

## Farklı Yöntemlerle Ölçme Değişmezliğinin İncelenmesi: Pisa 2012 Örneği\*

### Examining Measurement Invariance with Different Methods: Example of Pisa 2012

Alperen YANDI\*\*, İbrahim Alper KÖSE\*\*\*, Ömür UYSAL\*\*\*\*

**Öz:** Bu çalışmada istatistiksel varsayımlar açısından farklı olan yöntemler kullanılarak aynı veri setinin ölçme değişmezliği ile ilgili sonuçların incelenmesi amaçlanmıştır. Ayrıca normallik varsayımını gerektiren ve gerektirmeyen yöntemlerin normallik varsayımı sağlanamayan durumlarda farklı sonuçlar gösterip göstermediğine bakılmıştır. Bu amaca göre PISA 2012 alt ölçeklerinden beş maddeden oluşan Problem Çözmeye Açıklık ölçeği Türkiye ve Finlandiya örneklemi veri seti üzerinde yapısal eşitlik modellemesi çatısı altındaki ortalama kovaryans yapılarının değişmezliği analizi ve örtük sınıf analizi çatısı altındaki çoklu grup örtük sınıf analizi yöntemi analizler gerçekleştirilmiştir. Ortalama kovaryans yapılarının değişmezliği analizi için Lisrel 8.72; çoklu grup örtük sınıf analizi için ise Latent Gold 5.1 programları kullanılmıştır. Ortalama kovaryans yapılarının değişmezliği analizi için yapısal değişmezlik ile başlayan ve katı değişmezlik aşaması ile biten aşamalı test etme yöntemi adımları takip edilmiştir. Örtük sınıf analizi için ise örtük sınıf sayısının belirlenmesinden sonra heterojen, kısmi homojen ve homojen model test edilmiştir. Analizlerden önce Kolmogorov Smirnov testi kullanılarak her bir maddenin normalliği incelenmiştir. Alt örneklem ve tüm grup için hiç bir madde normal dağılım göstermemiştir. Normallik varsayımı gerektiren ortalama kovaryans yapılarının değişmezliği analizi sonuçlarına göre katı değişmezlik kabul edilmiştir. Bir başka deyişle ölçme değişmezliği sağlanmıştır. Fakat normallik varsayımı gerektirmeyen çok gruplu örtük sınıf analizi için kısmi homojen model kabul edilmiştir. Kısmi homojen model ortalama kovaryans yapılarının analizi adımlarından zayıf değişmezliğe karşılık gelmektedir. Elde edilen sonuçlara göre varsayımlar açısından farklılaşan metodlar kullanıldığında ölçme değişmezliği bulguları değişiklik göstermiştir. Bu bağlamda yöntemler için gerekli varsayımlar mutlaka incelenmeli ve gizli ve gözlenen değişken yapıları göz önünde bulundurularak uygun yöntem seçmeye dikkat edilmelidir.

**Anahtar Kelimeler:** Ölçme değişmezliği, ortalama kovaryans yapılarının değişmezliği yöntemi, çoklu grup örtük sınıf analizi, normallik varsayımı

**Abstract** In this study, it is aimed to examine results about measurement invariance of the same data set by using different methods in the way of statistical assumptions. Also the methods requiring the normality assumption and the methods not requiring this assumption were examined whether they show different results or not when the normality assumption cannot be supplied. In accordance with this aim, analyses were carried out by using mean and covariance structures invariance analysis under roof of the structural equation modeling and multi-group latent class analysis under roof of the latent class analysis on the data set Openness for Problem Solving Scale consisting of five items, subscale of PISA 2012 of Turkey and Finland samples. For mean and covariance structures method, Lisrel 8.72 program and for the multi-group latent class method Latent Gold 5.1 program were used. Steps of stepwise testing method starting with configural invariance and ending with testing strict invariance step were followed for mean and covariance structures invariance method. After determining number of class, heterogeneous, partial homogenous and homogenous models were tested for multi-group latent class analysis. Before the analyses, normality of each item was examined by using Kolmogorov Smirnov test. None of the items demonstrated normal distribution for all subsamples and the whole group. According to the results of mean and covariance structures invariance analysis requiring normality assumption, strict invariance was accepted. In other words measurement invariance was supplied. But for multi-group latent class analysis not requiring normality assumption, partial homogenous model was accepted. Partial homogeneous model is equal to weak invariance in steps of mean and covariance structures invariance analysis. According to the results, findings of measurement invariance changed when methods differing from the point of assumptions were used. In this context, certainly the assumptions considered necessary for methods

\*Bu çalışma 5. Eğitimde Ölçme ve Değerlendirme Kongresi'nde sözel bildiri olarak sunulmuştur.

\*\*Araş. Gör., Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, Bolu-Türkiye, e-posta: alperenyandi@gmail.com

\*\*\*Doç. Dr., Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, Bolu-Türkiye, e-posta: i.alper.kose@gmail.com

\*\*\*\*Öğretmen, MEB, Muğla-Türkiye, e-posta: omuruysal@gmail.com

must be examined and choosing the suitable method must be considered by bearing in mind the structures of latent and observed variables.

