

Gruplararası Karşılaştırmalarda Ölçme Değişmezliğinin Test Edilmesi: PISA Öğrenci Anketi Örneği

Examination of Measurement Invariance at Groups' Comparisons: A Study on PISA Student Questionnaire

N. Bilge BAŞUSTA*, Selahattin GELBAL**

ÖZET: Anlamli karşılaştırmalar yapmak için kullanılan ölçme araçlarının birde fazla grupta aynı yapıda olması gerekmektedir. Ölçmelerin değişmezliği ile ilgili çalışmalar çok maddeli ölçme araçlarında iç tutarlılık ve dışsal ilişkileri araştırmaktadır. Bu çalışmada PISA öğrenci anketinde yer alan fen bilgisi ve teknolojileri ile ilgili maddelere ait temel faktörleri ortaya çıkarmak ve bu faktörlerin cinsiyet bağlamında ölçme değişmezliğine sahip olup olmadığı incelenmiştir. Bu çalışmanın örneklemini Türkiyeden PISA 2009'a katılan 4942 öğrencinin ankete vermiş olduğu yanıtlar oluşturmuştur. Veriler yapısal eşitlik modellemesi (YEM) tekniğini kullanarak analiz edilmiştir. Oluşturulan ölçme modeli aşama aşama sınırlandırılarak ölçme değişmezliği test edilmiştir. Değişmezliğin testinde uyum indeksleri farkı kullanılmıştır. Dört aşama için elde edilen LISREL çıktıları, standart değerlerden yola çıkılarak karşılaştırılmış ve cinsiyet bakımından elde edilen değerlerin ölçme değişmezliği bakımından sorun teşkil etmediği tespit edilmiştir.

Anahtar sözcükler: Ölçme Değişmezliği, Çoklu Grup Doğrulamalı Faktör Analizi, Yapısal Eşitlik Modeli (YEM), Cinsiyet

ABSTRACT: In order to make a meaningful comparison, instruments must have similar measurement qualities across settings. Research on measurement invariance (MI) has typically focused on internal structure of multi-item instruments and external relations. The purpose of this study is to investigate the factors of PISA student questionnaire's items and assessing the measurement invariance of these factors across gender. This study examined the PISA data for Turkish students with the sample size of 4942 through the analysis of Structural Equation Modeling (SEM). First, it is designed a measurement model over the factors from student questionnaire with Structural equation modeling. It was understood that the designed measurement model's goodness fit index output is good. Then four forms of invariance were assessed for each latent variables progressively and assessed MI. The variables were examined using comparative fit index (CFI). The MI results revealed that there were no invariance problems across gender in PISA.

Keywords: Measurement Invariance, Multi Group Confirmatory Factor Analysis (MGCFA), Confirmatory Factor Analysis, Gender

1. GİRİŞ

Sosyal bilimler alanında yapılan birçok araştırmada, belirli psikolojik yapılar açısından gruplar arası (kültür, cinsiyet, sınıf düzeyi ve sosyoekonomik düzey vb.) karşılaştırmalara dayanan çalışmaların göz ardı edilmeyecek kadar fazla olduğu bilinmektedir. Yapı kavramı, psikolojik özelliklerin ölçülmesi söz konusu olduğunda ölçülmek istenilen değişkeni ifade etmektedir (Cronbach ve Meehl, 1955). Birçok araştırmada psikolojik yapılar açısından gruplar arası farklılıkların niteliğine ve niceliğine yönelik anlamlı çıkarımlarda bulunmak amaçlanmaktadır. 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. Örneğin Amerikalı işçilerden oluşan bir grup ile Çin işçilerden oluşan bir grubun iş memnuniyet derecesi araştırılırken kullanılan ölçekten elde edilen yanıtlar, grupların ölçek maddelerini yorumlama biçimine göre değişecektir (Cheug ve Rensvold; 1998). Söz konusu ölçeğin farklı ülkelerde farklı dil formları kullanılarak, uzmanlarla

