Türk Psikoloji Dergisi, Aralık 2009, *24* (64), 61-75

Ölçme Eşdeğerliğinin Yapısal Eşitlik Modellemesi ve Madde Cevap Kuramı Kapsamında İncelenmesi

Oya Somer, Ege Üniversitesi
Mediha Korkmaz, Ege Üniversitesi
Seda Dural, Ege Üniversitesi
Seda Can, İzmir Ekonomi Üniversitesi

Özet

Bu çalışmanın amacı kadın ve erkek karşılaştırma grupları için bir kişilik alt ölçeğinden elde edilen veriler kullanılarak, ölçme eşdeğerliğinin MACS ve DIF yöntemleri kapsamında incelenmesi ve bu yöntemlerden elde edilen parametrelere ilişkin sonuçların tartışılmasıdır. Çalışmada katılımcı olarak Somer ve arkadaşları (2004) tarafından geliştirilen Beş Faktör Kişilik Envanteri'nin (5FKE) yetişkin norm örnekleminin Endişeye Yatkınlık alt ölçeğinden alınan 500 kadın ve 500 erkek yer almıştır. MACS analizlerinde LISREL 8.8; DIF analizlerinde PARSCALE 4.1 programları kullanılmıştır. MACS analizleri sonucunda faktör yüklerine karşılık gelen parametrelerin tüm maddeler için kadın ve erkek gruplarında eşdeğer olduğu bulunurken, DIF analizleri sonucunda ölçekte yer alan 10'uncu maddenin ayırtetme parametresinin farklılaştığı görülmüştür. Ayrıca, MACS analizlerinde regresyon sabitinin ve DIF analizlerinde madde yerleşim parametrelerinin değişmezliğine ilişkin bulgular her iki yaklaşımda da 5., 6., 9., ve 11'inci maddelerin gruplar arasında farklılaştığını göstermiştir. Çalışmada her iki yöntemle elde edilen bulgular benzer sonuçlar vermekle birlikte, söz konusu yöntemlerin avantaj ve dezavantaj durumları göz önünde bulundurulduğunda ölçme eşdeğerliği çalışmalarında her iki yöntemden de faydalanılması önerilmektedir.

Anahtar kelimeler: Ölçme eşdeğerliği, ortalama ve kovaryans yapısı modeli, madde işlevsel farklılığı

Abstract

The purpose of the present study was to investigate measurement equivalence for male and female comparison groups in the DIF and MACS methods by using the data obtained from a personality subscale and to discuss the estimated parameter results obtained from these two methods. The participants were 500 females and 500 males from the adult norm sample of Big Five Personality Inventory Proneness to Anxiety subscale developed by Somer and her colleagues (2004). MACS analyses were conducted by using LISREL 8.8; DIF analyses were conducted by using PARSCALE 4.1 computer programs. MACS results showed that the factor loadings were found to be invariant for male and female groups however it was seen in the DIF results that the 10th item was functioning differentially between comparison groups. Moreover, the results concerning the invariance of the intercepts estimated in the MACS and the item location parameters estimated in the DIF analyses showed that the 5th, 6th, 9th and 11th items were differentiated significantly between males and females in both methods. Although the results obtained from both methods revealed similar results, when the advantages and disadvantages of these methods were taken into consideration, the application of both methods together in measurement equivalence studies is suggested.

Key words: Measurement equivalence, mean and covariance structure model, differential item functioning

Yazışma Adresi: Yrd. Doç. Dr. Mediha Korkmaz, Ege Üniversitesi Edebiyat Fakültesi Psikoloji Bölümü, Bornova - İzmir.
E-posta: mediha.korkmaz@ege.edu.tr

Gruplar arası karşılaştırmalar psikoloji alanında yapılan çalışmalarda önemli bir yer tutmaktadır. Söz konusu karşılaştırmalar genellikle bilişsel yetenekler, kişilik özellikleri, düşünme stilleri gibi örtük özellikler (*latent trait*) üzerinden yapılmaktadır. Bu tür karşılaştırmaların geçerli olabilmesi için ilgili yapılar bakımından gruplar arasında ölçme eşdeğerliğinin (*measument equivalence*) sağlanmış olması gerekmektedir. Mellenbergh (1989), Meredith (1993) ve Meredith ve Millsap (1992) tarafından ölçme eşdeğerliği şu şekilde tanımlanmaktadır: ölçme eşdeğerliği, herhangi bir bireyin belirli bir gözlenen puana sahip olma olasılığının hangi grupta yer aldığından bağımsız olma durumudur. Bu tanımdan hareketle ölçme eşdeğerliğinin sağlandığı koşulda, farklı gruplarda yer alan ama aynı gerçek puana sahip olan bireyler aynı gözlenen puana sahip olacaktır. Bu koşulun sağlanmadığı durumda yapılacak olan grup karşılaştırmalarından elde edilecek farklılıkların ölçmedeki bir yanlılıktan mı, yoksa gerçek grup farklılıklarından mı kaynaklandığını yorumlamak problematik olabilmektedir (Chan, 2000; Somer, 2004; Stark ve ark., 2006).

Ölçme eşdeğerliğinin incelenmesinde literatürde sıklıkla iki yaklaşımın kullanıldığı görülmektedir. Bu yaklaşımlardan birisi Madde Cevap Kuramı'na (*Item Response Theory* - IRT) dayalı Madde ve Test İşlev Farklılıklarını inceleyen (*Differential Item and Test Functioning* - DIF ve DTF) modeller, diğeri ise Yapısal Eşitlik Modellemeleri'dir (*Structural Equation Modeling* - SEM). SEM kapsamında ölçme eşdeğerliği çalışmalarında iki tür yaklaşım kullanılabilmektedir. Bunlardan en yaygın olarak kullanılanı kovaryans yapılarının eşdeğerliğinin test edildiği Çoklu Grup Doğrulayıcı Faktör Analizleri'dir (*Multi Group Confirmatory Factor Analysis* - MGCFA). İkincisi ise kovaryans yapılarıyla birlikte ortalama yapılarının da karşılaştırıldığı Ortalama ve Kovaryans Yapılarının (*Mean and Covariance Structure* - MACS) eşdeğerliğini inceleyen yaklaşımlardır.

IRT literatüründe işlevsel farklılık; metrik eşitlemesi yapıldıktan sonra farklı alt grup üyelikleri olan deneklerin, "aynı" yetenek ya da psikolojik özellik düzeyinde maddeyi doğru yanıtlama/onaylama olasılıklarının farklılık göstermesidir (Camilli ve Shepard, 1994; Hambleton ve ark., 1991; Raju ve ark., 2002; Thissen ve ark., 1988). Eğer bu işlevsel farklılık, madde düzeyinde gerçekleşiyorsa DIF; toplam test puanı düzeyinde gerçekleşiyorsa DTF olarak adlandırılmaktadır (Collins ve ark., 2000; Flowers ve ark., 1999; Maurer ve ark., 1998; Raju ve ark., 1995). SEM literatüründe ise ölçmedeki değişmezlik kovaryans yapılarının ve/veya ortalama yapılarının karşılaştırma grupları arasında eşdeğer olmasını ifade etmektedir ve sırasıyla MGCFA ve MACS model olarak anılmaktadır (Lubke ve ark., 2003; Raju ve ark., 2002).

Genel anlamda ölçme eşdeğerliği çalışmaları farklı popülasyonların (kültürlerarası), aynı evrenin alt örneklem gruplarının (cinsiyet, yaş, sosyo-ekonomik düzey gibi) karşılaştırılması veya aynı popülasyondaki zamana bağlı olarak ölçmenin durağanlığının (önsontest, tekrarlı ölçümler) incelenmesi gibi değişik koşullarda gerçekleştirilebilmektedir.

Bu araştırmada, kadın ve erkek karşılaştırma grupları için bir kişilik alt ölçeğinden elde edilen veriler kullanılarak, ölçme eşdeğerliğinin MACS ve DIF yöntemleri kapsamında incelenmesi ve bu yöntemlerden elde edilen bulguların karşılaştırılması amaçlanmıştır. Bu kısımda sırasıyla, MACS ve DIF yöntemleri kısaca ele alınacaktır.

SEM Kapsamında Ölçme Eşdeğerliği: MACS

Literatüre bakıldığında grup karşılaştırmalarını içeren SEM çalışmalarının yaklaşık olarak % 80'inde (Vandenberg ve Lance, 2000) ölçme eşdeğerliğinin kovaryans yapıları temelinde MGCFA kapsamında incelendiği görülmektedir. Ancak son yıllarda ölçme eşdeğerliği çalışılırken regresyon sabitinin (*intercept*) de karşılaştırma grupları bakımından eşdeğer olup olmadığının test edilmesi gerektiği vurgulanmaktadır. Bu nedenle, kovaryans yapılarının yanı sıra ortalama yapılarının da analizini içeren MACS model kullanılarak gözlenen puan ortalamaları da analize dahil edilmektedir.