*Keywords:* Measurement invariance, mean and covariance structures invariance method, multi-group latent class analysis, normality assumption

## Giriş

Sosyal bilimler alanındaki yapılan araştırmaların odak noktasında bireylerin olduğu görülmektedir. Yapılan araştırmaların birçoğunda bireyler arası ve gruplar arası farkların belirlenmesinin amaçlandığı görülmektedir. Bireylerden elde edilen veriler analiz edilmekte ve analiz sonuçlarına dayalı olarak çeşitli çıkarımlar yapılmaktadır. Bu çıkarımlar doğrultusunda önemli kararlar verilmektedir.

Eğitim yöneticileri gerek geniş gerek küçük ölçekli uygulamalarda elde edilen sonuçları göz önüne alarak okul bazında veya ülke bazında düzenlemelere ve değişikliklere gidilmesini kararlaştırmaktadır. Yapılan bu düzenlemelerle eğitim öğretim sürecinin daha verimli hale getirilmesi var olan eksikliklerin giderilerek yola devam edilmesi sağlanmaya çalışılmaktadır. Bu bağlamda araştırmalarda elde edilen verilerin gerçekten ölçülmek istenen özellikle ilgili olması durumunun gerekliliği ve önemi ortaya çıkmaktadır.

Sosyal bilimler alanında araştırmacılar bireylere ilişkin psikolojik özellikler üzerinde durmaktadır. Bir veya daha çok psikolojik özellikle ilgili ölçme araçları kullanılarak veriler toplanmaktadır. Toplanan veriler üzerinde ise ilişki analizleri, gruplar arası farkların farklı yöntemlerle belirlenmesi, modellemeler yapılmaktadır. Bu noktada elde edilen sonuçların doğruluğuna ilişkin olarak psikolojik ölçme araçlarının sağlanması gereken psikometrik özellikler karşımıza çıkmaktadır. Psikolojik ölçme aracı, bireyleri yalnızca ölçülmek istenen özellik bakımından uyaran ve ölçülen bireyde sadece söz konusu özelliğin bulunmuşluk düzeyini yansıtacak tepki düzeylerinin yer aldığı maddeler takımı olarak tanımlanabilir (Erkuş, 2012). Çalışmalarda kullanılan psikolojik ölçme araçları belli aşamalar çerçevesinde gerçekleştirilen uyarılma ve geliştirme çalışmaları ile elde edilmektedir. Bir ölçme aracında aranan en önemli psikometrik özellik geçerliktir. Messick (1994) geçerlilik kavramını uygulamalar sonucu elde edilen puanlara dayalı olarak yapılan yorumların, bu yorumlar doğrultusunda ortaya çıkan sonuçların, deneysel ve kuramsal boyutlarla desteklenmesi şeklinde ele almaktadır. Ölçme araçlarının geçerliğine ilişkin doğru kanıtların toplanması gerekmektedir. Bu nedenle farklı grup ve bireyleri karşılaştırmayı amaçlayan çalışmalarda kullanılan ölçme araçlarının doğrudan ölçülmek istenen özelliğe ilişkin ölçümleri ortaya koymalıdır. Bu doğrultuda ölçme araçlarının geçerliğine dair doğru kanıtların toplanabilmesi, elde edilen sonuçlar doğrultusunda atılacak adımların uygun olması içi ölçme değişmezliği çalışmaları mutlaka yapılmalıdır.

Ulusal alanyazın incelendiğinde çeşitli değişkenler açısından farklı gruplara üye bireylerin karşılaştırıldığı çalışmalar sayıca fazla olsa da, kullanılan ölçme araçlarının ölçme değişmezliğine dair kanıtların ortaya konmadığı dikkat çekmektedir. Ölçme araçlarının geliştirilme sürecinde “araç uygulandığı her grupta aynı özelliği ölçer” (Başusta ve Gelbal, 2015) varsayımı ile hazırlandığı düşünüldüğünde ölçme değişmezliğinin sınanmadığı çalışmaların sonuçların birçok soru işareti taşıdığı düşünülebilir.

Ölçme değişmezliği ölçülmek istenen özelliğin bireyin başka farklı bir özelliğinden etkilenmeden ortaya çıkarılması anlamına gelmektedir. Ölçme değişmezliğine ilişkin olarak alanyazında farklı tanımlamalar bulunmaktadır. Bryne ve Watkins (2003) ölçme değişmezliğini belli bir değişken bakımından farklı alt gruplarda bulunan bireylerin ölçme aracındaki maddeleri aynı şekilde algılama ve yorumlaması olarak tanımlamıştır. Öte yandan Flowers, Raju ve Oshima (2002) ise ölçülmek istenen psikolojik özellik bakımından aynı düzeyde fakat farklı alt gruplarda bulunan bireylerin aynı ölçme aracı ile yapılan uygulamalardan aynı gözlenen puanı alma durumları olarak ölçme değişmezliğini açıklamışlardır. Bir başka tanıma göre ise ölçme değişmezliği, farklı alt grupların katılımıyla gerçekleştirilen çalışmalarda, ölçme aracı ile ölçülmek istenen bir yapının ölçme parametrelerinin katılan alt gruplar arasında matematiksel olarak birbirine denk olmasıdır (Little, 1997).