* Yrd. Doç. Dr., Mersin Üniversitesi, Eğitim Fakültesi, Mersin-Türkiye, n.bilgeuzun@gmail.com

** Prof. Dr., Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Ankara-Türkiye, gelbal@hacettepe.edu.tr

dille bağlantılı olan ayrılıkların giderilmesi yönünde çalışmalar yapılırsa da, ölçeğin farklı ülkelerde aynı anlama geldiği ve ölçeği cevaplayan bireylerce aynı biçimde yorumlandığı sağlanmış olmayacaktır (Cheung ve Rensvold; 1999). Bir ölçme aracı hazırlanırken; “araç uygulandığı her grupta aynı özelliği ölçer” varsayımı ile hazırlanmaktadır. Oysa yukarıdaki örnekte ölçme aracının uygulandığı gruplarda elde edilecek ölçümlerde farklılaşmasından ötürü güvenilirlik ve geçerlik çalışmalarının farklılaşabileceği anlaşılmaktadır. Diğer bir ifadeyle gruplardan elde edilen ölçümlerin eşit/denk psikometrik niteliklere sahip olmayabileceği belirtilmektedir. Dolayısıyla gruplar için genellenebilirlikten söz etmek yanlış olacaktır.

Ölçme aracı ile bir özelliğin oluş derecesi ortaya konmaya çalışılır. Çeşitli alt gruplarda bu özellik açısından karşılaştırmalar yapılır ve bu biçimde gruplar hakkında bilgi sağlama yoluna gidilir. Bu tanımlamanın doğruluğu anlaşılacağı gibi ölçme aracı ve elde edilen ölçüm ile doğrudan ilgilidir. Çalışmalarda gruplar arasında bulunan farklılıklar ilgili psikolojik yapı açısından gruplar arası “gerçek” değişim olarak kabul edilmektedir (Mark ve Wan, 2005). Bir yapı ortaya konmaya çalışılırken söz konusu yapıyı ölçmeye yönelik kullanılan ölçme aracı farklı gözlem ve çalışma koşulları altında aynı ölçümleri vermelidir. Benzer biçimde bir yapı bakımından gruplararası farklılıkları ortaya koymaya amaçlayan çalışmalarda gruplar için ölçme aracının benzer şekilde çalışacağı ve ölçme aracından elde edilecek ölçümlerin eşit psikometrik özelliklere sahip olduğu varsayılmaktadır. Başka bir ifadeyle istatistiksel olarak karşılaştırılan iki grupta gözlenen değişkenler ile ilişkilendiren örtük değişkenlere ait matematiksel bağıntı eşit olmak durumundadır. Bu açıklamalar; ölçme değişmezliğinin (Measurement Invariance: MI) temel konusudur. Ölçme değişmezliği ölçülen bireylerin karakteristikleri ile değil ölçme aracının direkt kendisiyle ilgilidir. Birden fazla grup söz konusu olduğunda her grup için ölçme aracının önemli niteliklerinden biri olan yapı geçerliliğini sağlamanın yollarından biridir. Her grubun yapı geçerliliği kanıtları, bu grupların hepsinde geçerli olduğu düşünülen (beklenen) faktör yapısıyla uyum göstermelidir.

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 (Byrne, 1998; Jöreskog ve Sörbom, 1993). Değişmezlik çalışmaları ile yeni bir ölçme aracı geliştirilmez. Ölçmelerin değişmezliği gruplar arası karşılaştırmalarda bu karşılaştırmaların anlamlılığında bir koşul olarak ortaya çıkar (Bollen, 1989; Cheung ve Rensvold, 2000). Bu çalışmalar ölçme aracının elde edilen ölçümlerde eşit yapıyı ortaya koyup koymadığını ortaya koymaya çalışır. Farklı grupların aynı ölçme aracı ile karşılaştırıldığı çalışmalarda, “ölçme aracından kaynaklanabilecek farklılıklar yoktur” biçimindeki bir kabul, karşılaştırmaların ve yorumların eksik kalmasına neden olabilmektedir (Başusta Uzun ve Öğretmen; 2010). Bir ölçme aracının faktör yapılarının iki grupta farklı olması her grupta farklı bir psikolojik özelliği ölçüldüğüne ilişkin bilgi sağlar. Böyle bir durumda aynı ölçme aracıyla ölçme yapıyor olmasına rağmen iki gruptan elde edilen sonuçların aynı anlama gelmediği yorumu yapılır.