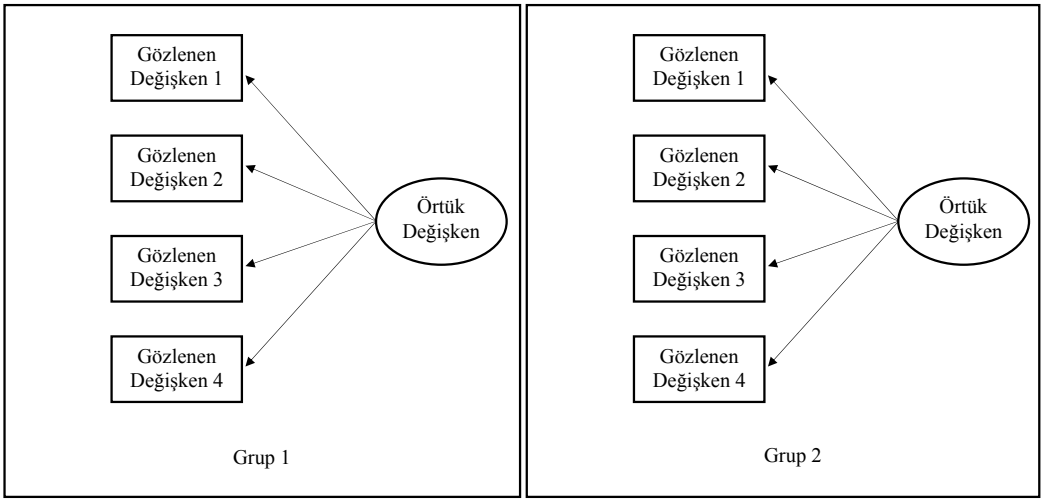
Genel olarak MACS kapsamında ölçme eşdeğerliğinin incelenmesi içiçe geçmiş (nested) 4 hiyerarşik modelin test edilmesini içermektedir (Byrne ve ark., 1989; Byrne ve Stewart, 2006; Chan, 2000; Little, 1997; Stark ve ark., 2006; Vandenberg ve Lance, 2000; Wu ve ark., 2007):

1. Yapısal değişmezlik modeli (*configural invariance model*)
2. Zayıf değişmezlik modeli (*weak invarinace model*)
3. Güçlü değişmezlik modeli (*strong invariance model*)
4. Katı değişmezlik modeli (*strict invariance model*)

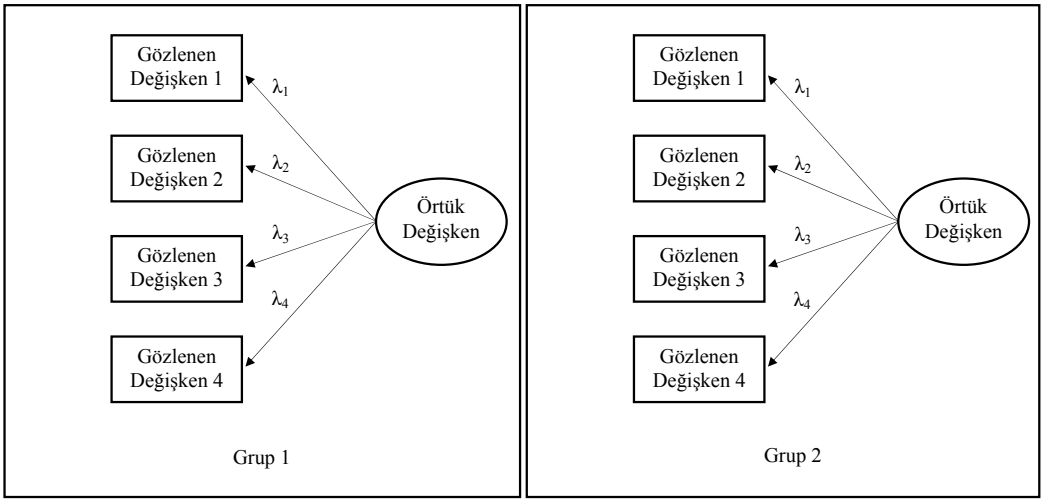
Yapısal Değişmezlik Modeli. Yapısal değişmezliğin incelendiği ilk aşamada, grupların aynı faktör yapısına sahip olup olmadığı incelenir. Bu nedenle,

[1] Parametrelerin her iki grupta ayrı ayrı tahminlenmesine izin verildiği durumlarda parametreler "serbest" olarak anılmakta, ikinci gruptaki parametre değerlerinin birinci gruba eşitlenmesi durumunda parametrelerin "sabitlenmesi" olarak anılmaktadır.

ölçme modeli için yapısal değişmezlik test edilirken faktör yükleri, regresyon sabitleri ve hata varyanslarının serbest tahminlenmesine izin verilerek, yalnızca gruplar için faktör sayısı ve yüklenme örüntüsü (loading pattern) sınırlandırılmaktadır[1] (Vandenberg ve Lance, 2000; Wu ve ark., 2007). Örneğin, iki farklı grup için belirli bir örtük değişkene ilişkin dört gözlenen değişkenin olduğu hipotetik bir model düşünelim. Bu durumda yapısal değişmezlik modeli test edilirken, faktör sayısının ve örüntüsünün her iki grup için de aynı şekilde tanımlandığı bir model oluşturulur. Bu örnek durum için yapısal değişmezlik modeli Şekil 1'de ve söz konusu modele ilişkin LISREL sentaksı Ek-1a'da gösterilmiştir. Ek-1a'da görülebileceği gibi yapısal



Şekil 1. Yapısal Değişmezlik Modeli



Şekil 2. Zayıf Değişmezlik Modeli

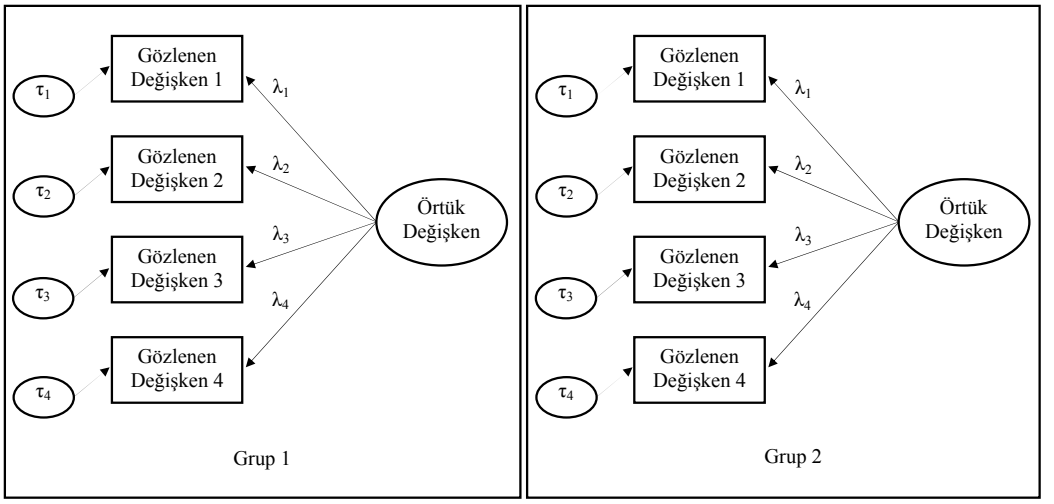
değişmezlik test edilirken, regresyon sabitleri (sentaksta CONST), faktör yükleri (sentaksta od-örtük değişken) ve hata varyansları (sentaksta Set the error variance of GD1-2-3-4 free) her iki grup için de eşitliklere yazılmak suretiyle serbest bırakılmaktadır.

Yapısal değişmezliğin sağlanması, farklı gruplarda aynı yapının ölçüldüğüne işaret etmektedir. Gruplar için aynı faktör sayısı ve yüklenme örüntüsüne sahip modellerin elde edilmiş olması, ölçme eşdeğerliğinin sonraki aşamalarının test edilmesine olanak sağlamaktadır. Eğer yapısal değişmezlik koşulu sağlanamaz ise, bu durum gruplarda farklı yapıların ölçüldüğünü gösterdiği için ölçme eşdeğerliğinin sonraki aşamalarında model karşılaştırmalarının yapılmasının bir anlamı olmayacaktır.

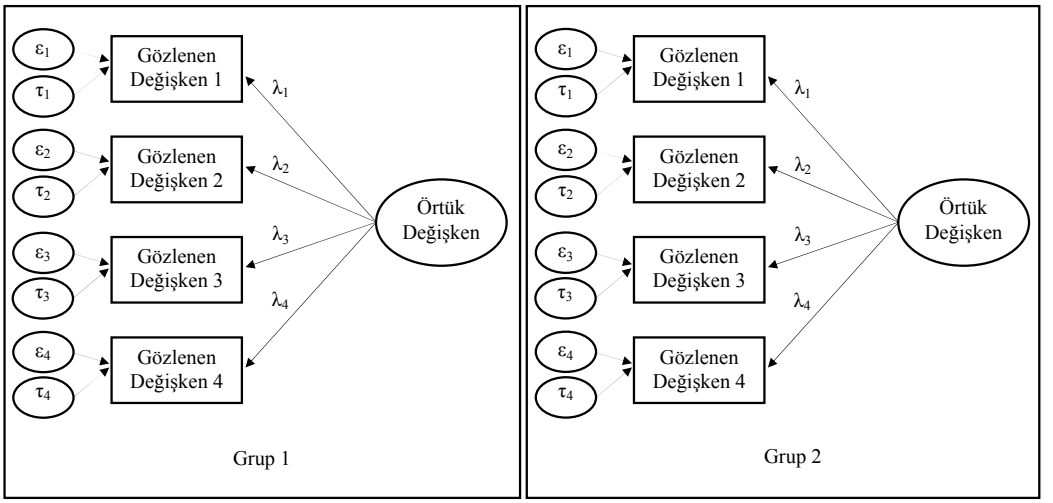
Zayıf Değişmezlik Modeli. Zayıf değişmezlik modelinde, grupların örtük değişkene ilişkin ölçme biriminin benzer olup olmadığı test edilir. Bu nedenle zayıf değişmezlik aynı zamanda metrik değişmezlik (*metric invariance*) olarak da adlandırılır. Zayıf değişmezlik test edilirken faktör sayısı ve yüklenme örüntüsü ile birlikte faktör yükleri de (λ_i) sınırlandırılmaktadır (Vandenberg ve Lance, 2000; Wu ve ark., 2007). Hipotetik örnek için zayıf değişmezlik modeli Şekil 2'de ve söz konusu modele ilişkin LISREL sentaksı Ek-1b'de gösterilmiştir. Ek-1b'den görülebileceği gibi zayıf değişmezlik test edilirken ikinci grupta yer alan eşitliklerden faktör yükleri (od) silinerek birinci grubun parametre değerlerine sabitlenmektedir. Eğer zayıf değişmezlik sağlanamaz ise, bu durum grupların farklı ölçme birimlerine sahip olduğuna işaret eder.

