veri setine ait ortalama sağdakinden daha yüksektir, - = anlamlı fark vardır, soldaki veri setine ait ortalama sağdakinden daha düşüktür

Tablo 5'te madde güçlük indeksinin KTK'ya dayalı gerçekleştirilen kestiriminde tam veri seti ile üçüncü (kümeleme analizi sonucunda başarılı) ve yedinci (mutlak başarı sonucunda başarılı) veri setlerinden hesaplanan madde güçlük indekslerinin anlamlı bir şekilde farklılaştığı; tüm veri setinden elde edilen madde güçlük indekslerinin daha

yüksek düzeyde olduğu tespit edilmiştir. MTK'ye dayalı gerçekleştirilen 1PL, 2PL ve 3PL modeline dayalı hesaplamalarda tüm veri seti ile dördüncü (ortalamanın üstünde başarılı grup) veri setinden hesaplanan madde güçlük indekslerinin anlamlı bir farklılık gösterdiği tespit edilmiştir. Bunun dışında 2PL modelde tüm veri seti ile ikinci (kümeleme analizi sonucunda başarısız) ve altıncı (mutlak başarıya göre başarısız) veri setlerinin de farklılaştığı tespit edilmiştir. 3PL modelde de beşinci veri setinde (ortalamanın altında başarısız) veri seti ile tüm veri setinin madde güçlük indekslerinin anlamlı bir farklılık gösterdiği belirlenmiştir.

Tablo 5'te yer alan madde ayırt edicilik indekslerinin değişimi incelendiğinde tüm veri setinden hem KTK hem de MTK'ya dayalı gerçekleştirilen hesaplamaların; diğer veri setlerinden hesaplanan değerlerle anlamlı bir farklılık gösterdiği saptanmıştır.

KTK ve MTK'dan elde edilen yetenek parametreleri arasındaki ilişkilerin incelenmesinde Pearson korelasyon katsayısı kullanılmıştır. Bu amaçla 2 parametrelili lojistik model aracılığıyla en çok olabilirlik yöntemine göre yetenek parametresi her bir veri seti için tahmin edilmiştir. KTK için ise toplam puanlar elde edilmiştir. Hesaplanan ilişki katsayıları Tablo 6'da gösterilmiştir.

Tablo 6. KTK ve MTK'ya dayalı hesaplanan yetenekler arasındaki ilişkiler

	VS-1	VS-2	VS-3	VS-4	VS-5	VS-6	VS-7
KTK - MTK	0.970	0.940	0.936	0.931	0.934	0.947	0.911

Yetenek parametreleri arasındaki ilişkiler incelendiğinde, tüm veri setine ait ilişkilerin en yüksek düzeyde olduğu, yüzde 75'lik ölçüte göre ayrılmış veri seti-7'de ise en düşük ilişkilerin ortaya çıktığı belirlenmiştir. Elde edilen tüm katsayıları çok yüksek bir ilişkiyi göstermekte ve tüm katsayılar .001 düzeyinde anlamlıdır.

TARTIŞMA ve SONUÇ

KTK ve MTK'den elde edilen madde ve kişi parametrelerinin karşılaştırıldığı birçok araştırma alan yazında yer almaktadır (DeMars, 2001; Fan, 1998; Hwang, 2002; Macdonald ve Paunonen, 2002). Bu araştırmaların ortak noktası, dağılım özellikleri ve farklı kesme noktalarına göre karşılaştırmalar yapmaması ve KTK ve MTK karşılaştırmalarından benzer bulgular elde etmeleridir. Ancak bu araştırma, farklı ölçütlere göre ayrılmış başarılı-başarısız grupların veri setleri ve kendi içinde normal dağılım gösteren farklı dağılım özelliklerinin parametreler üzerindeki etkisi karşılaştırılmıştır. Araştırma sonucunda farklı veri setlerinde yer alan bireylerin yetenek parametrelerinin KTK ve MTK'ya dayalı olarak benzer sonuçlar ürettiği belirlenmiştir. Başka bir anlatımla ölçme kuramlarının farklı dağılım özellikleri ve nitelikleri bulunan veri setlerindeki bireylerin yeteneklerini benzer şekilde kestirdiği tespit edilmiştir.

Xu ve Jia (2011), madde parametrelerinin yetenek dağılımından etkilenmediği, ancak özellikle çok çarpık veri setlerinden elde edilen madde parametrelerinin nispeten etkilendiğini belirtmişlerdir. Ayrıca, madde parametrelerinin normallikten uzaklaşan dağılımlardan etkilendiğini belirten çeşitli araştırma bulguları da alan yazında yer almaktadır (Roberts, Donoghue ve Laughlin, 2002; Sass, Schmitt ve Walker, 2008; Sen, Cohen ve Kim, 2016). Tüm bu araştırmalar, normal olmayan dağılımın etkisinin sınırlı olduğunu ve örneklem büyüklüğü ve test uzunluğu ile ilgili alınabilecek önlemler ile birlikte sorunun çözülebileceğini belirtmişlerdir. Bu araştırmanın bulguları, ilgili literatürdeki bulgularla benzerlik göstermektedir.