Gruplar arası karşılaştırma araştırmalarının temel amacının, araştırmalarda kullanılan ölçümlerin gruplar arası eşitliğinde yattığı belirtilmektedir. Horn ve Mc Ardle (1992) gruplararası karşılaştırmaların yapılabilmesi için, ölçme değişmezliğinin incelenmesinin mantıksal bir ön koşul olduğunu belirtmektedir. Ölçme değişmezliğinin temelinde, “ölçülen özellik açısından belirli bir düzeyde bulunan bireylerin söz konusu araçtan alması beklenen puanların, grup üyeliğinden bağımsız olması” düşüncesinin yattığı belirtilmektedir (Wicherts, 2007). Eğer ölçme değişmezliği gösterilemezse, araştırmanın gruplar arası niceliksel karşılaştırmalardan ziyade niteliksel yorumlamalarla sınırlandırılmasının daha uygun olacağı ifade edilmektedir. Meredith (1993), ölçme değişmezliğinin gösterilmesi için dört aşamalı mantıksal bir süreç ve hipotez test etme yöntemlerinin izlenmesini önermektedir. Bunlar;

Biçimsel değişmezlik (configural invariance): Ölçmelerin değişmezliğindeki en temel düzeydir. Bu aşamada bir psikolojik ölçme aracının faktör yapısının gruplar arası eşit/değişmez olduğu şeklindeki bir hipotez test edilmektedir. Biçimsel değişmezliğe ilişkin kanıt elde edilmesi, ölçme aracının maddelerinin –gruplararası- aynı psikolojik yapıyı temsil ettiği anlamına gelmektedir.

Metrik değişmezlik (metric invariance): Farklı grupların maddelere aynı biçimde cevap verdiğini ve böylece farklı gruplardan elde edilen puanların karşılaştırma da anlamlı olduğu sayıtlıları metrik değişmezlikte karşılanmalıdır (Steenkamp ve Baumgartner, 1998). Bu aşamada, bir psikolojik ölçme aracını oluşturan maddelere ilişkin regresyon eğimlerinin diğer adıyla faktör yüklerinin gruplar arası eşit/değişmez olduğu şeklindeki bir hipotez test edilmektedir.

Ölçek değişmezliği (scalar invariance): Bu aşamada, psikolojik ölçme aracını oluşturan maddeler için oluşturulan regresyon denklemlerindeki sabit sayının, gruplar arası eşit/değişmez olduğu şeklindeki bir hipotez test edilmektedir. Ölçek değişmezliği hem metrik değişmezliği hem de ölçme işleminde eşit orjinleri gerektirmektedir. Ölçme değişmezliğinin bu aşaması; aynı değere sahip konuların örtük yapı üzerindeki değerinin gözlenen değer üzerindeki değerine eşit olduğunu ifade eder. Bu ifade “gözlenen maddeler üzerindeki grup farklılıkları örtük yapıların ortalamalarından kaynaklanır” demektir.

Katı değişmezlik (strict invariance): Bu aşamada, ölçme aracını oluşturan maddelere ilişkin özgül varyansların diğer bir deyişle hata terimlerinin karşılaştırma grupları arasında eşit/değişmez olduğu şeklindeki bir hipotez test edilmektedir.

Bu aşamalardan her birinde, o düzeyde ölçme değişmezliği için oluşturulan bir hipotez sırayla test edilmektedir. Bu süreçte her aşamadaki model, bir önceki aşamadaki modele dayalı olarak oluşturulmaktadır. Dolayısıyla belirli bir aşamadaki ölçme değişmezliği, o aşamadaki model ile bir önceki aşamadaki modelin veriye uyum düzeylerinin karşılaştırılması yoluyla incelenmektedir.