Güçlü Değişmezlik Modeli. Güçlü değişmezlik modelinde, grupların faktör puanı sıfır olduğunda elde edilen regresyon sabitinin eşit olup olmadığı test edilir. Bu nedenle güçlü değişmezlik aynı zamanda skalar değişmezlik (*scalar invariance*) olarak da adlandırılır. Güçlü değişmezlik test edilirken faktör örüntüsü ve faktör yüklerine ek olarak regresyon sabiti de (τ_i) sınırlandırılmaktadır (Vandenberg ve Lance, 2000; Wu ve ark., 2007). Hipotetik örnek için güçlü değişmezlik modeli Şekil 3'te ve söz konusu modele ilişkin LISREL sentaksı Ek-1c'de gösterilmiştir. Ek-1c'den görülebileceği gibi güçlü değişmezlik test edilirken ikinci grupta yer alan eşitlikler kaldırılarak faktör yüküne ek olarak regresyon sabitleri de birinci grubun parametre değerlerine eşitlenmektedir.

Katı Değişmezlik Modeli. Son aşamada yani katı değişmezlik modelinde ise, hata varyanslarının gruplarda farklılaşıp farklılaşmadığı test edilir. Ölçme modelindeki katı değişmezlik test edilirken bütün parametre sınırlamaları ile birlikte hata varyansları (ε_i) sınırlandırılır (Vandenberg ve Lance, 2000; Wu ve ark., 2007). Hipotetik örnek için güçlü değişmezlik modeli Şekil 4'te ve söz konusu modele ilişkin LISREL sentaksı Ek-1d'de gösterilmiştir. Ek-1d'de görülebileceği gibi katı değişmezlik test edilirken ikinci grupta yer alan ve hata varyanslarını serbest bırakan komutlar da kaldırılarak hata varyansları birinci grubun parametre değerlerine sabitlenmektedir.



Şekil 3. Güçlü Değişmezlik Modeli



Şekil 4. Katı Değişmezlik Modeli

Geleneksel olarak ölçme eşdeğerliğinin sağlanıp sağlanamadığı test edilirken içiçe geçmiş iki modelden elde edilen Ki-kare "χ^2" değerinden ve Karşılaştırmalı Uyum İndeksinden (*Comparative Fit Index* - CFI) yararlanılmaktadır (Byrne ve Stewart, 2006; Vandenberg ve Lance, 2000; Wu, ve ark., 2007). İki model için söz konusu değerlerin farkları alınarak $\Delta\chi^2$ ve ΔCFI hesaplanır. Elde edilen $\Delta\chi^2$'nin istatistiksel anlamlılığı test edilirken, iki modelin fark serbestlik derecesindeki kritik Ki-kare değeriyle karşılaştırılır. Bu karşılaştırmanın sonucunda istatistiksel olarak anlamsız bir $\Delta\chi^2$ değerinin elde edilmesi, ölçme eşdeğerliğinin sağlandığını göstermektedir. ΔCFI için herhangi bir istatistiksel anlamlılık testi yapılamamakla birlikte, iki modelin karşılaştırılması sonucunda elde edilen ΔCFI değerinin -.01 değerine eşit ya da bu değerden küçük olması, ölçme eşdeğerliğinin sağlandığına ilişkin bir kanıt olarak kullanılabilmektedir (Byrne ve ark., 1989; Wu, ve ark., 2007).

IRT Kapsamında Ölçme Eşdeğerliği: DIF

IRT'de, gözlenen değişkenler ile örtük özellik arasındaki işlevsel ilişki olasılığa dayalı matematiksel fonksiyonlar kullanılarak tanımlanmakta ve bu ilişki Madde Karakteristik Eğrisi (*Item Characteristic Curve* - ICC) ile temsil edilmektedir (Camilli ve Shepard, 1994; Chernyshenko ve ark., 2001; Hambleton ve Swaminathan, 1989; Hambleton ve ark., 1991; Somer, 1998, 1999; Zickar, 1998). IRT'de 1, 2 ya da 3 parametreli modeller kullanılarak parametre tahminleri yapılabilmektedir. Bu parametreler; (1) a_i - eğim / madde ayırt edicilik parametresi (*slope / item discrimination parameter*), (2) b_i - yerleşim / madde güçlük parametresi (*location / threshold*) ve (3) c_i - maddenin doğru yanıtını tahmin parametresidir (*guessing*). Psikolojik yapıların incelendiği çoğu araştırmada sıklıkla iki parametreli model tercih edilmektedir. Madde ayırt edicilik parametresi, madde karakteristik eğrisinin eğimini yani dikliğini belirlemektedir ve maddenin ölçülen yapı ile olan ilişkisinin düzeyini göstermektedir. Madde güçlük parametresinin değeri, *i* maddesini .50 oranında doğru cevaplayan deneklerin bulundukları örtük özellik (*theta* - θ) düzeyine karşılık gelmektedir.

IRT'de karşılaştırma grupları referans ve fokal gruplar olarak adlandırılır. Karşılaştırma gruplarının ICC'leri arasındaki fark, örtük özellik üzerinde belirli bir konumda bulunan referans grup ve fokal grup deneklerinin maddeye doğru/olumlu yanıt verme olasılıklarının eşdeğer olup olmadığını gösterir. Madde işlev farklılığı iki grupta doğru cevabın koşullu (*conditional*) olasılığının, $P_i(\theta)$, farklılık gösterdiği her durumda ortaya çıkmaktadır (Camilli ve Shepard, 1994). Başka bir ifadeyle aynı yetenek ya da tutum (θ) düzeyindeki deneklerin maddeyi aynı yönde yanıtlama olasılıklarının farklı olmasıdır. Bu farklılıklar karşılaştırma gruplarının madde karakteristik eğrilerinde düzgün (*uniform*) ve düzgün olmayan (*nonuniform*) olarak iki farklı formda ortaya çıkmaktadır. Düzgün formlu madde işlev farklılığında, sadece madde güçlük parametresi farklılaşmaktadır.

Diğer bir deyişle, maddenin yapı ile olan ilişkisi gruplar arasında farklılaşma göstermezken, maddeye doğru yanıtlama olasılığı grup üyeliğinden etkilenmektedir. Bu durum maddenin örtük özellik ile ilişkisinin her iki grupta da aynı yönde olduğu ancak maddenin gruplardan biri için daha zor iken diğeri için daha kolay olduğu anlamına gelir ve gruplardan biri için göreceli bir yanlılığa yol açar (Reise ve ark., 2001; Smith, 2002). Düzgün olmayan formlu madde işlev farklılığında ise, referans ve fokal grupların madde karakteristik eğrilerinin biçimleri farklılık göstermekte ve θ ölçeğinin bazı noktalarında kesişmektedir. Burada madde ile ölçülen özellik arasındaki ilişki bir grupta diğer gruba nazaran daha güçlüdür ya da ilişki grupların θ düzeylerine göre farklılaşmaktadır, çünkü maddenin hem ayırt edicilik hem de güçlük parametreleri gruplar arasında farklılık göstermektedir (Orlando ve Marshall, 2002; Smith, 2002; Van de Vijyer ve Leung, 1997).

Bu çalışmada IRT kapsamında ölçme eşdeğerliğinin incelenmesi amacıyla madde parametrelerini karşılaştırma yöntemi kullanılmıştır. Yöntem, karşılaştırma gruplarındaki (kadın-erkek) madde ayırtetme ile madde güçlük parametrelerinin farklılıklarına temellenmektedir. Bu yöntemde inceleme altına alınan bir maddenin, karşılaştırma grupları için ayrı ayrı tahminlenen madde ayırtetme ve madde güçlük parametre değerlerinin (metrik eşitliği sağlandıktan sonra) birbirlerinden çıkarılmasıyla madde işlevsel farklılığının bir ölçümü elde edilir. Daha sonraki aşamada ise elde edilen bu fark değerleri her iki karşılaştırma grubunun standart sapmalarına bölünmesiyle standardize edilmiş madde işlevsel farklılık istatistikleri hesaplanır (Korkmaz, 2006; Morales ve ark., 2000; Reise ve ark., 2001). Örneğin; referans grubun kadın ve fokal grubun erkek olduğu gruplar arası bir karşılaştırmada, bir maddenin standardize madde işlevsel farklılık değeri (SDIF) pozitif (+) olduğunda, bu maddenin fokal (erkek) grup üyeleri için kolay olduğunu ve standardize madde işlevsel farklılık (SDIF) değeri negatif (-) olduğunda da referans (kadın) grup üyeleri için daha kolay olduğunu göstermektedir (Smith, 2002). Thissen ve arkadaşları (1993), standardize madde işlevsel farklılık değerinin karesi alındığında "1" serbestlik derecesinde χ^2 istatistiği olarak değerlendirilebileceğini önermektedirler (akt., Reise ve ark. 2001). Bu ölçüte göre, eğer bir madde .01 ya da .05 nominal alfa düzeylerinde anlamlı bir χ^2 değerine sahipse, maddenin gruplar arasında işlevsel farklılık gösterdiğine karar verilir.