Geçerlik, alınan kararların doğruluğu gibi kritik özellikler bakımında bu denli önem taşıyan bir özellik olan ölçme değişmezliğinin yapılan çalışmalarda ihmal ediliyor olmasına ilişkin olarak Steenkamp ve Baumgartner (1998) bazı sebepler ortaya koymuşlardır. Bunlar:

- 1- Ölçme değişmezliğinin aşamaları hakkında birçok farklı görüşün bulunması,
- 2- Ölçme değişmezliği aşamaları ile ilgili olarak bilim insanları arasında görüş birliğinin sağlanamamış olması,
- 3- Araştırmacıların örtük ve gözlenen değişkenlere yönelik ölçme modellerine ve bunların test edilmesine olan uzaklığı,
- 4- Ölçme değişmezliği aşamalarının test edilmesinde kullanılan yöntemler ile ilgili karmaşıklıklar ve bu aşamaların test edilmesi ile ilgili örneklerin az olması,
- 5- Kültürlerarası karşılaştırmaların anlamlı olabilmesi için hangi ölçümlerle ilgili eşitliğin sağlanması gerektiği ile ilgili bilgi eksikliği,
- 6- Kültürlerarası ölçme değişmezliği için ölçütleri gösteren bir rehberin olmamasıdır.

Belirtilen bu nedenler arasında da belirtildiği gibi araştırmacıların yöntemlere ilişkin eksiklikleri ölçme değişmezliğine ilişkin incelemelerin yapılmasını sekteye uğratmaktadır. Ölçme değişmezliğinin test edilmesine ilişkin olarak yapısal eşitlik modellemesi (YEM), madde tepki kuramı (MTK) ve örtük sınıf analizleri (ÖSA) çatısı altındaki yöntemler göze çarpmaktadır. Ölçme değişmezliği ile ilgili çalışmaların çoğunda YEM ve MTK çatısı altındaki yöntemler tercih edilirken, son yıllarda ÖSA altındaki çoklu grup örtük sınıf analizi (ÇGÖSA) yöntemi de kullanılmaya başlanmıştır.

YEM ve MTK altındaki yöntemler örtük ve gözlenen değişkenlerin sürekli yapıda olduğu durumlarda kullanılırken ÇGÖSA örtük ve gözlenen değişkenlerin kategorik olduğu durumlarda tercih edilmektedir. Aynı zamanda ÇGÖSA'nin değişkenlerin sürekli yapıda olduğu durumlarda da iyi bir alternatif olduğu belirtilmektedir (Kankaras, Vermunt ve Moors, 2011). Ayrıca YEM ve MTK yöntemleri ile ÇGÖSA arasında varsayımsal olarak da belli farklılıklar vardır. Kankaras ve Moors'a (2009) göre bu farklılıklar kısaca şu şekilde açıklanabilir:

- 1- ÇGÖSA, ikili ve çoklu puanlanan değişkenler için YEM altındaki yöntemlere göre daha iyi uyum göstermektedir. Ayrıca ÇGÖSA normallik, homojenlik ve doğrusal ilişki varsayımlarının sağlanmasını gerektirmemektedir.
- 2- ÇGÖSA'de grup karşılaştırmalarının yapılması için eş değer madde sayısı ve faktör yapısının önemli bir kriter değilken, YEM altındaki yöntemler için ise farklı alt gruplar için faktörlerdeki en az iki madde eş değer olmalıdır.
- 3- Öte yandan ÇGÖSA MTK ile karşılaştırıldığında ise çok boyutlu yapılarla daha kolay şekilde baş edebilmektedir (Güngör, Korkmaz ve Somer, 2013).

Vandenberg ve Lance (2000) ölçme değişmezliği çalışmalarının %80'nin YEM altındaki yöntemlerle yapıldığını belirtmiştir. YEM altındaki yöntemler Çoklu Grup Doğrulamalı Faktör Analiz (ÇGDFA) ve Ortalama Kovaryans Yapılarının Değişmezliği (OKYD) olarak söylenebilir. Bu yöntemler kullanılarak yapılan çalışmalarda varsayımların test edilmesi oldukça dikkat edilmesi gereken bir husustur. Değişken yapılarının sürekli olmasının gerekliliği ve normallik varsayımının sağlanması elde edilen sonuçların kesinliği açısından gereklidir. Ölçme değişmezliğinin elde edilen sonuçların geçerliği ve bu sonuçlara dayalı kararların doğruluğu için yapılmasının önemli olduğu kadar kullanılan ölçme değişmezliği yönteminin doğru seçilmesi ve bu yöntemin gerekliliklerinin yerine getirilmesi bu denli önemlidir. Bu bağlamda varsayımsal olarak farklılaşan yöntemlerin işleyiş süreci, bu yöntemlerle elde edilen sonuçlara dair araştırmacılara katkı sağlanabilir.

### ***Araştırmanın Amacı***

Bu çalışmada varsayımsal olarak birbirinden farklılaşan iki farklı yöntem olan ortalama kovaryans yapılarının değişmezliği ve çoklu grup örtük sınıf analiz yöntemleri kullanılarak ölçme değişmezliğinin test edilmesi amaçlanmıştır. Gerçekleştirilen analizler sonucunda elde edilen sonuçlar incelenerek bu iki yöntemle elde edilen ölçme değişmezliği sonuçları ele alınmıştır. Bu doğrultuda normallik varsayımı incelemeleri yapılarak bu varsayıma göre elde

edilen sonuçlarda farklılaşma olup olmadığı tartışılmıştır. Bu amaçlar doğrultusunda araştırma soruları şu şekilde sıralanabilir.