KTK ve MTK'dan elde edilen güçlük parametrelerinin farklı veri setinde genel olarak benzer bulgular ürettiği belirlenmiştir. Ancak aynı durum ayırt edicilik katsayıları için geçerli değildir. 2 parametrelili lojistik modelden elde edilen a parametresinin farklı KTK ayırt edicilik parametreleri ile genellikle yüksek bir ilişki gösterdiği; ancak 3 parametrelili lojistik modelin KTK'dan elde edilen farklı ayırt edicilik parametreleri ile düşük ilişki içinde olduğu belirlenmiştir. 2 parametrelili lojistik model ile KTK'dan elde

edilen ayırt edicilik parametrelerinin farklı dağılım özellikleri ve kesme noktalarında da benzer olduğu söylenebilmektedir.

Tüm örneklemden elde edilen KTK ve MTK'ya dayalı ayırt edicilik parametrelerinin diğer tüm veri setlerinden daha yüksek düzeyde elde edildiği belirlenmiştir. Diğer altı veri setinden elde edilen ayırt edicilik katsayıları ise genellikle kendi içinde benzer değerler üretmiştir.

Farklı veri setlerinden elde edilen güçlük parametreleri arasında genellikle yüksek ilişkiler belirlenmiştir. Ayırt edicilik katsayılarında ise oldukça geniş ranjda bulgular elde edilmiştir. Farklı veri setlerinde 2 parametrelili lojistik modelden elde edilen parametrelerinin genellikle orta ve yüksek düzeyde ilişki gösterdiği söylenebilir. Ayrıca, tüm veri setlerinde KTK ve MTK'den elde edilen yetenek parametreleri arasındaki ilişkiler yüksek ve istatistiksel olarak anlamlıdır.

Araştırma sonucunda farklı ölçütlere göre belirlenmiş başarılı-başarısız gruba ilişkin veri setlerinde yer alan maddelerin güçlük indekslerinde hem KTK hem de MTK'ya dayalı gerçekleştirilen hesaplamalarda benzerlikler olduğu gibi farklılıklar da hesaplanmıştır. Güçlük indeksleri arasındaki ilişkiler KTK'ya dayalı gerçekleştirilen kestirimlerde MTK'ya dayalı kestirime göre daha yüksek hesaplanmıştır. Bu durum, KTK'nin hesaplamalarının gruba bağlı olduğunun bir göstergesi olmakla birlikte MTK kestirimlerinin de gruba bağlı değişim gösterdiğini göstermektedir. Benzer şekilde madde ayırt edicilik indekslerinin hem KTK hem de MTK'ya dayalı modellerle hesaplamaları arasında anlamlı farklılıklar olduğu tespit edilmiştir. Bu durum da hem KTK hem de MTK'nin parametre kestiriminde gruptan etkilendiğini göstermektedir.

Bu araştırma sonuçlarına dayalı olarak KTK ve/veya MTK'ya dayalı gerçekleştirilecek parametre kestirimlerinde veri setinin kendi içinde normal dağılım göstermesinin yanı sıra heterojen olmasının da dikkate alınması önerilmektedir. Belli bir yetenek düzeyinin altında ya da üstündeki grupla gerçekleştirilen kestirimlerin heterojen bir gruptan gerçekleştirilen kestirimlerle farklı sonuçlar verdiği göz önünde bulundurulmalıdır. Veri setinin normal dağıldığı durumlarda KTK ve MTK modellerinden herhangi biri ile

madde parametrelerinin kestirilmesi uygun iken özellikle heterojen veri setlerinde 2 parametrelili lojistik modele uygun MTK'ya ait madde parametresi kestirimleri yapmak önerilmektedir.

KAYNAKLAR

- Adedoyin, O. O., Nenty, H. J., & Chilisa, B. (2008). Investigating the invariance of item difficulty parameter estimates based on CTT and IRT. *Educational Research and Reviews*, 3(3), 83.
- Awopeju, O. A., & Afolabi, E. R. I. (2016). Comparative analysis of classical test theory and item response theory based item parameter estimates of senior school certificate mathematics examination. *European Scientific Journal, ESJ*, 12(28).
- Çıkrıkçı-Demirtaşlı, N. (2002). A study of raven standard progressive matrices test's item measures under classic and item response models: an empirical comparison1. *Ankara University, Journal of Faculty of Educational Sciences*, 35(1-2).
- DeMars, C. (2001). Group differences based on IRT scores: Does the model matter?. *Educational and Psychological Measurement*, 61(1), 60-70.
- 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.
- Gelbal, S. (1994). pMadde güçlük indeksi ile rasch modelinin b parametresi ve bunlara dayalı yetenek ölçüleri üzerine bir karşılaştırma. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 10(10).
- Hambleton, R. K., & Jones, R. W. (2012). Comparison of classical test theory and item response theory and their applications to test development, Instructional Topics in Educational Measurement Series 16.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.
- Harris, D. (1989). Comparison of 1-, 2-, and 3-Parameter IRT Models. *Educational Measurement: Issues and Practice*, 8(1), 35-41.
- Hernandez, R. (2009). Comparison of the item discrimination and item difficulty of the quick-mental aptitude test using CTT and IRT methods. *The International Journal of Educational and Psychological Assessment*, 1(1), 12-18.
- Hwang, D. Y. (2002). Classical Test Theory and Item Response Theory: Analytical and Empirical Comparisons. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Austin, TX.
- Lawson, S. (1991). One parameter latent trait measurement: Do the results justify the effort. *Advances in educational research: Substantive findings, methodological developments*, 1, 159-168.