Her iki yaklaşımın madde parametreleri birlikte değerlendirildiğinde, MACS modelde τ, regresyon doğrusunun sabitidir ve örtük özellik "0" olduğunda gözlenen değişkenin aldığı değer olarak tanımlanır. IRT'de ise madde güçlük parametresi (b_i) olarak yorumlanmaktadır. MACS modelde λ değeri ise faktör yüküne karşılık gelmektedir ve IRT'de madde ayırtetme parametresi (a_i) olarak yorumlanmaktadır (Mellenbergh, 1994).

Yukarıda ifade edilen açıklamalar çerçevesinde bu çalışmanın amacı, kadın ve erkek karşılaştırma grupları için bir kişilik alt ölçeğinden elde edilen veriler kullanılarak, ölçme eşdeğerliğinin MACS ve, DIF yöntemleri kapsamında incelenmesi ve bu yöntemlerden elde edilen parametrelere ilişkin sonuçların tartışılmasıdır.

Yöntem

Örneklem

Bu çalışmada yer alan katılımcılar Somer ve arkadaşları (2004) tarafından geliştirilen Beş Faktör Kişilik Envanteri'nin (5FKE) yetişkin norm örnekleminden alınan 500 kadın ve 500 erkek olmak üzere toplam 1000 kişiden oluşmaktadır. Kadın katılımcıların yaş ortalaması 27.23 (S = 11.20) olup erkek katılımcıların yaş ortalaması ise 28.85'dir (S = 11.70).

Veri Toplama Araçları

Araştırmada kullanılan ölçme aracı Beş Faktör Kişilik Envanteri "Duygusal Tutarsızlık" faktörünün "Endişeye Yatkınlık" alt ölçeğidir (Somer ve ark., 2004). Ölçek 5 dereceli (tamamen uygun-hiç uygun değil) toplam 14 maddeden oluşmaktadır. Ölçeğin maddeleri Ek2'de verilmiştir. Ölçekten alınan yüksek puanlar, endişeli, kuruntulu, kötümser, gergin, kaygılı, kolay incinen, başkalarının onayına ihtiyaç duyan, kendini suçlamaya yatkın, hassas yapılı olma özelliklerini temsil ederken, düşük puanlar ise huzurlu, rahat, dirençli, gerçeklerle yüzleşebilen, ego gücünün yüksekliğine işaret etmekte, psikolojik dayanıklılığı temsil etmektedir.

Bu araştırmada kullanılan veriler için ölçeğin Cronbach-Alpha iç tutarlık güvenirlik katsayısı kadın katılımcılar için .84, erkek katılımcılar için de .83 olarak saptanmıştır. Ayrıca endişeye yatkınlık ölçeğinin tek bir boyutu ölçmesine ilişkin olarak yapılan açımlayıcı faktör analizi temel bileşenler yöntemi (herhangi bir faktör döndürme işlemi yapılmadan) sonucunda kadın katılımcıların birinci faktör özdeğerinin 4.76 ve açıklanan toplam varyans oranının da % 33.97 olduğu, erkek katılımcılar için de birinci faktör özdeğerinin 4.58 ve açıklanan toplam varyans oranının da % 32.74 olduğu bulunmuştur.

Ölçeğin gruplara göre dağılım özellikleri incelendiğinde, hem kadın (kayışlılık = -.04 ve basıklık = -.11) hem de erkek (kayışlılık = .36 ve basıklık = .00) gruplarında normal dağılımın elde edildiği ve varyans

homojenliğinin sağlandığı görülmüştür (F_{998} = .215, p = .64).

Normal dağılım özellikleri, içtutarlık güvenirlik analizi ve açımlayıcı faktör analizi sonuçları araştırmada kullanılan endişeye yatkınlık ölçeğinin analizler için gerekli olan temel varsayımları karşıladığına işaret etmektedir.

İşlem

Endişeye yatkınlık alt ölçeğinden elde edilen verilerin kadın ve erkek katılımcı grupları için ölçme eşdeğerliğinin incelenmesi amacıyla, söz konusu veriler MACS ve DIF yöntemleri kullanılarak analiz edilmiştir. MACS analizlerinde LISREL 8.8 (Jöreskog ve Sörbom, 2006); DIF analizlerinde PARSCALE 4.1 (Muraki ve Bock, 2002) programları kullanılmıştır.

Bulgular

MACS Sonuçları

Kadın ve erkek katılımcı gruplarında endişeye yatkınlık alt ölçeği için ölçme eşdeğerliği test edilirken önce temel model (*baseline model*) olarak kadın ve erkek katılımcı grupları için ayrı ayrı doğrulayıcı faktör analizi yapılmıştır. Doğrulayıcı faktör analizi yapılırken, ön analizler sonucunda gruplar arası farklılık gösterme olasılığı en az olan maddeler arasından madde 1 referans değişken olarak seçilmiş ve söz konusu maddenin faktör yükü her iki grup için modelde 1'e bağlanmıştır. Kadın ve erkek grupları için yapılan doğrulayıcı faktör analizine ilişkin model uyum indeksleri Tablo 1'de Temel Model 1 olarak verilmiştir.

Tablo 1'den de görülebileceği gibi, her iki cinsiyet grubu için de ölçme modelinin uyum indeks değerleri, model ile verinin orta düzeyde uyum gösterdiğine işaret etmektedir. Bu duruma ek olarak program, bazı maddelerin hata varyanslarının ilişkilendirilmesine yönelik düzeltme indeksleri (*modification index*) önermiştir. Temel modelde bu tür düzeltmelerin yapılması başlangıç düzeyindeki model uyumunu arttırmak açısından önerilmektedir (Byrne ve ark., 1989). Program tarafından kadın ve erkek grupları için önerilen ortak düzeltme indekslerinde en yüksek değere sahip olan 4 madde çifti seçilerek (madde 1-7, madde 6-7, madde 7-8, madde 10-12) bu maddelerin hataları ilişkilendirilmiş ve modeller yeniden test edilmiştir. Madde içeriklerine bakıldığında, bu madde çiftlerinin diğerlerine nazaran daha fazla benzerlik gösterdiği görülmektedir (örn., Moralim çabuk bozulur - Derin umutsuzluklara kapılırım). Bu nedenle hata ilişkilendirmeleri sadece görgül bir bulgu nedeniyle yapılmamış, içerik ve kuramsal açıdan da anlamlılığı göz önünde bulundurulmuştur. İkinci modele ilişkin uyum indeksleri Tablo 1'in alt kısmında Temel Model 2 olarak verilmiştir. Bu şekilde, kadın ve erkek grupları için söz konusu ölçeğe ilişkin iyi uyum değerleri elde edilmiş ve ölçme eşdeğerliğini incelemek üzere hiyerarşik analizlere geçilmiştir.

Ölçme eşdeğerliğinin incelenmesi amacıyla yapılan hiyerarşik analizlerin ilk aşamasında yapısal değişmezlik test edilmiştir. Yapısal değişmezlik test edilirken (giriş bölümünde örnek sentakslara atıflarla belirltildiği gibi), faktör sayısı ve gözlenen değişkenlerin örtük değişkene yüklenme örüntüsü sınırlandırılmıştır. Başka bir ifadeyle, bu aşamada grupların faktör yapılarının eş değer olup olmadığı aynı model içerisinde test edilmiştir. Tablo 2'de görülebileceği gibi, söz konusu model uyum indeksleri açısından yapısal değişmezliği sağlamaktadır.

Tablo 1. Temel Modellerde Doğrulayıcı Faktör Analizine İlişkin Uyum İyiliği İstatistikleri

		χ^2	sd	χ^2/sd	RMSEA*	CFI
Temel Model 1	Erkek	311.98	77	4.05	.08 (.07-.09)	.94
	Kadın	318.87	77	4.14	.08 (.07-.09)	.95
Temel Model 2	Erkek	239.04	73		.07 (.06-.08)	.95
	Kadın	198.73	73		.06 (.05-.07)	.97

* RMSEA değerlerine ilişkin güven aralıkları (% 90) parantez içerisinde verilmiştir.

Not: χ^2/sd için iyi uyum kriterleri = 3 ve altı, RMSEA için iyi uyum kriterleri = .08 ve altı, CFI için iyi uyum kriterleri = .95 ve üzeri

Tablo 2. MACS Modele İlişkin Uyum İyiliği ve Model Farkı İstatistikleri

	χ^2	sd	RMSEA*	CFI	$\Delta\chi^2$	Δsd**	ΔCFI
Yapısal Değişmezlik Modeli	438.66	147	.06 (.06-.07)	.96	-	-	-
Zayıf Değişmezlik Modeli	454.74	160	.06 (.05-.07)	.96	16.08	13 (27.7)	.00
Güçlü Değişmezlik Modeli	575.54	174	.07 (.06-.07)	.95	120.80	14 (29.1)	-.02
Kısmi Güçlü Değişmezlik Modeli	475.20	169	.06 (.05-.07)	.96	20.46	9 (21.7)	.00
Katı Değişmezlik Modeli	488.68	183	.06 (.05-.06)	.96	13.48	14 (29.1)	.00

* RMSEA değerlerine ilişkin güven aralıkları (% 90) parantez içerisinde verilmiştir.
** Δsd değerlerine ilişkin kritik χ^2 değerleri parantez içerisinde verilmiştir.

Analizin ikinci aşamasında test edilen zayıf değişmezlikte, yapısal değişmezliğe ek olarak grupların örtük değişkene ilişkin ölçme biriminin benzer olup olamadığı test edilmiştir. Zayıf değişmezlik test edilirken, faktör sayısı ve yüklenme örüntüsü ile birlikte faktör yükleri de sınırlandırılmıştır. Daha sonra zayıf değişmezlik modelinin χ^2 ve CFI değerlerinden yapısal değişmezlik modeli için elde edilen χ^2 ve CFI değerleri çıkarılarak zayıf değişmezlik modeli için ölçme eşdeğerliği incelenmiştir. Elde edilen $\Delta\chi^2$ ve ΔCFI değerleri modelde anlamlı düzeyde bir kötüleşme olmadığını göstermektedir (Tablo 2). Bu sonuç gruplar arasında zayıf değişmezliğin sağlandığına yani faktör yüklerinin iki grup için eşdeğer olduğuna işaret etmektedir. Ayrıca model faktör yükleri için herhangi bir düzeltme indeksi önermemiştir.