- 1- Ortalama kovaryans yapılarının değişmezliği yöntemi,
  - 2- Çoklu grup örtük sınıf analizi yöntemi,
- kullanılarak elde edilen ölçme değişmezliği sonuçları nelerdir?

### **Araştırmanın Önemi**

Bu çalışma ile birlikte farklı ölçme değişmezliği belirleme yöntemleri ile ilgili olarak araştırmacılara bilgi sağlanabilir. Böylece farklı özellikte veri setleri için araştırmacıların “hangi yöntemi kullanmalıyım?” sorusuna yanıt arama sürecin katkı sağlanabilir. Araştırmacıların ölçme değişmezliği belirleme yöntemlerine dair bilgilerine katkı sağlanarak ölçme değişmezliğinin test edilmesinin yaygınlığı artırılabilir.

Bu çalışma kapsamında varsayımsal olarak farklılaşan iki yöntemin kullanılması ile ölçme değişmezliğinin doğru şekilde test edilmesi gerekliliği vurgulanmaya çalışılmıştır. Bu durum özellikle varsayımların test edilmesinin ve bu doğrultuda uygun yöntemin seçilmesinin gerekliliğini göstermiştir. Bu bağlamda bu çalışma ölçme değişmezliğine ilişkin elde edilecek sonuçların doğruluk derecesine katkı sağlamak için önem arz etmektedir.

### **Yöntem**

Bu bölümde araştırma kapsamında kullanılan verilerin özelliklerine ilişkin olarak çalışma grubu, veri toplama aracı ve veri analizi sürecinden söz edilmiştir.

### **Çalışma Grubu**

Bu çalışmada PISA 2012 veri setinden kayıp ve geçersiz verilerin tümü silindikten sonra geriye kalan 3159 Türk ve 5676 Fin ST94 kodlu Problem Çözmeye Açıklık alt ölçeğine verdiği yanıtlar kullanılmıştır. Çalışma kapsamında PISA sınavlarının tümünde yüksek başarı elde etmiş olan Finlandiya ve öğretim programı açısından en yoğun programlardan birine sahip olan, PISA sınavlarında beklenen ve istenilen başarıyı elde edememiş Türkiye örneklemini seçilmiştir.

### **Veri Toplama Aracı**

PISA 2012 alt ölçeklerinden Problem Çözmeye Açıklık ölçeği beş maddeden oluşmaktadır. Bu ölçek hiç bana göre değil(1) - tam bana göre(5) arasında derecelenen şekilde 5’li likert tipi bir yapıdadır. Problem Çözmeye Açıklık ölçeği öğrencilerin çok miktarda bilgiyle başa çıkabilme, karşılaştıkları anlayabilme hızı ve açıklayabilme durumu, gerçekler arasında ilişkiler kurabilme düzeylerini ve karmaşık problemlerden hoşlanma seviyelerini ölçmek için kullanılmıştır. Problem Çözmeye Açıklık ölçeği için Finlandiya ve Türkiye örneklemini için gerçekleştirilen doğrulayıcı faktör analizi (DFA) sonuçları Tablo 1’de verilmiştir.

**Tablo 1. Problem çözmeye açıklık ölçeği DFA sonuçları**

Örneklem	Madde	t değerleri	Uyum indeksi	Değer	Karar
TÜRKİYE	m1	37,090*	Ki-Kare/sd	16,077	Kabul edilebilir
	m2	44,400*	RMSEA	0,069	Kabul edilebilir
	m3	38,870*	GFI	0,99	Mükemmel
	m4	35,400*	CFI	0,99	Mükemmel
	m5	27,070*	NNFI	0,98	Mükemmel
FİNLANDIYA	m1	53,910*	Ki-Kare/sd	16,650	Kabul edilebilir
	m2	58,540*	RMSEA	0,053	İyi
	m3	59,950*	GFI	0,990	Mükemmel
	m4	70,970*	CFI	1,000	Mükemmel
	m5	54,180*	NNFI	0,990	Mükemmel

\*p<0,01

Tablo 1 incelenerek Türkiye ve Finlandiya örneklemelerine ait verilerle gerçekleştirilen analiz sonuçlarına Problem Çözmeye Açıklık ölçeğinin üç grup için iyi uyum gösterdiği yorumu yapılabilir. Ayrıca ölçekte bulunan maddelere ait t değerlerine bakıldığında, tüm t değerlerinin 2,58'den daha yüksek olduğu görülmektedir. T değerleri, gizil değişkenlerin gözlenen değişkenleri açıklama durumlarına ilişkin bilgi verir. Eğer t değeri 1,96'nın üzerindeyse 0,05 düzeyinde, 2,58'in üzerindeyse 0,01 düzeyinde manidardır. 1,96'dan daha düşük bir değerse manidar değildir (Şimşek, 2007). Buna göre tüm maddelere ilişkin değerler 0,01 düzeyinde anlamlıdır.

Problem Çözmeye Açıklık ölçeğine ilişkin olarak gerçekleştirilen güvenilirlik analizlerinde elde edilen Cronbach  $\alpha$  katsayıları Türkiye örnekleme için 0,785; Finlandiya örnekleme için 0,854; Şangay-Çin örnekleme için ise 0,835 olarak hesaplanmıştır. Bu sonuçlara ölçeklerin güvenilirlik düzeylerinin oldukça yüksek düzeyde olduğu söylenebilir.

