

Ölçme Eşdeğerliği

N. Bilge (UZUN) BAŞUSTA*

Hacettepe Üniversitesi

Özet

Bu çalışma; ölçmelerin eşdeğerliği konusunun teorik yapısının anlaşılmasını ve çok örneklemlili doğrulayıcı faktör analizi(MGCFA) yönteminin ölçme eşdeğerliğinde nasıl işe koşulduğunu betimlemeyi amaçlayan bir derlemedir. Anlamli karşılaştırmalar yapmak için kullanılan ölçme araçlarının birde fazla grupta aynı yapıda olması gerekmektedir. Ölçme eşdeğerliği ile ilgili çalışmalar çok maddeli ölçme araçlarında iç tutarlılık ve dışsal ilişkileri araştırmaktadır. Son yıllarda yapılan ölçmelerin eşdeğerliği çalışmalarında araştırmacılar eşdeğerliğin dört biçimine vurgu yapmaktadır. Ölçme eşdeğerliği ile yürütülen çalışmalar kültürel yanlılıklardan kaynaklanan hataları, çevirmeden kaynaklanan hataları ve diğer faktörlerin meydana getireceği hataları önler. Ayrıca belli bir teorik tanımın çerçevesinin daha doğru belirlenmesinde ve geliştirilmesinde yararlı bir istatistiksel tekniktir.

Anahtar Sözcükler: Ölçme Eşdeğerliği, Çok Örneklemlili Doğrulayıcı Faktör Analizi (MGCFA), Yapısal Eşitlik Modeli (YEM)

Abstract

In order to make a meaningful comparison, instruments must have similar measurement qualities across settings. Research on measurement invariance has typically focused on internal structure of multi-item instruments and external relations. In the recent literature, researchers have paid attention to four forms of invariance. The researchs which include the measurement invariance tests can detect psychometric problems that result from cultural biases in item content, translation errors, acquiescence biases, and other factors. Also Statistical techniques to define and assess measurement invariance are helpful and useful to develop a framework for theoretical definition.

Keywords: Measurement Invariance, MGCFA, Structural equation modeling (SEM)

Araştırmalarda farklı kültürlere, bölgelere, şehirlere vs. karşılaştırmalara dayanan çalışmaların göz ardı edilmeyecek kadar fazla olduğu bilinmektedir. Literatürde makalelerin büyük bir bölümünün alt grup karşılaştırmaları ile ilgili olduğunu söylemek yanlış olmayacaktır. Çünkü temeli bireye dayanan araştırmaların temel noktalarından biri bireye ya da bireylere dair sosyal yapının ortaya konmasıdır. Bir yapının ortaya konmasında en temel ilke ise o yapıya dair ölçmeler yapmaktan geçer.

Ölçme konusu olan; açığa çıkarılmak istenen belli bir özelliktir. Varlıkların birden çok özelliği vardır. Oysa ölçme belli bir özelliğin varlıklarda sahip oluş derecesini ortaya çıkarmak için ortaya çıkan bir kavramdır. Varlıkların belli bir özellik bakımından bulunduğu yeri tespit etmek için bu özelliği ölçmeye yönelik ölçme araçları kullanılır.

Bir ölçme aracı modeli 'farklı gruplarda aynı özelliği ölçer' kabulü ile hazırlanır. Eğer bu kabul onay alabilirse yapılan karşılaştırma ve puanlara ait analizlerin doğruluğu anlamlı olacaktır. Benzer şekilde bu kabul doğrulanamıyorsa o zaman yapılan analiz ve karşılaştırmalar anlamlılığını kaybedecektir.

Ölçme aracı özelliğin oluş derecesini ortaya koyarken, çeşitli alt gruplarda bu özellik açısından karşılaştırmalar yapma ve bu biçimde bilgi sağlama yoluna gider. Ardından farklı varlıkların aynı özellik bakımından karşılaştırılması ile varlıklar belli özellikleri bakımından farklı biçimde tanımlanır. Bu tanımlamanın doğruluğu anlaşılacağı gibi ölçme aracı ve elde edilen ölçüm ile doğrudan ilgilidir.