Analizin üçüncü aşaması olan güçlü değişmezlik test edilirken, ilk iki aşamada yapılan sınırlandırmalara ek olarak regresyon sabitleri de sınırlandırılmıştır. Zayıf ve güçlü değişmezlik modellerinden elde edilen ΔCFI ve $\Delta\chi^2$ değerleri modelde anlamlı bir kötüleşme olduğuna işaret etmektedir (Tablo 2). Bu bulgu maddelerden "en az bir tanesinin" sabit değerinin gruplar arasında ölçme eşdeğerliğini bozduğunu göstermektedir. MACS modellerinde hangi madde ya da maddelerde sabit değerinin gruplar arasında farklılık gösterdiğini incelemek üzere genellikle düzeltme indeksleri kullanılmaktadır. Eğer sabit değerler bakımından bir madde için gruplar arasında bir farklılık söz konusu ise, program o madde için regresyon sabitinin ikinci grupta serbest bırakılmasına ilişkin bir düzeltme indeksi (örn., madde1 = CONST) önermektedir. Güçlü değişmezlik modeli için önerilen düzeltme indeksleri incelendiğinde, dört maddenin (madde 5, 6, 9 ve 11) sabit değerlerinin gruplar arasında farklılaştığı görülmüştür.

Bu aşamada MACS modelleri için iki seçenek söz konusu olmaktadır. Birinci seçenek, hiyerarşik modeli test etmeyi bu aşamada sonlandırarak ölçme aracı için sadece zayıf değişmezlik sağlandığını rapor etmektir. İkinci seçenek ise, düzeltme indeksleri doğrultusunda farklılaşan maddelerin regresyon sabitlerinin ikinci grupta serbest bir şekilde tahminlenmesine izin vermektir. Bu şekilde düzenlenmiş modeller kısmi değişmezlik (*partial invariance*) olarak adlandırılmaktadır. Bu çalışmada, hiyerarşik modelin son aşamasına geçebilmek amacıyla kısmi değişmezlik modeli tercih edilmiştir.

Kısmi güçlü değişmezlik modeli sonuçları Tablo 2'de verilmiştir. Kısmi güçlü değişmezlik modelinin χ^2 ve CFI değerlerinden zayıf değişmezlik modelinin χ^2 ve CFI değerleri çıkarıldığında elde edilen $\Delta\chi^2$ ve ΔCFI değerleri modelde anlamlı düzeyde bir kötüleşme olmadığını göstermiştir (Tablo 2). Bu sonuç gruplar arasında kısmi güçlü değişmezliğin sağlandığı anlamına gelmektedir. Bu modelde serbest bırakılan yani eşdeğer olmayan dört maddenin regresyon sabitleri kadın ve erkek grupları için sırasıyla, madde 5 için 2.68 ve 2.93, madde 6 için 2.06 ve 2.24, madde 9 için 3.58 ve 3.21, madde 11 için 2.72 ve 2.42'dir.

MACS modeline ilişkin analizlerin son aşaması olan katı değişmezlik test edilirken, bütün parametre sınırlamaları ile birlikte hata varyansları da sınırlandırılmıştır. Kısmi güçlü değişmezlik ve katı değişmezlik modellerinden elde edilen ΔCFI ve $\Delta\chi^2$ değerleri modelde anlamlı bir kötüleşme olmadığına işaret etmektedir (Tablo 2).

Hiyerarşik analiz sonuçları genel olarak değerlendirildiğinde, endişeye yatkınlık alt ölçeğinde faktör yapısı ve örüntüsünün, faktör yüklerinin ve 4 madde dışında regresyon sabitlerinin kadın ve erkek grupları için eşdeğer olduğu görülmektedir. Elde edilen sonuçlara göre, yapısal ve zayıf değişmezliğin

tam olarak sağlandığı, diğer bir ifadeyle ölçek birimlerinin her iki grupta eşdeğer olduğu ancak skalar değişmezliğin kısmi olarak sağlandığı görülmüştür. Katı değişmezlik aşamasında da hata varyanslarının gruplar arasında farklılık göstermediği bulunmuştur.

DIF Sonuçları

Muraki ve Bock (1996) tarafından geliştirilen PARSCALE programı, referans ve fokal grupların madde parametre değerlerini doğrudan karşılaştırmaya ve bu parametrelerdeki farkların anlamlılığını Ki-kare ile test edilmesine imkan sağlamaktadır. Bu araştırmada DIF incelemeleri için özellikle likert tipi maddelere uygun olan ağırlıklandırılmış cevaplar modeli (*graded response model*; Samejima, 1997) çerçevesinde iki parametreli model kullanılarak madde parametreleri tahminlenmiştir. Karşılaştırma gruplarının madde ayırtetme ve madde güçlük parametre tahminleri yapılırken; önce referans (erkek) ve fokal (kadın) gruplar için madde parametreleri serbest olarak tahminlenmiş daha sonra ise, MACS modelin hiyerarşik basamaklarıyla uyum sağlayabilmesi amacıyla, eğim parametresi her iki grupta sabitlenerek yalnızca yerleşim parametresi tahminlenmiştir (Tablo 3). MACS modelde zayıf değişmezlik aşamasında faktör yükleri tahminlenirken regresyon sabitleri serbest bırakılmakta; regresyon sabitlerinin sınırlandığı güçlü değişmezlik aşamasında ise faktör yükleri modele halihazırda sınırlandırılmış olarak girmektedir (bkz., Ek-1b-c'deki hipotetik sentakslar). Madde ayırtetme diğer bir ifadeyle eğim parametresi açısından karşılaştırma gruplarının ölçekteki 10'uncu (χ^2_1 = 13.79, p < .001) maddede istatistiksel olarak anlamlı fark gösterdiği saptanmıştır. Bu madde (çabucak telaşlanırım) için kadın grubunun madde ayıretme değerinin 1.352, erkek grubunun madde ayırtetme değerinin de 1.732 olduğu görülmüştür. Bu değerler maddenin erkek grubunda kadın grubuna göre yüksek derecede ayırt etme düzeyine sahip olduğunu göstermektedir.

Karşılaştırma gruplarının örtük özellik yani θ üzerindeki konumunu belirleyen yerleşim parametresi tahminlerine bakıldığında; madde 5 (χ^2_1 = 14.27, p < .001), madde 6 (χ^2_1 = 9.43, p < .001), madde 9 (χ^2_1 = 20.96, p < .001) ve madde 11 (χ^2_1 = 10.75, p < .001) de gruplar arasında istatistiksel olarak anlamlı farklılaşmalar bulunmuştur. Yerleşim parametresi bakımından DIF bulunan maddelerin parametre değerleri kadın ve erkek grupları için sırasıyla, madde 5 için 0.616 ve 0.154, madde 6 için 1.225 ve 0.921, madde 9 için -1.233 ve -0.371, madde 11 için 0.631 ve 1.134'dür. Madde yerleşim parametre değerlerine bakıldığında; özellikle ölçeğin 9'uncu ve 11'inci maddelerinde kadın ve erkeklerin endişeye yatkınlık

Tablo 3. DIF Analizlerine İlişkin Madde Parametre Tahminleri ve Fark İstatistikleri

	Eğim (a_i) Parametresi				Yerleşim (b_i) Parametresi			
	Kadın	Erkek	Fark	χ^2	Kadın	Erkek	Fark	χ^2
M1	1.585	1.714	0.924	1.110	0.666	0.496	0.170	3.262
M2	0.946	0.955	0.990	0.017	1.738	1.618	0.120	0.699
M3	0.722	0.802	0.901	2.319	0.678	0.244	0.435	7.163
M4	1.616	1.528	1.057	0.510	0.406	0.424	-0.028	0.034
M5	1.105	1.143	0.967	0.229	0.616	0.154	0.463	14.270**
M6	1.579	1.805	0.874	3.060	1.225	0.921	0.304	9.429**
M7	1.635	1.606	1.018	0.042	-0.180	-0.190	0.010	0.010
M8	2.133	1.841	1.159	3.228	0.312	0.160	0.151	3.118
M9	0.678	0.697	0.973	0.149	-1.233	-0.371	-0.862	20.956**
M10	1.352	1.732	0.781	13.796**	0.242	0.321	-0.079	0.627
M11	0.890	0.791	1.125	2.480	0.631	1.134	-0.504	10.749**
M12	1.629	1.469	1.109	1.881	0.586	0.594	-0.008	0.008
M13	1.311	1.065	1.231	5.241	1.422	1.608	-0.186	2.357
M14	0.866	0.878	0.986	0.037	1.667	1.663	0.004	0.001

** p < .01

düzeylerini farklı konumlandırdıkları, diğer bir ifadeyle maddelerin puanlama anahtarı (tamamen uygun-hiç uygun değil) üzerinden değerlendirildiğinde kadın grubunun maddeyi daha kolay onaylama yönünde yanıtladıkları buna karşın erkek katılımcıların uygun değil yönüne doğru yanıt verdikleri söylenebilir.