### **Verilerin Analizi**

Problem Çözmeye Açıklık ölçeği ile elde edilen veriler üzerinde Ortalama Kovaryans Yapılarının Değişmezliği ve Çoklu Grup Örtük Sınıf Analizi yöntemleri kullanılarak ikili şekilde ülke örneklemeleri verileri üzerinde ölçme değişmezliği analizleri gerçekleştirilmiştir. Ölçme değişmezliği analizlerine geçilmeden önce Kolmogorov Smirnov testi sonuçları incelenerek maddelerin normal dağılım gösterip gösterilmediği alt gruplar ve tüm grup bazında incelenmiştir. Normallik varsayımı için manidarlık düzeyi 0,01 olarak alınmıştır. Bu doğrultuda normallik testi sonuçlarına göre bu varsayımı gerektiren ve gerektirmeyen yöntemlerin kullanılmasının elde edilen sonuçlar üzerinde bir farklılık ortaya çıkarıp çıkarmadığına dair incelemeler yapılmıştır.

Ölçme değişmezliği test etme sürecine ilişkin olarak yapılan sınıflamada geriye doğru aşamalı ve aşamalı test etme süreçlerinin karşımıza çıkmaktadır. Bunlardan ilkinde ölçme değişmezliği testleri, en az sınırlandırılmış modelden en fazla sınırlandırılmış modele doğru (aşamalı test etme); ikincisinde ise en fazla sınırlandırılmış modelden en az sınırlandırılmış modele doğru (geriye doğru aşamalı test etme) gerçekleştirilmektedir. Geriye doğru aşamalı test etme yöntemi ile ilgili olarak bazı dezavantajlar olduğu belirtilmektedir (Önen, 2009). Bu dezavantajlardan en önemlisi en sınırlandırılmış modelden başladığında ölçme değişmezliğinin sağlanmadığı bulgusuna ulaşıldığında, bu durumun hangi parametreden kaynakladığı oldukça güç olmasıdır. Bu nedenle OKYD yöntemi ile yapılan analiz sürecinde Meredith (1993) tarafından önerilen aşamalı test yöntemine uygun olarak analizler ilerletilmiştir. Aşamalı test etme yönteminden en az sınırlandırılmış olan ve yalnızca gruplar arası faktör yapılarının eş değerliliğinin incelendiği heterojen model test edilerek analize başlanır. İkinci basamakta ilk madde faktör yüklerinin, üçüncü basamakta regresyon sabitlerinin ve son basamakta ise hata varyanslarının eş değerliliğinin incelendiği sırasıyla zayıf değişmezlik, güçlü değişmezlik modeli ve katı değişmezlik basamakları takip edilerek analiz yürütülür. Her bir adımda bir önceki adımlardaki sınırlamalara ek olarak yeni sınırlamalar getirilerek ilerlenir. Bu çalışmada bu basamaklara uygun şekilde analizler gerçekleştirilmiştir.

Analiz sürecinde aşamaların karşılaştırılmasında birbirini takip eden aşamalar için elde edilen Ki-kare değerleri farklarının ( $\Delta\chi^2$ ) ardışık iki aşamanın serbestlik dereceleri farkındaki 0,05 düzeyindeki anlamlılığına, karşılaştırmalı ve iyilik uyum indeksi farklarına ( $\Delta$ GFI ve  $\Delta$ CFI) bakılabileceği belirtilmektedir (Vandenberg ve Lance, 2000; Cheung ve Rensvold, 2002; Wu, Li ve Zumbo, 2007). Bu çalışmada  $\Delta\chi^2$  değeri örneklem büyüklüğünden etkilendiği için OKYD bulguları karşılaştırmalı ve iyilik uyum indeksleri üzerinden yorumlanmıştır. Ölçme değişmezliğinin sağlandığının rapor edilmesi için karşılaştırmalı ve iyilik uyum indeksi farklarının ise -0,01 ile 0,01 arasında değişmesi gerekmektedir (Cheung ve Rensvold, 2002; Wu, Li ve Zumbo, 2007). Bu doğrultuda OKYD ile gerçekleştirilen ölçme değişmezliğine ilişkin sonuçlar rapor edilmiştir.

ÇGÖSA süresiz veriler üzerindeki kategorik yapıdaki örtük değişkenlerin tanımlanması amaç edinen örtük sınıf analizi yöntemlerinden biridir. (Samuelsen ve Dayton, 2010; Akt: Güngör, Korkmaz ve Somer, 2013). Örtük sınıf analizi sürecinde veri seti üzerindeki

örtük sınıf sayısının belirlenmesi ile analize başlanır. Örtük sınıf sayısının belirlenmesi için Pearson Ki-Kare ( $\chi^2$ ), ençok olabilirlik ki-kare oranı (likelihood ratio chi square) (L2) Bayesian Bilg kriteri ve türevleri (BIC ve DBIC) ve Akaike bilgi kriteri ve türevleri (AIC, AIC3, CAIC) kullanılabilir (Silvia, Kaufman ve Pretz, 2009; Akt: Güngör, Korkmaz ve Sazak, 2015). Bu bilgi kriterlerinden küçük örneklerde AIC'nin, büyük örneklerde ise BIC'nin AIC'den daha iyi sonuçlar verdiği belirtilmektedir (Kankaras, Vermunt ve Moors, 2011). Bu bilgi kriterleri doğrultusunda en küçük değerin elde edildiği değer örtük sınıf sayısı olarak belirlenir (Moors ve Wennekers, 2003). Bu çalışmada örtük sınıf sayısı ve en iyi modelin belirlenmesinde BIC kriterinden yararlanılmıştır. Ayrıca çoklu grup örtük sınıf analizi sürecinde uygun modelin belirlenmesine ek olarak her bir madde için örtük sınıf olasılıkları belirlenerek grupta bulunan bireylerin maddelerdeki tepki düzeylerine ilişkin eğilimleri kestirilebilir