* Arş. Gör., Hacettepe Üniversitesi Tıp Fakültesi, Tıp Eğitimi ve Bilişimi Anabilim Dalı
buzun@hacettepe.edu.tr

Ölçme Eşdeğerliği

Yapılan bir ölçmeden elde edilen sonuçlar bireylerin farklı özellikler taşıması nedeniyle farklı olacaktır. Ancak bu farklılığın sadece birey özelliklerine bağlanması ve bu biçimde açıklanması çoğu zaman doğru değildir. Çünkü bireyler arasındaki bu farklılık ölçme aracının kendinden kaynaklanıyor olabilir.

İki farklı tedavi yönteminin uygulandığı bir çalışmada gruplardan biri üzerinde yeni bir tedavi yöntemi uygulanırken, diğer grup üzerinde hastalığın standart tedavisi uygulanmaktadır. Bu tip çalışmalarda ilgilenilen, yaşam ya da tedavi kalitesinde artış sağlanıp sağlanmadığıdır. Bu her iki grup için aynı anlama gelmektedir. Aynı biçimde gruplar arasındaki farklılığın nedeni de açıktır. Fakat birçok çalışmanın ilgi konusu daha karmaşık olabilmektedir. Örneğin Amerikalı işçilerden oluşan bir grup ile Çin işçilerden oluşan bir grubun iş memnuniyeti derecesi araştırılıyor olsun. İş memnuniyetini değerlendirmek için anket tipi ölçme araçlarından biri kullanılacaktır. Bu ankette elde edilen yanıtlar ise, grupların anket maddelerini yorumlama biçimine göre değişecektir (Cheug & Rensvold; 1998). Söz konusu ölçeğin farklı ülkelerde farklı formları kullanılarak, uzmanlarla dille bağlantılı olan ayrılıkların giderilmesi yönünde çalışmalar yapılırsa da, çeviri çalışmaları ile ölçeğin farklı ülkelerde aynı anlama geldiği ve anketi cevaplayan bireylerce aynı biçimde yorumlandığı garanti edilmiş olmayacaktır (Brislin, Lonner, Thorndike; 1978; aktaran Cheug&Rensvold; 1998).

Türkiye’de, belli bir sektörde çalışma verimini artırmak amacıyla, farklı bölgelerde aynı birimde çalışan bireylerden belli bilgiler sağlayacak bir anket uygulandığını düşünelim. Bu ankette çalışan memnuniyetine yönelik şöyle bir madde kullanılsın.

“Sabahları işe gitmek bana zulüm geliyor” (Örnek 1)

Çeviri ile ilgili eşitsizliği ortadan kaldırdığımız durumlarda bile anket maddelerinden kaynaklanan ve bireyin yanıtını değiştirebilecek durumlar söz konusu olabilmektedir. Yukarıdaki maddenin bulunduğu ölçek iki farklı ülkede değil, ülkemizin iki farklı bölgesinde uygulanmasına rağmen ölçme aracına bağlı farklılıklar ortaya çıkabilecektir. Ölçeğin bu maddesi; İstanbul’ da yaşayan biri için, yoğun trafikten duyduğu endişe anlamına gelirken; Ağrı şehrinde ölçeği yanıtlayan biri için dondurucu soğukun sabahları onda yarattığı rahatsızlık anlamına gelebilir. Görüldüğü gibi aynı madde iki farklı şehirdeki insanlarca koşulları doğrultusunda cevaplandırılacaktır. Yani sabahları işe gitmenin zorluğu farklı nedenler çerçevesinde ve anket maddesinin kendi özelliğiymiş gibi algılanıp farklı yorumlanabilir.