Tartışma

Bu çalışmada ölçme eşdeğerliğini incelemek üzere ele alınan IRT'ye ve SEM'e dayalı iki yöntemin birçok ortak yönü olmakla birlikte bazı farklı yönleri de bulunmaktadır. Raju ve arkadaşlarının (2002) ifade ettiği gibi örtük özellik üzerinde aynı konumda olan kişilerin aynı gerçek puana sahip olması ölçme eşdeğerliğini sağlayan bir koşuldur. Bu yöntemlerde ölçme eşdeğerliği sağlanması için grupların ilgili boyut üzerindeki dağılımlarının eşdeğer olması bir gereklilik değildir. Her iki yöntemde de örtük bir yapıyla gözlenen değişkenler arasındaki ilişki incelenmektedir. SEM'de faktör yükü (λ) ve IRT'de ayırtetme (a_i) bu ilişkiyi tanımlayan temel parametrelerdir ve kavramsal olarak modellerde birbirlerine karşılık gelir.

Bu ilişkiyi incelerken IRT doğrusal olmayan matematiksel fonksiyonları kullanırken, SEM doğrusal fonksiyonlar üzerinde modellenmiştir. Bunun bir sonucu olarak orta puanlarda bireylerin gerçek puanlarını tahmin etmede iki yöntem yakın sonuçlar verirken; uç puanlar söz konusu olduğunda IRT modellemeleri daha hassas ve daha doğru kestirimlerde bulunmaktadır. Çünkü doğrusal olmayan fonksiyonlar psikolojik değişkenlerin dağılımlarına daha uygundur ve günlük hayat verilerini daha iyi temsil ederler. Dolayısıyla IRT modelleri ölçme eşdeğerliğini uç puanlarda regresyon modellerine göre daha doğru olarak test etmektedir.

IRT'de madde güçlük ya da yerleşim parametresi (b_i) ve SEM'de regresyon sabiti (τ) benzer olarak yorumlanmakla birlikte bazı farklılıklar göstermektedir. Ferrando'nun (1996) belirttiği gibi her iki yaklaşımda da bu değerler madde ortalamalarının örtük özellikte belirli bir değere sabitlendiğinde elde edilen değerlerdir. IRT'de yerleşim parametresi doğru cevap verme olasılığı .50 olduğunda elde edilen değerken, SEM'de regresyon sabiti örtük özellik ortalaması 0 iken elde edilen değere karşılık gelir (Chan, 2000).

SEM'de skalar değişmezlik için (regresyon sabiti eşdeğerliği) metrik değişmezliğin (faktör yüklerinin-ölçme birimlerinin eşdeğerliği) sağlanması bir ön koşul iken IRT'de parametre tahminleri eş zamanlı olarak yapılabilmektedir. Bu çalışmada SEM'in bu özelliği göz önünde bulundurularak IRT'den elde edilen parametreleri daha karşılaştırabilir hale getirmek için IRT analizlerinde parametre tahminlerinin basamakları da bulgular kısmında ifade edildiği gibi MACS'ın basamaklarına göre uyarlanmıştır.

Ölçme eşdeğerliği çalışmalarında tek boyutluluk varsayımının faktör analizi yoluyla incelenmesi IRT'ye dayalı modellerde model parametrelerinin tahminlenmesi ile eş zamanlı olarak yapılamazken, bu varsayım SEM'de eş zamanlı olarak test edilmektedir. SEM'in bir diğer avantajı ise hata varyansları arasındaki ilişkilerin incelenmesi yoluyla halihazır modelle açıklanamayan alt yapılar hakkında ipucu sağlamasıdır. Öte yandan IRT incelemeleri bu tür bir bilgi sağlayamamaktadır.

Bu çalışmadan elde edilen sonuçlara göre MACS ve DIF kapsamında kadın ve erkek karşılaştırma grupları arasında bazı maddelerde ölçme eşdeğerliğinin sağlanamadığı bulunmuştur. Ölçme eşdeğerliğinde farklılık gösteren maddeler her iki yöntemde de büyük oranda ortak maddeleri içermektedir. MACS analizleri sonucunda Endişeye Yatkınlık Ölçeğinin faktör yüklerine karşılık gelen λ parametrelerinin tüm maddeler için kadın ve erkek gruplarında eşdeğer olduğu bulunurken, DIF analizleri sonucunda ölçekte yer alan 10'uncu maddenin ayırtetme parametresinin (a_i) farklılaştığı görülmüştür. Ayırtetme parametresine baktığımızda bu maddenin ayırt edicilik düzeyi her iki grupta da yüksek olmasına rağmen, erkek katılımcıları kadın katılımcılara göre daha yüksek düzeyde ayırt ettiği, diğer bir ifadeyle, maddenin endişeye yatkınlık özelliğine bağlanması erkek grubunda daha kuvvetli iken, kadın grubunda daha düşük düzeydedir. Maddenin içeriği (Çabucak telaşlanırım.) değerlendirildiğinde; kadın denekler erkeklere göre kendilerini daha geniş bir ranjda telaşlı olarak tanımlarken (ICC daha yaygın), erkek denekler bu maddeyi onaylamakta daha kesin bir tutum göstermektedirler (ICC daha dik). Bu farkın "telaşlı" olma sıfatının kültürel açıdan erkeklerden ziyade kadınlara yakıştırılan bir özellik olmasından kaynaklandığı düşünülebilir. Sonuç olarak her iki yöntemden elde edilen bulgular bir madde dışında gruplar arasında metrik eşitliğinin sağlandığına işaret etmektedir.

MACS analizlerinde regresyon sabitinin (τ) ve DIF analizlerinde madde yerleşim parametrelerinin (b_i) değişmezliğine ilişkin bulgular ele alındığında; her iki yaklaşımda da Endişeye Yatkınlık Ölçeğinin 5., 6., 9., ve 11'inci maddelerinin regresyon sabiti/yerleşim parametrelerinin gruplar arasında farklılaştığı görülmüştür. 5. madde (Bir şeylerin kötü sonuçlanacağını düşünürüm.) için yerleşim parametreleri kadın için .62 erkek için ise .15 iken, 6. madde için (Kendimi kolayca tehdit altında hissederim.) yerleşim parametreleri kadın için 1.23 erkek için ise

.92 bulunmuştur. Söz konusu maddeler için regresyon sabit değerleri ise 5. madde için kadın grubunda 2.68 erkek grubunda 2.93, 6. madde için ise kadın grubunda 2.06 erkek grubunda 2.24 olarak bulunmuştur. Söz konusu iki maddenin madde yerleşim/sabit parametrelerinin erkekler yönünde yanlı olduğu görülmektedir. Öte yandan MACS modelde tahminlenen sabit değerleri yanlılığın olduğu grupta daha yüksek değerlerde karşımıza çıkmakla beraber, DIF'de tahminlenen yerleşim parametre değerleri yanlılığın olduğu grupta daha düşük değerlerde karşımıza çıkmaktadır. Bu aradaki farklılık her iki yaklaşımdaki ölçekleme yönteminden kaynaklanmaktadır. Elde edilen bulgular her iki yöntemde de maddelerin erkekler yönünde daha popüler maddeler olduğuna işaret etmektedir. 9. (Başkalarının onayına ihtiyaç duyarım.) ve 11. (Korunmaya ihtiyaç duyarım.) maddelerde ise bu durumun tam tersi gözlenmiş yani bu maddeler her iki yöntemde de kadınlar yönünde yanlılık olduğuna ilişkin sonuçlar vermiştir. "Korunmaya ihtiyaç duyarım" ve "başkalarının onayına ihtiyaç duyarım" maddelerini kadınların daha kolay bir şekilde onayladığı görülürken "bir şeylerin kötü sonuçlanacağını düşünürüm" ve "kendimi kolayca tehdit altında hissederim" maddelerini ise erkeklerin daha kolay bir şekilde onayladığı gözlenmiştir. Bulgular, sözü edilen 4 madde dışında ölçek için skalar düzeyde değişmezliğin sağlandığına işaret etmektedir.

Sonuç olarak, kadın ve erkek karşılaştırma grupları için bir kişilik alt ölçeğinden elde edilen veriler kullanılarak, ölçme eşdeğerliğinin MACS ve DIF yöntemleri kapsamında incelendiği bu çalışmada, ölçme eşdeğerliğinin sağlanamadığı durumlarda her iki yöntemin de bunun kaynaklarının araştırılmasında önemli ipuçları sağladığı görülmektedir.