Ülkemizde ölçme değişmezliği ile çalışmaların sayıları sınırlıdır. Oysa bir ölçme aracının en önemli niteliklerinden biri olan geçerlik için benzer çalışmaların artırılması; ulusal ve uluslararası sınavların geçerlik kanıtı olarak sunulması oldukça önemlidir. Özellikle PİOSA, TIMSS, PIRLS gibi uluslararası sınav verileri ile yapılan çalışmalar; eğitimi yönlendirmesi bakımından önem kazanmaktadır.

Önen'in (2009) çalışmasında Epistemolojik İnançlar Envanteri'nin Türkçe'ye uyarlaması yapılarak, ardından envanterin Türkçe Formu için cinsiyet grupları arası ölçme değişmezliği incelenmiştir. Geçerlik ve güvenilirlik düzeylerine ilişkin kanıtlar elde edildikten sonra uyarlama çalışmalarına katılmayan 818 öğrenciden elde edilen verilerden yararlanılmıştır. Envanterin Türkçe Formu için cinsiyet grupları arası kapsam ve ölçme eşitliğinin incelenmesi amacıyla tek gruplu ve çok gruplu DFA ve olası madde yanlılıklarını taramak amacıyla da Lagrange Multiplier Test uygulanmıştır. Yapılan analizler sonucunda envanterin Türkçe formu için cinsiyet grupları arası şekilsel değişmezliğe, tam metrik değişmezliğe, kısmi ölçek değişmezliğine, kısmi değişmez özgüllüğe ve faktör varyanslarının değişmezliğine ilişkin kanıtlar elde edilmiştir. Ayrıca metrik ve ölçek değişmezliklerinin test edilmesi sonucunda, dört maddenin cinsiyet grupları arası tek biçimli yanlılık sergiledikleri gözlenmiştir.

Uzun ve Öğretmen (2010) yapmış oldukları çalışmada TIMSS Türkiye verilerini, Yapısal Eşitlik Modellemesi (YEM) tekniği kullanarak analiz etmiştir. Modelde yeralan örtük değişkenler (özyeterlik, önem, tutum ve sınıf içi öğrenci etkinlikleri) için aşamalı bir şekilde

değişmezlik testleri yapılmıştır. Analizlerde hiçbir sınırlandırmanın yapılmadığı (biçimsel değişmezlik) durum ile daha fazla sınırların şart konulduğu diğer değişmezlik testleri arasındaki CFI karşılaştırmacı uyum iyiliği kriterleri farkına bakılmıştır. Ölçmelerin değişmezliği çalışmalarından modele alınan tüm değişkenlerin metrik değişmezlik koşulunu yerine getirdiği, ancak cinsiyet gruplarında katı değişmezlik koşulunu hiçbirinin sağlayamadığı görülmüştür.

Bu konu ile ilgili yapılan çalışmalardan birini Akyıldız (2011) yürütmüştür. Akyıldız yapmış olduğu çalışmada, PIRLS 2001 başarı testinin, 35 ülkede yapılmış olan uygulamasından elde edilen faktör yapılarının birbirine denkliği çok gruplu doğrulayıcı faktör analizi ile incelemiştir. PIRLS 2001 başarı testinin faktör yapısının 35 ülkede değişmezliği tüm ülkelerde uygulanmış olan dört ayrı alt ölçek bağlamında ayrı ayrı test edilmiştir. Elde edilen bulgular PIRLS 2001 testinin faktör yapısının uygulamaya katılan tüm ülkelerde birbirine orta düzeyde denk sayılabileceğini göstermiştir.