Ebeveyn rolleri ve bebek bakımına ilişkin hazırlanmış bir ölçekte kullanılan bir madde şu biçimde olsun:

“Bebeğe yapılan baş ve boyun masajının anne bebek ilişkisini kurmanın en iyi yolu olduğunu düşünürüm” (Örnek 2)

Genel bir bakım önerisinin yer aldığı bu madde büyük bir olasılıkla Endonezya ülkesinde düşük oranda çıkacaktır. Çünkü bu ülkede “Küçük çocukların başını okşamayın, yoksa zekaları gelişmez.” gibi bir inanışın yaygın olarak kabul görmesi; bu ülkeden ankete katılan bireylerin yanıtlarına yansıtacaktır. Verilen bu örnekler; ölçme eşdeğerliğinin (Measurement Invariance: MI) temel konusudur. Ölçme eşdeğerliği ölçülen bireylerin karakteristikleri ile değil ölçme aracının direk kendisiyle ilgilidir.

Ölçme eşdeğerliği için temel prensip; gruplar arasında ölçekte ölçülebilen maddelerle belli özelliklerin ilişkisinin aynı olup olmadığını tespit etmektir. Gözlenen değişkenin durumsal dağılımı,

değerlendirilen örtük değişkenin verilen değeri alt örneklerde ya da gruplarda eşitse o zaman ölçme eşdeğerliğinden bahsedilir.

Bir ölçme modelinin birden fazla grupta aynı yapıya sahip olması demek; söz konusu ölçeğin maddelerinin faktör yüklerinin, faktörler arası korelasyonların ve hata varyanslarının aynı olması demektir (Bollen, 1989; Byrne, 1998; Jöreskog & Sörbom, 1993). Farklı gruplarda karşılaştırılacak modelin bu gruplar için geçerli olması gerekir. Yapılan karşılaştırmalar sonucu tüm parametrelerin eşitliği söz konusuysa modelin gruplar için değişmez olduğu söylenebilir. Karşılaştırmaların yapılacağı alt gruplarda ölçmelerin eşdeğerliği sağlandığında alınan testin ya da teste ait herhangi bir maddenin özel bir tipi oluşturulmaz. Ölçme eşdeğerliğine yönelik çalışmalar; ölçme aracının yapılan ölçümlerde eşit yapıyı ortaya koyup koymadığını garantilemeye çalışır.

Ölçme eşdeğerliği; gruplar arası karşılaştırmalarda bu karşılaştırmaların anlamlılığında bir koşul olarak ortaya çıkar (Bollen, 1989; aktaran: Cheung & Rensvold, 1998). Oysa araştırmalara bakıldığında grupların karşılaştırılmasında daha doğru ve güvenilir sonuçlar vermesine rağmen hala eski geleneksel yöntemlerle grup karşılaştırmalarının ağırlıkta olduğunu görülmektedir. Bunun olası nedenleri aşağıdaki biçimde açıklanmaya çalışılmıştır (Steenkamp, E.M. ve diğerleri; 1998):

1. Ölçme eşdeğerliği çeşitleri ve terminolojisi hakkında görüş birliğinin sağlanmaması,
2. Araştırmacıların örtük ve gözlenen değişkenlere yönelik ölçme modellerine ve bunların test edilmesine olan uzaklığı,
3. Ölçme eşdeğerliği formlarının test edilmesindeki metodolojik karışıklıklar ve incelenecek örtük yapıların bu biçimde örneklerinin az olması ya da olmaması,
4. Araştırmacıların kültürler arası karşılaştırmaların anlamlılığı için hangi ölçümlerin eşit olması gerektiği konusundaki bilgi eksikliği,
5. Kültürler arası eşdeğerlik için ölçütleri gösteren açık bir rehberin bulunmaması.

Özellikle kültürler arası karşılaştırma yaparken kullanılan ve bir ülkede geliştirilmiş bir ölçeğin faktör yapısının başka bir ülkede aynı olup olmadığına yönelik çalışmalar mevcuttur (Şimşek, 2007:20). Bu çalışmalarda kültür, moderatör denilen farklılaştırıcı değişkendir.