Ölçme eşdeğerliğinin incelenmesinde DIF yöntemleri çoğunlukla madde parametrelerinin karşılaştırılmasına dayanmaktadır. Elde edilen sonuçlar her madde için ayrı ayrı parametreleri bazında fark (*contrast*) ve bu farklara ilişkin anlamlılık değerleri vermektedir. Madde parametreleri bazında fark düzeylerinin program sonuçlarında izlenebilmesi etki büyüklüklerini görmek açısından da önem taşımaktadır. Bu bağlamda maddelerin tek tek psikometrik özelliklerinin incelemek açısından DIF analizleri ayrıntılı bilgiler sağlamaktadır. Maddelerin farklı gruplar için farklı yönlerde sağladığı avantaj ve dezavantajları göz önünde bulunduran telafi edici DIF yöntemlerinde (Raju ve ark., 1995; 2002) madde bazında sonuçların yanı sıra toplam test düzeyinde de ölçme eşdeğerliği sonuçlarını değerlendirmek mümkün olmaktadır. Yapısal eşitlik modellemelerine dayalı ölçme eşdeğerliğinin incelenmesinde ise empoze edilen modelin ve eşitliklerin uygunluğu elde edilen uyum indekslerinin incelenmesi sonucu değerlendirilmektedir. Model uyumlu bulunduğunda ölçme eşdeğerliğinin sağlandığı varsayılmaktadır. Ölçme eşdeğerliğinin farklı düzeyleri modeller arası farklar yoluyla incelenmektedir. Parametreler bazında sonuçlar ise düzeltme indekslerinden çıkarsanmaktadır. Ancak bu farklılıklara ilişkin etki büyüklüklerinin görülmesi IRT program çıktılarındaki kadar net ve pratik değildir. Buna karşılık toplam test bazındaki ölçüm eşdeğerliği, SEM yöntemlerinin hepsinde uyum indeksleri yoluyla daha net bir şekilde değerlendirilebilmektedir. Ayrıca çok boyutlu yapılar söz konusu olduğunda SEM yöntemleri hem maddelerin boyutlara bağlanmalarını hem de madde parametrelerinin birlikte aynı model içerisinde eş zamanlı olarak değerlendirilmesini mümkün kılmaktadır. Test geliştirme çalışmalarında örneğin başarı testlerinde olduğu gibi geniş bir madde havuzu üzerinde maddelerin gruplara ilişkin yanlılık özellikleri tek tek incelenmek istenildiğinde çalışmalara IRT yöntemleri ile başlamak daha pratik ve uygun olmaktadır. Madde analizleri sonucu oluşturulmuş ölçek ya da alt ölçeklerin incelenmesi aşamasında ise SEM yöntemlerinden yararlanmak daha uygundur. Mevcut ölçeklerin gruplar ve kültürler arası farklılıklarının incelenmesinde toplam test bazında sonuçlar vermesi açısından SEM yöntemleri daha pratik görünmektedir. Ancak Yapısal Eşitlik Modellerinde madde sayısı çok olduğunda toplam ölçek düzeyinde tahminlenmesi gereken parametre sayısının çoğalmasına bağlı olarak uyumlu modeller elde etmek güçleşmektedir. Bu güçlüğü gidermek maddelerin parseller altında toplanarak yeni değişkenler olarak modele girilmesiyle mümkün olabilmektedir. Bu durumda da maddelerin tek tek etkileri görülememektedir. Bu bağlamda yapılan çalışmanın özelliklerine ve amacına göre tek tek IRT ya da SEM yöntemleri kullanılabileceği gibi, her iki yöntemden birlikte birbirini destekleyici şekilde yararlanmak da uygun olabilmektedir.

Çalışmamız sonuçlarında her iki yöntemle elde edilen bulgular benzer sonuçlar vermekle birlikte, söz konusu yöntemlerin yukarıda söz edilen avantaj ve dezavantajları göz önünde bulundurulduğunda, ölçme eşdeğerliği çalışmalarında her iki yöntemden birlikte yararlanmanın birbirini destekleyen sonuçlar elde etme bakımından önemli olduğu düşünülmektedir.

Kaynaklar

Byrne, B. M., Shavelson, R. J. ve Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466.

Byrne, B. M. ve Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order

structure: A walk through the process. *Structural Equating Modeling, 13*(2), 287-321.

Camilli, G. ve Shepard, L. A. (1994). *Methods for identifying biased test items*. Sage Publication: London.

Chan, D. (2000). Detection of differential item functioning on the Kirton Adaption - Innovation Inventory using multiple-group mean and covariance structure analysis. *Multivariate Behavioral Research, 35*(2), 169-199.

Chernyshenko, O. S., Stark, S., Chan, K. Y., Drasgow, F. ve Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36*(4), 523-562.

Collins, W. C., Raju, S. N. ve Edwards, J. E. (2000). Assessing differential functioning in a satisfaction scale. *Journal of Applied Psychology, 85*(3), 451-461.

Ferrando, P. J. (1996). Calibration of invariant item parameters in a continuous item response model using the extended lisrel measurement submodel. *Multivariate Behavioral Research, 31*(4), 419-439.

Flowers, C. P., Oshima, T. C. ve Raju, N. S. (1999). A description and demonstration of the polytomous-DFIT framework. *Applied Psychological Measurement, 23* (4), 309-326.

Hambleton, R. K, Swaminathan, H. ve Rogers, H. J. (1991). *Fundamentals of item response theory.* Sage Publications: London.

Jöreskog, K. G. ve Sörbom, D. (2006). *LISREL (Version 8.8) [Computer software]*. Chicago: Scientific Softare International Inc.

Korkmaz, M. (2006). Test ve ölçek geliştirmede yeni yaklaşımlar: Madde cevap kuramı kapsamında madde işlevsel farklılığı (madde yanlılık) yöntemleri. *Türk Psikoloji Yazıları, 9*(18), 63-80.

Little, T. D. (1997). Mean and covariance structure (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*(1), 53-76.

Lubke, G. H., Dolan, C. V., Kelderman, H. ve Mellenberg, G. J. (2003). Weak measurement invariance with respect to unmeasured variables: An implication of strict factorial invariance. *British Journal of Mathematical and Statistical Psychology, 56*, 231-248.

Maurer, T. J., Raju, S. N. ve Collins, W. C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology, 83*(5), 693-702.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*, 127-143.

Mellenbergh, G. J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research, 29* (3), 223-236.

Meredith, W. (1993). MI, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.

Meredith, W. ve Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement invariance. *Psychometrika, 57*(2), 289-311.

Morales, L. S., Reise, S. P. ve Hays, R. D. (2000). Evaluating the equivalence of health care ratings by Whites and Hispanics. *Medical Care, 38*(5), 517-527.

Muraki, E. ve Bock, R. D. (2002). *PARSCALE (Version 4.1) [Computer software]*. Chicago: Scientific Software International Inc.

Orlando, M. ve Marshall, G.N. (2002). Differential item functioning in a Spanish Translation of the PTSD checklist: Detection and evaluation of impact. *Psychological Assessment, 14*(1), 50-59.

Raju, N. S., Laffitte, L. J. ve Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*(3), 517-529.

Raju, N. S., Van der Linden, W. J. ve Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement, 19* (4), 353-368.

Reise, S. P., Smith, L. ve Furr, R. M. (2001). Invariance on the NEO PI-R neuroticism scale. *Multivariate Behavioral Research, 36*(1), 83-110.

Samejima, F. (1997). Graded response model. Van der Linden W. J. ve Hambleton R. K., (Ed), *Handbook of Modern Item Response Theory*. NewYork: Springer-Verlag.

Smith, L. L. (2002). On the usefulness of item bias analysis to personality psychology. *Personality and Social Psychology Bulletin, 28*(6), 754-763.

Somer, O. (1998). Kişilik testlerinde klasik ve modern test kuramları ile madde analizi. *Türk Psikoloji Dergisi, 13*(41), 1-15.

Somer, O. (1999). Çok kategorili (polytomous) maddelerde klasik ve modern test kuramlarının madde analizleri, güvenirlik ve bilgi kavramları açısından karşılaştırılması. *Türk Psikoloji Dergisi, 14*(44), 63-75.

Somer, O. (2004). Gruplararası karşılaştırmalarda ölçek eş-değerliğinin incelenmesi: Madde ve test fonksiyonlarının farklılaşması. *Türk Psikoloji Dergisi, 19* (53), 69-82.

Somer, O., Korkmaz, M. ve Tatar, A. (2004). *Kuramdan uygulamaya beş faktör kişilik modeli ve beş faktör kişilik envanteri (5FKE)*. Ege Üniversitesi Edebiyat Fakültesi, Yayın No: 128, İzmir.

Stark, S., Chernyshenko, O. S. ve Drasgow, F. (2006). Detecting differential item functioning with comfirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*(6), 1292-1306.

Thissen, D., Steinberg, L. ve Wainer, T. (1988). Use of item response theory in the study of group differences in trace lines. H. Wainer ve H. I. Braun, (Ed.), *Test validity* içinde (147-169). New Jersey: Lawrence Erlbaum Associates Inc.

Van de Vijyer, F. ve Leung, K. (1997). Methods and data analysis of comparative research. J. W. Berry ve Y. H. Poortinga, (Ed.), *Handbook of cross-cultural psychology, Vol.1: Theory and method* içinde (259-300). Needham Heights, MA: Alleyn & Bacon.

Vandenberg, R. J. ve Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.

Wu, A. D., Li, Z. ve Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation, 12*, 1-26.

Zickar, M. J. (1998). Modeling item-level data with item response theory. *Current Directions in Psychological Science, 7*(4), 104-109.

Ek 1a. Yapısal Değişmezlik Modeli için LISREL Sentaksı

```
Group 1: GRUP1

Observed Variables:
GD1[1] GD2 GD3 GD4

Covariance Matrix from File grup1.COV
Means from File grup1.DAT
Sample Size: 200

Latent Variables:
od[2]

Equations:
GD1 = CONST + 1*od
GD2 = CONST + od
GD3 = CONST + od
GD4 = CONST + od

Group 2: GRUP2

Covariance Matrix from File grup2.COV
Means from File grup2.DAT
Sample Size: 200

Latent Variables:
od

Equations:
GD1 = CONST + 1*od
GD2 = CONST + od
GD3 = CONST + od
GD4 = CONST + od

Set the error variance of GD1 free
Set the error variance of GD2 free
Set the error variance of GD3 free
Set the error variance of GD4 free

Path Diagram
End of Problem
```

Ek 1b. Zayıf Değişmezlik Modeli için LISREL Sentaksı

```
Group 1: GRUP1

Observed Variables:
GD1 GD2 GD3 GD4

Covariance Matrix from File grup1.COV
Means from File grup1.DAT
Sample Size: 200

Latent Variables:
od

Equations:
GD1 = CONST + 1*od
GD2 = CONST + od
GD3 = CONST + od
GD4 = CONST + od

Group 2: GRUP2

Covariance Matrix from File grup2.COV
Means from File grup2.DAT
Sample Size: 200

Latent Variables:
od

Equations:
GD1 = CONST
GD2 = CONST
GD3 = CONST
GD4 = CONST

Set the error variance of GD1 free
Set the error variance of GD2 free
Set the error variance of GD3 free
Set the error variance of GD4 free

Path Diagram
End of Problem
```