Bahadır (2012) tarafından yürütülen benzer bir çalışmada Türkiye'deki öğrencilerin okuma becerilerini etkileyen değişkenlerin PISA 2009 verilerini kullanarak modellenmesi ve modele alınan değişkenlerin bölgeler arasındaki farklılıkların incelenmesi amaçlanmıştır. Okuma becerisini etkilediği düşünülen değişkenler üç boyut altında toplanmıştır. Bunlar; anne-baba eğitim düzeyi ve öğrenciye sağlanan olanakların olduğu *ail*e değişkeni, evde bulunan kültürel varlıkların olduğu *kulvar* değişkeni ve öğrencilerin okumaya karşı tutumlarının olduğu *tutum* değişkenidir. Öğrencilerin okuma becerileriyle modele alınan değişkenler arasındaki ilişkileri yordayan yapısal modelin uyum indeksleri incelenmiş ve modelin iyi uyum gösterdiği sonucuna ulaşılmıştır. Modele alınan değişkenlerin coğrafi bölgeler arasında karşılaştırılabilirliği için ölçme değişmezliği test edilmiştir. Model skalar değişmezlik şartlarını sağlamadığından kısmi değişmezlikten faydalanılarak, değişmezlik aşamaları test edilmiştir. Modele alınan değişkenlerin faktör yükleri incelendiğinde, bölgeler arasında en büyük farklılığın anne-baba eğitim düzeyi ve öğrenciye sağlanan olanaklarda olduğu görülmüştür. Beş bölgede en yüksek faktör yükü gösteren değişkenler anne-baba eğitim durumları iken Doğu ve Güneydoğu Anadolu bölgesinde öğrencilere sağlanan imkânların daha yüksek faktör yüküne sahip değişkenler olduğu sonucuna ulaşılmıştır.

Bu çalışmada OECD'nin yürütmekte olduğu, OECD'ye üye ülkelerin katılımıyla gerçekleşen; ülkemizde de uygulanan PISA 2006 öğrenci anketinden elde edilen veriler kullanılarak cinsiyetlere ilişkin fen ve çevre konularına yönelik maddelerinin değişmezliği incelenmiştir. Temel amaç fen bilimlerine yönelik maddelerle oluşturulan ölçme modelinin cinsiyetler için aynı yapıda olup olmadığını test etmektir. Çalışmada ölçme değişmezliğinin test edilmesi; kurulan modele ilişkin cinsiyetler için yapılan herhangi bir karşılaştırmayı anlamlı kılabilecek nitelikteki diğer çalışmalara rehberlik edecektir. Aynı zamanda PISA 2006 öğrenci anketinin cinsiyete göre yapılan karşılaştırmaların anlamlılığı konusunda bize daha geçerli bilgiler sağlayabilecektir.

Araştırma problemimiz PISA öğrenci anketinin faktör yapısı için tanımlan modele ilişkin ölçme değişmezliği kanıtları bulunup bulunmadığıdır.

Uluslararası Öğrenci Değerlendirme Programı – PISA (**P**rogramme for **I**nternational Student Assessment), Ekonomik İşbirliği ve Kalkınma Örgütü OECD'nin üç yıllık aralarla düzenlemekte olduğu ve 15 yaş grubu öğrencilerin kazandıkları bilgi ve becerilerin değerlendirilmesine yönelik yapılan bir tarama araştırmasıdır. Üçer yıllık dönemler halinde uygulanan PISA projesinde, her bir dönemde bir konu alanına ağırlık verilmektedir. PISA projesinin ilki 2000 yılında uygulanmış, bu uygulamada okuma becerilerine ağırlık verilmiştir. 2003 yılında yapılan ikinci uygulamada Matematik okuryazarlığı alanına, 2006 yılında yapılan

uygulamada ise Fen Bilimleri alanına ağırlık verilmiştir. Yapılan bu uygulamalar ile üç konu alanını içeren 9 yıllık bir dönem sona ermiştir (PISA Ulusal Ön Rapor, 2007).