Ölçme eşdeğerliğini ile ilgili olarak farklı jenerasyonların zeka düzeyleri arasındaki fark, cinsiyetler arası etkililik farkı, kültürler arası kişilik farklarını konu alan çalışmalar ön plana çıkmış ve araştırılmıştır. Bu tarz karşılaştırmalarda belli kesinlik dereceleri ortaya konmadıkça yapılan karşılaştırmaların sonuç getirmesi ve anlamlı olabilmesi mümkün olamayacaktır. Bu bakımdan “ölçmelerin eşdeğerliği” gibi test edilebilir sonuçlarla bu tarz tartışmaların çerçevesini çizen çalışmalar büyük bir öneme sahiptir.

Farklı grupların (zenci-beyaz) farklı ham puanlarından yola çıkarak (IQ puanları) ortaya çıkarılması düşünülen genel teorik davranışa ait (zeka) grup farklılıkları, eşdeğerlik sınırlamaları ile ortaya konan modellerle test edilmedikçe ortaya koymaz. Yani davranış; farklı gruplarda aynı şekilde ve aynı gözlem seti ile değerlendirilmelidir. İstatistiksel olarak karşılaştırılan iki grupta gözlenen değişkenler ile ilişkilendiren örtük değişkenlere ait matematiksel bağıntı eşit olmak durumundadır.

Ölçme Eşdeğerliği Çalışmaları Nasıl Yapılır?

Ölçme eşdeğerliğinin teorik tanımı oldukça geneldir ve farklı modellere uygulanabilir. Eşdeğerlik çalışmalarında her türlü veri ile uygulama yapma olanağı mevcuttur. Deneysel yöntemler kullanılarak ölçmelerdeki eşdeğerliğin ölçülmesi söz konusu değildir.

Son on yıldır ölçmelerin eşdeğerliği üzerinde yapılan çalışmalarda diferansiyel madde fonksiyonunu (DIF) değerlendirmek üzere birçok istatistiksel teknik ortaya konmuştur. Bu teknikle genellikle farklı kabuller yer alır. Bu kabullere örnek olarak; örtük değişkenler ve gözlenen değişkenler arasındaki ilişkinin biçimini göz ardı etmek, örtük değişken üzerindeki grup dağılımının

şeklini göz ardı etmek verilebilir. Bu nedenle belli bir teknik seçerek değerlendirmeye gitmek oldukça zordur. Çünkü seçilen tekniğin ölçmelerin eşdeğerliği teşhisindeki farklılığın ne kapsam ve büyüklükte olduğu her zaman açık olmamaktadır (Borsboom, 2006).

Eşdeğerlik testleri bir tür kovaryans yapı analizi olup farklı gruplarda belli bir yapının ölçülmesi üzerine tasarlanmıştır. İstatistiksel olarak ölçmelerin eşdeğerliği araştırılırken en çok ve yaygın kullanılan metot MG-CFA (Multi-Group Confirmatory Faktör Analizi)dir (Jöreskog & Sörbom, 1999). Çünkü bu metot olabilirlik oran testini kullanarak formal hipoteze dayanarak MI için karar verir (Byrne, Shavelson, & Muthén, 1989; Jöreskog & Sörbom, 1999; Zumbo, Sireci, & Hambleton, 2003). MG-CFA en az sınırlı modelden en sınırlı modellemeler arası karşılaştırma yaparak gruplar arası parametrelerin eşdeğerliğini sağlamaya çalışır (Horn & McArdle, 1992).

Birden fazla grubun bulunduğu grup karşılaştırmalarında çok örneklemli doğrulayıcı faktör analizi, grup parametrelerinin eşit kalarak, faktör ortalamalarının karşılaştırılmasını sağlamaktadır. Yani bu yöntem ölçme aracının gruplar arası eşdeğerliğinin sağlanmasında kullanışlılık sağlamaktadır.