Örtük sınıf sayısının belirlenmesinin ardından uygun örtük sınıf sayısının belirlendiği model üzerinden heterojen model, kısmi homojen model ve homojen model test edilerek ölçme değişmezliği analizleri gerçekleştirilmektedir. Heterojen modelde eğim ve sabit parametreleri serbest bırakılırken, kısmi homojen modelde eğim parametreleri alt gruplar için eşitlenmektedir. Homojen modelde ise eğim ve sabit parametrelerinin her ikisi alt gruplar için eşitlenerek analiz yapılmaktadır (Güngör, Korkmaz ve Somer, 2013).

Bu çalışmada OKYD yöntemi için analizlerde LISREL 8.72, ÇGÖSA için ise Latent Gold 5.1 programları ve normallik testleri için SPSS 20.0 paket programı kullanılmıştır.

## Bulgular

### *Maddelerin Normalliklerine İlişkin Sonuçlar*

Çalışma kapsamında kullanılan veriler üzerinde yapılan Kolmogorov Smirnov testi sonuçları Tablo 2'de sunulmuştur.

**Tablo 2.** Kolmogorov Smirnov testi sonuçları

Madde	Türkiye	Finlandiya	Tüm Grup
m1	0,00	0,00	0,00
m2	0,00	0,00	0,00
m3	0,00	0,00	0,00
m4	0,00	0,00	0,00
m5	0,00	0,00	0,00

Tablo 2 incelendiği ölçekte yer alan 5 maddenin alt gruplar ve tüm örneklem için normal dağılım göstermediği görülmektedir ( $p_{m1-m5} < 0,05$ ). Bu sonuç doğrultusunda OKYD analizinin ölçme değişmezliğinin test edilmesi için uygun olmadığı söylenebilir. Ancak bu çalışmada normal dağılım varsayımına ilişkin incelemelerde yapılmak istendiğinden analizlere devam edilmiştir.

### *Ortalama Kovaryans Değişmezliği Yöntemine İlişkin Sonuçlar*

Problem Çözmeye Açıklık ölçeği için Türkiye ve Finlandiya alt grupları için OKYD yöntemi kullanılarak elde edilen ölçme değişmezliği sonuçları Tablo 3'de sunulmuştur.

**Tablo 3.** Türkiye – Finlandiya örnekleme OKYD yöntemi ile elde edilen analiz sonuçları

Aşama	$\chi^2$	sd	CFI	GFI	$\Delta\chi^2$	$\Delta$ sd	$\Delta$ CFI	$\Delta$ GFI
Yapısal değişmezlik	1145,750	15	0,950	0,950				
Zayıf değişmezlik	1160,960	19	0,950	0,950	15,210	4	0,000	0,000
Güçlü değişmezlik	1510,560	24	0,940	0,950	349,600	5	-0,010	0,000
Katı değişmezlik	1513,820	25	0,940	0,950	3,260	1	0,000	0,000

Tablo 3'te verilen sonuçlar incelendiğinde yapısal değişmezlik aşamasında elde edilen uyum indekslerine göre yapısal değişmezliğin sağlandığı görülmektedir. İkinci adımda yapılan incelemelerde ise modelin karşılaştırmalı ve iyilik uyum indekslerine göre zayıf değişmezliği sağladığı fakat ki-kare değerleri farkının anlamlılığına göre sağlamadığı görülmektedir. Daha öncede belirtildiği ki-kare uyum indeksinin serbestlik derecesi ve örneklem büyüklüğü parametrelerinden etkileniyor olması nedeniyle diğer iki uyum indeksine göre elde edilen sonuçlar göz önüne alındığında zayıf değişmezliğin sağlandığı yorumu yapılmıştır. Güçlü değişmezlik aşaması ile ilgili olarak elde edilen sonuçlara göz atıldığında ise yine benzer sonuçlar elde edildiği görülmektedir. Son olarak aşamada ise yine karşılaştırmalı, iyilik uyum indekslerine ve ki kare değerleri farkının anlamlılığına göre katı değişmezliğin sağlandığı görülmektedir.

### **Çoklu Grup Örtük Sınıf Analizi Yöntemine İlişkin Sonuçlar**

ÇGÖSA yöntemi ile yapılan analizlerden uygun sınıf sayısının belirlenmesi ile ilgili olarak elde edilen sonuçlar tablo 4'te verilmiştir.