[1] GD = Gözlenen Değişken
[2] od = Örtük Değişken

Ek 1c. Güçlü Değişmezlik Modeli için LISREL Sentaksı

```
Group 1: GRUP1

Observed Variables:
GD1 GD2 GD3 GD4

Covariance Matrix from File grup1.COV
Means from File grup1.DAT
Sample Size: 200

Latent Variables:
od

Equations:
GD1 = CONST + 1*od
GD2 = CONST + od
GD3 = CONST + od
GD4 = CONST + od

Group 2: GRUP2

Covariance Matrix from File grup2.COV
Means from File grup2.DAT
Sample Size: 200

Latent Variables:
od

Equations:

Set the error variance of GD1 free
Set the error variance of GD2 free
Set the error variance of GD3 free
Set the error variance of GD4 free

Path Diagram
End of Problem
```

Ek 1d. Katı Değişmezlik Modeli için LISREL Sentaksı

```
Group 1: GRUP1

Observed Variables:
GD1 GD2 GD3 GD4

Covariance Matrix from File grup1.COV
Means from File grup1.DAT
Sample Size: 200

Latent Variables:
od

Equations:
GD1 = CONST + 1*od
GD2 = CONST + od
GD3 = CONST + od
GD4 = CONST + od

Group 2: GRUP2

Covariance Matrix from File grup2.COV
Means from File grup2.DAT
Sample Size: 200

Latent Variables:
od

Equations:

Path Diagram
End of Problem
```

Ek 2. Endişeye Yatkınlık Alt Ölçeğinin Maddeleri

1. Derin umutsuzluklara kapılırım.
2. Her yerde tehlike görürüm.
3. Geçmiş hatalarımı düşünerek zaman harcarım.
4. Her şeye endişelenirim.
5. Bir şeylerin kötü sonuçlanacağını düşünürüm.
6. Kendimi kolayca tehdit altında hissederim.
7. Moralim çabuk bozulur.
8. Kolayca huzursuz olurum.
9. Başkalarının onayına ihtiyaç duyarım.
10. Çabucak telaşlanırım.
11. Korunmaya ihtiyaç duyarım.
12. Kolayca kendimi baskı altında hissederim.
13. Gözüm kolayca korkar.
14. Genelde rahatımdır.

Turkish Journal of Psychology, December 2009, *24* (64), 76-77

Summary
Detection of Measurement Equivalence by Structural Equation Modeling and Item Response Theory

Oya Somer, Ege Üniversitesi
Mediha Korkmaz, Ege Üniversitesi
Seda Dural, Ege Üniversitesi
Seda Can, İzmir University of Economics

Measurement equivalence is one of the important prerequisites to make valid across groups' latent trait comparisons. In measurement equivalence condition, the probability of the individual's having a specific observed score is independent from his/her group membership (Mellenbergh, 1989; Meredith, 1993; Meredith & Millsap, 1992). According to this definition, individuals from different groups with the same true score will get the same observed score. Otherwise, the differences obtained from group comparisons can be controversial (Chan, 2000; Somer, 2004; Stark et al., 2006).

Two approaches are frequently used in detection of measurement equivalence. One of them is based on Item Response Theory and referred as Differential Item and Test Functioning (DIF and DTF) and the other is based on the Structural Equation Modeling (SEM). In the framework of SEM, two models are generally used in measurement equivalence studies. The most widespread one is the Multi Group Confirmatory Factor Analysis - MGCFA that is based on testing the equivalence of covariance structures. The second one is the models that analyze the equivalence of Mean and Covariance Structure - MACS by using both covariance and mean structures.

In this study, the data of a personality scale obtained from male and female groups was subjected to MACS and DIF analysis for detection of measurement equivalence.

Generally, the investigation of measurement equivalence in MACS includes the testing of four nested hierarchical models (Byrne et al., 1989; Byrne & Stewart, 2006; Chan, 2000; Little, 1997; Stark et al., 2006; Vandenberg & Lance, 2000; Wu, et al., 2007):

Configural Invariance: In the configural invariance model, whether the groups have the same factorial structure is investigated. While testing the configural invariance, only numbers of factor and loading patterns are constrained across the groups (Vandenberg & Lance, 2000; Wu et al., 2007).

Weak Invariance: In the weak invariance model, whether the groups have the same factor loadings is investigated. While testing the weak invariance, in addition to the numbers of factor and loading patterns, factor loadings are also constrained across the groups. Weak invariance is also called metric invariance (Vandenberg & Lance, 2000; Wu et al., 2007).

Strong Invariance: In the strong invariance model, whether the groups have the same intercept values is investigated. While testing the strong invariance, in addition to the numbers of factor, loading patterns and factor loadings, intercepts are also constrained across the groups. Strong invariance is also called scalar invariance (Vandenberg & Lance, 2000; Wu, et al., 2007).

Strict Invariance: In the strict invariance model, whether the groups have the same error variances is investigated. While testing the strict invariance, in addition to the numbers of factor, loading patterns, factor loadings and intercepts, error variances are also constrained across the groups (Vandenberg & Lance, 2000; Wu, et al., 2007).

While estimating item discrimination and item difficulty parameters in comparison groups in the DIF analysis, firstly, item parameters were freely estimated for two groups and then only location parameter was estimated by fixing the slope parameter in both groups in order to make meaningful comparisons with the hierarchical stages of MACS model.

Method

Participants

The sample has consisted of 500 female and 500 male participants who were selected from the adult norm sample of Big Five Personality Inventory-BFPI (Somer et al., 2004). The mean age of female and male participants were 27.23 (SD = 11.20) and 28.85

Address for Correspondence: Yrd. Doç. Dr. Mediha Korkmaz, Ege Üniversitesi Edebiyat Fakültesi Psikoloji Bölümü, Bornova - İzmir.
E-mail: mediha.korkmaz@ege.edu.tr

(*SD* = 11.70), respectively.

Materials

"Proneness to Anxiety" subscale of "Emotional Instability" factor of BFPI was used in this study (Somer et al., 2004). The subscale has consisted of 14 items with a 5-point Likert scale (1 = strongly disagree and 5 = strongly agree). Cronbach Alpha coefficients of the subscale were .84 for female participants and .83 for male participants.

Normal distribution properties and the results of factor analysis indicated that the unidimensionality and normality assumptions for analysis were provided by the proneness to anxiety subscale used in the study.

Procedure

In order to investigate measurement invariance for the female and male participants, the data obtained from the Proneness to Anxiety subscale were analyzed by using MACS and DIF methods. In the MACS analysis LISREL 8.8 (Jöreskog & Sörbom, 2006) and in the DIF analysis PARSCALE 4.1 (Muraki & Bock, 2002) programs were used.

Results and Discussion

MACS Results

In order to test measurement invariance, firstly confirmatory factor analysis (CFA) was performed for female and male groups separately as a baseline model. Item 1 was selected as a reference indicator (which has the lowest probability of differing between the groups in the result of preliminary analysis) and the factor loading of this item was fixed to 1 in the model for the both groups.

Fit index values indicated that the data fit the model moderately for both groups. Modification indexes were suggested by the program to correlate error variances of some item pairs. Such kind of modifications was offered to establish a well-fitting baseline model (Byrne et al., 1989).

Among the suggested modifications, 4 item pairs were selected (item 1-7, item 6-7, item 7-8 and item 10-12) which are common for two groups with the highest index values. The models were tested again by correlating the error variances of these item pairs and the obtained fit indexes were found to be quite satisfactory to begin hierarchical analysis.

Both the configural invari-ance model and the weak invariance model provided good fits to the data. On the other hand, ΔCFI and $\Delta\chi^2$ values obtained from the difference between weak and strong invariance model indicated that the intercept of at least one of the items worsened measurement invariance between groups. When the suggested modification indexes for the strong invariance model were examined, it was seen that the intercepts of the 4 items (item 5, 6, 9 and 11) were not invariant across groups. For this reason, the intercepts of these items were freely estimated in the groups; thus, the model turned out to be partial strong invariance model. ΔCFI and $\Delta\chi^2$ values obtained from the difference between weak and partial strong invariance model indicated that the strong invariance was hold partially. The intercept values of the 4 items in the partial strong invariance model were found to be 2.68 and 2.93 for item 5, 2.06 and 2.24 for item 6, 3.58 and 3.21 for item 9, 2.72 and 2.42 for item 11 in the female and male groups, respectively. The strict invariance model which is the last level of the hierarchical analysis indicated that the error variances were invariant across the groups.

DIF Results

The PARSCALE program which was developed by Muraki and Bock (1996) allows to compare item parameter estimates of reference and focal groups directly and to test the significance of these parameter contrasts by using Chi-square. Item parameters were estimated using two parameter graded response model (Samejima, 1997). In the estimation of item discrimination and item difficulty parameters, firstly item parameters were freely estimated for the reference (male) and the focal (female) groups and then only location parameter was estimated by fixing the slope parameter in both groups in order to be concordant with the hierarchical stages of MACS model. For the slope parameter (item discrimination parameter) it was found that item 10 in the subscale was functioning significantly different in comparison groups. Item discrimination values of female and male groups were found to be 1.35 and 1.73 respectively for this item. These values indicated that this item has higher discrimination value in the male group than females.

The location parameters of the items 5, 6, 9 and 11 were found to be differently functioning in comparison groups. The location parameter values of the 4 DIF items were 0.616 and 0.154 for item 5, 1.225 and 0.921 for item 6, -1.233 and -0.371 for item 9, 0.631 and 1.134 for item 11 in the female and male groups, respectively.

Our findings indicated that the two approaches are supplying quite similar results. On the other hand, both methods have some advantages and disadvantages (Chan, 2000; Ferrando, 1996; Raju et al., 2002; Stark et al., 2006). Depending on the goal of the studies, both methods can be used separately or can be used simultaneously to support the findings of each other.