Yapısal eşitlik modeline bağlı olarak yapılan grup karşılaştırması gözlenen değişkene dair yapılan ölçmelerden ziyade örtük değişkenlerin içinde bulunduğu hipotezleri test etmeye olanak verecektir. Örtük ortalama analizi geleneksel istatistik tekniklerine göre daha duyarlı bir tekniktir ve iki grup arasındaki farklılığı daha kesin ortaya koyar (Sehee, Mary, Malik, 2003). Aynı zamanda ortalamalar arasındaki farklılığı değerlendirirken varyans analizindeki karıştırıcı kaynakların ön değerlendirmesini de yapmış olur. Örtük değişkenlerle yapılan yol (path) analizi çalışmalarında hatayı elimine edebildiği için değişkenler arası ilişki miktarları da daha doğru biçimde hesaplanır.

Çoklu grup uygulamaları ile Yapısal Eşitlik Modeli (YEM) kullanılarak kurulmuş bir modeldeki örtük değişkenlerin gruplar arasında farklılık yaratıp yaratmadığı sorusuna yanıt aranmaktadır. Temel olarak yapısal eşitlik modeli uygulamalarında elde edilen eşitlikte yer alan değerler şunlardır:

1. Faktörlerin sayısı ve yük değerleri
2. Regresyon katsayısı,
3. Regresyon yorum değeri,
4. Artık regresyon varyansı,
5. Ortak faktörlerin ortalamaları,
6. Ortak faktörlerin varyansları
7. Ortak faktörlerin kovaryansları.

İlk dördü ölçme modeli ile alakalıdır ve bunlar gözlenenlerin örtük değişkenlere nasıl bağlanacağı hakkında bilgi verir. Diğer üçü ise yapısal model ile ilgilidir. Bunlar ise örtük değişkenlerin dağılımı ve ilişkileri hakkında bilgi verir. Araştırma literatüründe bu son üç elemanın ölçmelerin eşdeğerliği için gerekmediği konusunda bir anlaşma söz konusudur. Bunun nedeni ise ölçmelerin eşdeğerliğinin maddeler ya da faktörler arasındaki ilişkilerle ilgili olmamasından kaynaklanmaktadır (Little, 1999; Meredith & Millsap, 1992; Millsap, 1998; Widaman & Reise, 1997).

Eşdeğerlik Aşamaları

Ölçmelerin eşdeğerliği aşamalılık içerir. Test edilmesi gereken dört aşama vardır (Meredith, 1993). Bunlar;

- i) Yapısal eşdeğerlik (Configural invariance)
- ii) Metrik eşdeğerlik (Weak factorial invariance= Metric invariance)
- iii) Sayıl eşdeğerlik (Strong invariance= Scalar invariance)
- iv) Tam eşdeğerlik (Strict invariance).

Eşdeğerlik çalışmaları; geçerli değerlendirme ve bu çıktıların ölçümleri açısından kritik bir değere sahiptir. Aşamalılık içeren eşdeğerlik biçimleri aşağıda kısaca anlatılmıştır.

Yapısal eşdeğerlik: Ölçmelerin eşdeğerliğindeki en temel düzeydir (McArdle, Mason; 1983; aktaran: Chen, Sausa, West; 2005). Temel model yapısı gruplar için değişmezse biçimsel eşdeğerlikten söz edilebilir. Bu model kestirilen parametreler üzerinde gruplar arası eşdeğerlik sınırlandırması yapmaz. Yani grupların farklı parametre değerleri almasına izin verir. Böylece, model eşdeğerlik hiyerarşisinde gerekli olan tüm modellerle temel bir karşılaştırma sağlar. Biçimsel eşdeğerlik modeli kritik bir önem sahiptir; çünkü veri gruplar etrafında sabit ve sabit olmayan parametre çiftlerinin benzerliğini desteklemese o zaman bu veri daha sınırlayıcı modelleri de desteklemeyecektir (Bollen, 1989).