**Tablo 4. ÇGÖSA örtük sınıf sayısı belirlemeye ilişkin sonuçlar**

	Tahminlenen parametre sayısı	L <sup>2</sup>	BIC	AIC	AIC3	CAIC	sd
1. SINIF	10	13860,00	9552,63	12911,60	12437,60	9078,63	474
2. SINIF	22	3543,80	-654,15	2619,81	2157,81	-1116,20	462
3. SINIF	34	1012,60	-2858,20	160,61	-265,38	-3284,20	450
4. SINIF	46	1251,90	-27280,00	375,86	-62,13	-3166,00	438
5. SINIF	58	1905,10	-2183,80	1005,12	555,12	-2633,80	426

Yapılan örtük sınıf sayısı belirleme için yapılan analizlerde en düşük BIC değerinin 3 sınıflı model için elde edildiği belirlenmiştir. Bu sonuca göre veri seti için en uygun modelin 3 sınıflı olduğuna karar verilmiştir. Bu durum problem çözmeye açıklık örtük değişkenin 3 seviyesinin olduğunun göstergesidir. Yapılan ölçme değişmezliği model çalışmaları sınıflı model üzerinden yürütülmüştür. Heterojen, kısmi homojen ve homojen model için gerçekleştirilen analiz sonuçları Tablo 5'te verilmiştir.

**Tablo 5. ÇGÖSA modellerine ilişkin sonuçlar**

Model	Tahminlenen parametre sayısı	L <sup>2</sup>	BIC	AIC	AIC3	CAIC	sd
Heterojen model	64	1659,45	-2156,87	819,45	399,45	-2576,87	420
Kısmi homojen model	32	1925,26	-2181,83	1021,26	569,26	-2633,83	452
Homojen model	22	2253,89	-1944,06	1329,89	867,89	-2406,06	462

Test edilen modellere ilişkin sonuçlar incelendiğinde en düşük BIC değerinin kısmi homojen model için olduğu görülmektedir. Bu sonuç bazı maddelerin sabit ya da eğim parametrelerinin eşdeğer olmadığının göstergesidir (Güngör, Korkmaz ve Somer, 2013). Bu noktada madde düzeyinde analizler yapılarak ve hangi maddenin eş değeri bozduğu belirlenebilir.

### **Tartışma ve Sonuç**

Elde edilen sonuçlar, YEM altındaki yöntemlerden OKYD ile ÖSA yöntemlerinden ÇGÖSA'nın ölçme değişmezliği için vermiş olduğu sonuçların farklılaştığını göstermektedir.

OKYD ile elde edilen sonuçlara göre katı değişmezliğin sağlandığı belirlenmişken, ÇGÖSA ile yapılan analizlere göre kısmi homojen model kabul edilmiştir. Kısmi homojen modellerin OKYD için yapılan basamaklar için karşılığı zayıf değişmezliktir. Bu durumda OKYD için ölçme değişmezliğinin sağlandığı fakat ÇGÖSA için bazı maddelerin eş değerliliği bozduğu bulgusuna ulaşıldığı söylenebilir.

Araştırmacıların çalışmalarında kullandığı veri setlerinin ve değişkenlerin yapısına göre ölçme değişmezliği yöntemlerini doğru bu şekilde seçmesi oldukça önemlidir. Varsayımları olan yöntemleri seçen araştırmacıların, analizin gerektirdiği varsayımları test etmesi ve sonuçlar doğrultusunda analizlere devam etmesi elde edilen sonuçların doğruluğunu etkileyebilir. Yapılan bu çalışmada normallik varsayımı gerektiren YEM çatısı altındaki yöntemlerin normallik varsayımı sağlanmayan durumlarda diğer yöntemlerden farklı sonuçlar verebileceğini göstermiştir. Elde edilen sonuçlara dayalı olarak yapılan yorumlamalarda da bu doğrultuda yanlışlar olabilir. Bu noktada araştırmacılar mutlaka eş değerliliğin test edilmesinde farklı yöntemler kullanılmalıdır. Nitekim Kankaras, Vermunt ve Moors (2011), eşdeğerlik çalışmaları ve grup karşılaştırmaları yapılan çalışmalarda farklı yöntemlerin kullanılmasını tavsiye etmektedir.

Bu çalışmayla birlikte ÇGÖSA analizlerinin sürekli yapıdaki değişkenler içinde iyi bir alternatif olduğu söylenebilir. ÇGÖSA'nin yalnızca süresiz yapıdaki değişkenler için kullanılması gerektiği gibi bir sınırlılık yoktur. Elde edilen sonuç Kankaras, Vermunt ve Moors (2011), ölçme değişmezliğinin test edilme sürecinde MTK, YEM ve ÖSA çatısı altındaki yöntemleri karşılaştırdığı çalışmada belirttiği sonuçlarla örtüşmektedir.

#### **Kaynaklar**

- Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme – I*. Ankara: PegemA Yayıncılık
- Messick, S. (1994). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749
- Başusta, N.B. ve Gelbal, S. (2015). Gruplararası karşılaştırmada ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği. *Hacettepe Eğitim Fakültesi Dergisi*, 30(4), 80-90.
- Bryne, B. M. and Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, 34(2), 155-175.
- Flowers, C.P., Raju, N.S. and Oshima, T. C. (2002). A comparison of measurement equivalence methods based on confirmatory factor analysis and item response theory. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.
- 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.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.
- Steenkamp E. M. and Baumgartner H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1). Oxford University Press.
- Kankaras, M., Vermunt, J. K. and Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research*, 40(2), 279-310.
- Kankaras, M. and Moors, G. (2009). Measurement equivalence in solidarity attitudes in Europe. Insights from a multiple group latent class factor approach. *International Sociology*, 24(4), 557-579.
- Güngör, D., Korkmaz, M. ve Somer, O. (2013). Çoklu-Grup Örtük Sınıf Analizi ve Ölçme Eş Değerliliği. *Türk Psikoloji Dergisi*. 28(72), 48-57.
- Güngör, D., Korkmaz, M. ve Sazak, H. S. (2015). Örtük Sınıf Analiziyle Yapılan Ölçme Eşdeğerliği Çalışmalarında Model Seçimi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*. 30(1), 90-105.