- 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.
- Progar, Š., Sočan, G., & Peč, M. (2008). An empirical comparison of item response theory and classical test theory. *Horizons of Psychology*, 17(3), 5-24.
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2002). Characteristics of MML/EAP parameter estimates in the generalized graded unfolding model. *Applied Psychological Measurement*, 26(2), 192-207.
- Sass, D. A., Schmitt, T. A., & Walker, C. M. (2008). Estimating non-normal latent trait distributions within item response theory using true and estimated item parameters. *Applied Measurement in Education*, 21(1), 65-88.
- Sen, S., Cohen, A. S., & Kim, S. H. (2016). The impact of non-normality on extraction of spurious latent classes in mixture IRT models. *Applied Psychological Measurement*, 40(2), 98-113.
- Skaggs, G., & Lissitz, R. W. (1988). Effect of examinee ability on test equating invariance. *Applied Psychological Measurement*, 12(1), 69-82.
- Stage, C. (1999). *A Comparison Between Item Analysis Based on Item Response Theory and on Classical Test Theory: A Study of the SweSAT Subtest WORD*. Department of educational measurement, Umeå univ..
- Xu, X., & Jia, Y. (2011). The sensitivity of parameter estimates to the latent ability distribution. *ETS Research Report Series*, 2011(2).

SUMMARY

The aim of this study is to determine the extent to which item parameters estimated via CTT and IRT depend on group properties (being master or non-master). Besides, the relationship between findings obtained from CTT and IRT is evaluated. The study group of this study consists of 1916 undergraduates who study at different departments (22) in educational faculty in a state university of Turkey. Within the scope of the study, the final exam of Turkish Language 1: Written Expression course that is one of the common courses in educational faculty during 2011-2012 educational year has been used as data collection tool. Data sets were created from the test containing 48 items in total according to researcher's purpose. After data sets were created, for every condition, three different item discrimination parameters (below/above discrimination, point biserial and biserial correlation), and two different item difficulty parameters (below/above difficulty and whole group difficulty) were estimated via CTT. In IRT, item difficulty parameters (b parameter) and item discrimination parameter (a parameter) were estimated for 2PL, 3PL and nominal response model. The relationship between item parameters obtained from data sets having different properties and the relationships between item parameters obtained via CTT and IRT have been investigated separately. Besides, the relationships between total scores for CTT and ability parameters for IRT have been examined for different data sets. The assumptions of normality, unidimensional, local independence and invariance for IRT have been checked. The relationship between point biserial correlation obtained from CTT and a parameter obtained from 2PL is moderate and significant. It has been determined that the other relationship coefficients are low or non-significant. It has been identified that item difficulty parameters are almost similar for the whole data set. It has been ascertained that there is a significant difference between item difficulty parameters for 1LP, 2PL and 3PL models of IRT for the whole data set and the fourth data set (master group above average). Besides, it has also been determined that the whole data set for 2PL model differs from second (non-master according to cluster analysis) and sixth (non-master according to absolute mastery) data sets. When the relationships between ability parameters have been examined, the relationships for the whole data set are at highest level and the relationships for the data set 7 that is separated according to 75% criterion are at lowest level. The coefficients obtained show high relationship and significant at the level of .001. According to literature, the effect of non-normal distribution is limited, and the problem can be solved with the precautions related to sample size and test length. The findings of this study resemble the findings in the literature. It has been determined that measurement theories estimate the ability of individuals in data sets having different distribution and properties similar. The high relationships have generally been obtained among difficulty parameters estimated from different data sets. The discrimination coefficients range widely. A parameter obtained from 2PL model in different data sets have generally moderate or high relationships. Besides, the relationships between ability parameters obtained from CTT and IRT in the whole data sets are high and significant. There are both similarities and differences between difficulty indexes of the items in data sets composed to different criteria (master-non-master) estimated for CTT and IRT in data sets. The relationships between difficulty indexes for CTT are higher than the estimates for IRT. This situation shows that the estimates obtained for CTT are group-dependent and the estimates for IRT are group-independent. Similarly, it has been observed that there is a significant

difference between item discrimination indexes estimated for CTT and IRT. This shows that both IRT and CTT parameters are affected by groups. According to the results of this study, it should be noted that the data set should have normal distribution and be heterogonous for the parameter estimation based on CTT and IRT.