Örnek 1’de açıklanmaya çalışılan biçimsel eşdeğerliğin sağlanmadığı bir durumdur.

$\Lambda_{\text{FORM}(\text{İstanbul})} \neq \Lambda_{\text{FORM}(\text{Ağrı})}$ biçiminde ifade edilebilir.

Metrik eşdeğerlik: (Weak factorial invariance) Farklı grupların maddelere aynı biçimde cevap verdiğini ve böylece farklı gruplardan elde edilen puanların karşılaştırma da anlamlı olduğu sayıtlıları metrik eşdeğerlikte karşılanmalıdır (Steenkamp&Baumgartner, 1998). Sınırlayıcı bir modeldir. Metrik eşdeğerliğin testi, faktör uyum katsayılarının gruplar arası eşit olmasına zorlanmasıyla sağlanabilir; çünkü uyum katsayıları örtük puanlarla gözlenen puanlar arasındaki ilişkiye dair bilgi verir. Sadece bu durumda maddelere ait puanların gruplar arasında anlamlı biçimde karşılaştırılması yapılabilir. Bu; “maddeye ait gözlenen puanlar arasındaki farklılık; örtük yapı altındaki gruplar arası farklılığı gösterir” demektir. Örnek 2 bu eşdeğerlik türüne girer, şu biçimde ifade edilir.

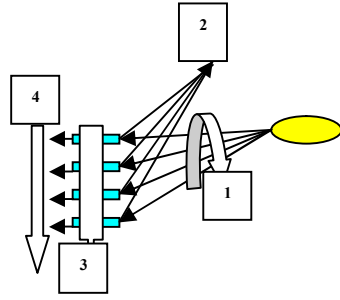
$\lambda_{ij} \neq \lambda_{ij}$

λ_{ij} : j örtük yapısı üzerindeki i maddesinin yükü

Skalar eşdeğerlik(Strong factorial invariance): Aynı değere sahip konuların örtük yapı üzerindeki değerinin gözlenen değer üzerindeki değerine eşit olduğunu ifade eder. Bu “gözlenen maddeler üzerindeki grup farklılıkları örtük yapıların ortalamalarından kaynaklanır” demektir. Gruplar arası genel ya da özel faktörlere verilen yanıtların durumsal anlamda beklenen yanıtlarının değişmez olduğunu vurgular. Özel faktör ortalamalarının da gruplar arasında benzer olması gerekir.

Sıcaklık ölçümünde kullanılan ölçek dönüşümü bunun için iyi bir örnek olabilir. Grup A’da sıcaklık ölçerken Celcius ölçeği, Grup B B’de ise Kelvin ölçeği kullanılsın. Her ikisinde de ölçmeye konu olan özellik aynıdır. Ancak ölçeklerde kullanılan başlangıç noktaları birbirinden farklıdır. Bu biçimi ile karşılaştırma yapılmaz. Ya Grup A’ da elde edilen ölçüme 273 eklenir, ya da Grup B de yapılan ölçümden 273 çıkarılır. Yapılan bu işlem ölçülen özelliğe yönelik farklı gruplarda elde edilen ölçümlerin karşılaştırılmasını olanaklı kılar (Chen, Sousa, West; 2005).

Tam eşdeğerlik (Strict factorial invariance): Faktörlere verilen durumsal yanıtların varyanslarının değişmez olmasını ister. Yani eş faktör yükleri, ve yorumlarının yanı sıra madde artıklarının varyanslarının (özel faktör + hata değişkeni) eş olmasını gerektirir. Gözlenen değişken ortalamalarında ki ve dağılım matrisindeki grup farklılıkları hesaplanır.