- Vandenberg, R. J. and 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.
- Şimşek, Ö. F. (2007). *Yapısal Eşitlik Modellemesine Giriş Temel İlkeler ve LISREL Uygulamaları*. Ankara: Ekinoks Yayınları.
- Önen, E. (2009). *Ölçme değişmezliğinin yapısal eşitlik modelleme teknikleri ile incelenmesi*. (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara.
- Cheung, G.W. and Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Wu, A. D., Li, Z. and Zumbo, B. D. (2007) Decoding the meaning of factorial invariance and updating the practice of multiple-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, 12, 1-26.
- Moors, G. and Wennekers, C. (2003 ). Comparing moral values in western european countries between 1981 and 1999. A multigroup latent-class factor approach. *International Journal of Comparative Sociology*, 44, 155-172.

## **Extended Abstract**

### **Introduction**

The basic data source for social sciences is individuals. In the literature, there are a lot of studies that aim to determine the differences between the individual members of different groups for a psychology structure. According to the results of these studies, the directors make important decisions and arrangements. Therefore measurement invariance of the instruments used in these studies ensues. The measurement invariance means revealing the levels of an individual's psychological features without having a bearing on other features. In other words the differences determined by comparing groups must be only predicated on the differences about the psychological feature asked for measuring. In this context measurement invariance is very important for the trueness of the comparisons and the decisions made from these comparisons. Measurement invariance is also important for validity of measurement instruments. However researchers do not attach importance to the measurement invariance, this may be the resulted from lack of necessary knowledge and different views about the testing process.

Some shortcomings are found in the studies where measurement invariance is tested. Some of those are: using inappropriate methods for the observed structures and latent variables and not testing the assumptions of the methods. Within this context, studies trying to solve complications and aiming to determine the differences between the methods diversifying as relevant to assumptions are important for helping for other researchers. In this study it is aimed to determine the measurement invariance differences for the same data by using the methods relevant to the assumptions. Besides measurement invariance results of two methods requiring normality assumption and not requiring normality assumption were compared to each other.

### **Method**

Data of Openness for Problem Solving scale consisting of five items and applied in PISA 2012 of Finland and Turkey samples were used. Analyses were carried out on 3159 Turkish and 5676 Finn students' data that remained after deleting the invalids and missing. Mean and Covariance Structure (MACS) analysis, a Structural Equation Modeling method and Multi-Group Latent Class analysis (MGLCA) and a Latent Class Analysis (LCA) method were used for testing measurement invariance.

Before testing measurement invariance, test of normality was checked for all items by Kolmogorov Smirnov. The stepwise method proposed by Meredith (1993) was used for MACS analysis. The analyses were carried out by keeping up with the steps consisting of the least restricted model named configural invariance to the most restricted model named strict invariance in this method. Significances of the differences between chi-squares of sequential models were used for making decisions. Comparative and goodness of fit indexes were also used. Both of the latent and observed variables must be continuous for MACS analysis. Besides all of items in measurement instrument must demonstrate normality. In this study LISREL 8.72 program was used for MACS analysis.

Multi-Group Latent Class Analysis is one of the methods of the LCA which aims to recognize the categorical latent variables for discontinuous data. In this method, researcher started analysis by determining the number of latent class. After that, heterogeneous, partial homogenous and homogenous models were tested for model suited admitted number of class. Bayesian and Akaike information criteria, as well as their derivatives, were checked for this method. The model has the minimal criteria which are accepted. In this study Bayesian information criteria, stated as giving the best results on big samples, was used. MGLCA do not require the normality for items. And MGLCA is a good alternative for analysis carried out on data consisting of the continuous variable. Latent Gold 5.1 program was used for MGLCA.

### **Result and Discussion**

According to results of Kolmogorov Smirnov test, none of the items demonstrated normality at the significant level of 0,01 for both of two subsamples and the whole group that using MACS analysis was not suitable for analyses. But analyses were continued for revealing the results based on shortcomings of testing assumptions process.

Strict invariance model was accepted as a result of MACS analysis. Differences of goodness and comparative fit indexes of sequential models were between -0,01 and 0,01. Firstly the number of class was determined to be three classes in the process of MGLCA. In other words we can say that the latent variable named Openness for Problem solving was triple level. After the determination of triple level model, partial homogenous model was accepted by examining the information criteria. The least Bayesian information criteria were calculated for partial homogenous model.

When results obtained from the two methods were compared, strict invariance step was accepted for MACS analysis, whereas partial homogenous model equaled to weak invariance for MACS analysis was accepted. The results can differentiate because of normality. The result shows that the assumptions of the methods certainly must be examined and proper method must be chosen according to verification of assumptions. Researchers can fail about measurement invariance if they do not choose suitable method for the data. Therefore researchers must choose at least two methods for testing measurement invariance and must compare the results obtained from different methods.