1. Biçimsel Eşdeğerlik
2. Faktöryel Eşdeğerlik
3. Skalar Eşdeğerlik
4. Tam Eşdeğerlik

Bu eşdeğerlik formları hiyerarşik bir sıra takip edilerek ve bir çok model uyum katsayısı kullanılarak test edilir. Meredith (1993) adil ve geçerli bir karşılaştırma için tam eşdeğerliğin gerekli bir aşama olduğunu özellikle vurgulamıştır. Ancak pratik uygulamalarda tam ölçme eşdeğerliği genelde kullanılmaz.

Tartışma

Ham ortalama örtük ortalamalar için aynı sonucu verdiğinde bile gerekli eşdeğerlik testlerinin yapılması gerekir. Çünkü ancak o zaman gruplar arası karşılaştırma anlam kazanır. Çeşitli özel formlara bağlı modeller oluştururken ve parametrelere dair uygun belirlemeler yaparken bunların gruplar arası farklılıklara sahip olacağı unutulmamalıdır. Hatalı karşılaştırmalar ölçmelerin eşdeğerliğinin sağlanamamasından kaynaklanmaktadır.

Gruplar arasındaki farklılıklar değerlendirilirken geleneksel yaklaşımlar kullanmak hatalı sonuçlar doğurabilir. Bu ise ortaya konmak istenen davranışın, özelliğin yanlış tanımlanmasına neden olabilir. Bu nedenle yapılarca dair gruplar arası karşılaştırmalar yaparken eşdeğerlik aşamalarının desteklenmesi gerekmektedir.

Çalışmaların eşdeğerlik testleri ile zenginleştirilmesi önemli bilgiler sağlayabilir. Bütününe yönelik anlamlı genellemeler yapılabilmesi için eşdeğerlik çalışması ile farklı faktörler bakımından gruplar arası karşılaştırmalar yapılması; özelliklerin grup bakımından açıklanmasında yararlı olacaktır.

Kaynaklar

- Cheung, G. W., & Rensvold, R B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
- D. WuZhen LiBruno D. Zumbo (2007) Decoding the Meaning of Factorial Invariance and Updating the Practice of Multi-group Confirmatory Factor Analysis: A Demonstration With TIMSS Data 12(3),
- Fang Fang Chen, Karen H. Sousa, and Stephen G. West. (2005). Testing Measurement Invariance of Second-Order Factor Models; *Structural Equation Modeling*, 12(3), 471-492
- George A. Marcoulides, Christin Emrich, Laura D. Marcoulides (2007) Educational And Psychological Measurement Testing For Multigroup Invariance Of The Computer Anxiety Scale
- Hoyle, R.H., (1995) *Structural Equation Modeling: Concept, Issues And Application*
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32(1), 53-76.
- M. Wicherts, Conor V. Dolana, David J. Hessena, Paul Oosterveld, G. Caroline M. van Baalb, Dorret I. Boomsma, Mark M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect *Journal of Personality and Social Psychology*, 86, 509-537

- Meredith, W. (1993), Measurement Invariance, Factor Analysis, And Factorial Invariance. *Psychometrika*, 58, 525-543.
- Öğretmen,T. (2006); Uluslar Arası Okuma Becerilerinde Gelişim Projesi (Pirls) 2001 Testinin Psikometrik Özelliklerinin İncelenmesi: Türkiye- Amerika Birleşik Devletleri Örneği; Yayınlanmış Doktora Tezi: Hacettepe Üniversitesi
- Roger B. Rensvold and Gordon W. (1998) ;Cheung Testing Measurement Models for Factorial Invariance: A Systematic Approach *Educational and Psychological Measurement*; 58; 1017-1018
- Sehee Hong And Mary L. Malik (2003); Testing Configural, Metric, Scalar, And Latent Mean Invariance Across Genders In Sociotropy And Autonomy Using A Non-Western Sample *Educational And Psychological Measurement* ; 63; 636
- Steenkamp E.M.&Baumgartner H.(1998); Assesing Measurement Invariance İn Cross-National Consumer Research