2. YÖNTEM

Bu araştırma, 2006 PISA'ya katılan öğrencilerden toplanan verilerle öğrencilerin fen ve çevre konularına yönelik maddelerinin değişmezliğini cinsiyetlere göre inceleyen ilişkisel bir araştırmadır.

2.1. Örneklem ve Veri Toplama Araçları

Çalışma grubu, PISA 2006 yılı uygulamasına Türkiye'de katılan öğrencilerden oluşmaktadır. Araştırmada 4942 (2290 kız öğrenci ve 2652 erkek öğrenci) öğrencinin verilerinden yararlanılmıştır. Çalışmaya katılan öğrencilerin yaş ortalaması 15,9 olarak bulunmuştur. PISA 2006 yılı uygulamasında yer alan öğrenci anketindeki fen bilgisi ve teknolojisine ilişkin maddelerin cinsiyet bağlamında değişmez olup olmadığı araştırılmıştır.

2.2. Veri Çözümleme Teknikleri

Çalışmada verileri analiz etmek için Lisrel 8.80 for Windows programı kullanılmıştır. Önce analizi yapılacak verilerin dağılımları betimsel istatistikler aracılığı ile incelenmiş, daha sonra çok gruplu doğrulayıcı faktör analizi uygulanmıştır. Çok gruplu doğrulayıcı faktör analizine başlamadan önce açıklayıcı faktör analizi ve bu analizin sonuçlarını kullanarak bir ölçme modeli kurulmuştur. Bu iki analiz ile Yapısal Eşitlik Modeli (YEM) ile regresyon modelindeki değişkenler arasındaki yordayıcı yapısal ilişki ve faktöranalizindeki gizil faktör yapılarını kapsamlı bir analizde birleştirmek amaçlanmaktadır.

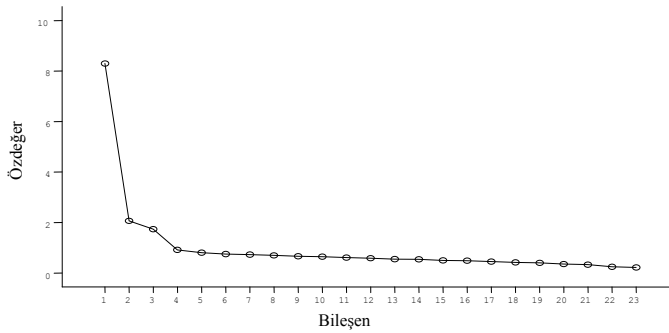
PISA projesi kapsamında uygulanan öğrenci anketi ile oluşturulan modelin faktör yapılarının cinsiyetler arası değişmezliği yapısal eşitlik modelleri yöntemlerinden çok gruplu doğrulayıcı faktör analizi yöntemi ile incelenmiştir. Birden fazla grubun bulunduğu grup karşılaştırmalarında çok örneklemlili doğrulayıcı faktör analizi grup parametrelerinin eşit kalarak, faktör ortalamalarının karşılaştırılmasını sağlamaktadır. Diğer bir ifadeyle bu yöntem ölçme aracının gruplar arası değişmezliğinin sağlanmasında kullanışlı sağlamaktadır. Yapısal eşitlik modeline bağlı olarak yapılan grup karşılaştırması gözlenen değişkene dair yapılan ölçümlerden 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 (Sehe, Mary, Malik; 2003). Özellikle psikolojik test geliştirilirken testin sahip olması istenen psikometrik özelliklerin testin uygulanacağı gruplarda değişmez olduğunun gösterilmesi için çok gruplu doğrulayıcı faktör analizi sıklıkla kullanılmaktadır (Thompson, 2004). Çoklu grup uygulamaları ile YEM kullanılarak kurulmuş bir modeldeki örtük değişkenlerin gruplar arasında farklılık yaratıp yaratmadığı sorusuna yanıt aranmaktadır. Çoklu grup uygulamalarında doğrulayıcı uyum katsayıları (CFI) arasındaki farklar kullanılmıştır. Uyum iyiliğinin değerlendirilmesinde örneklem büyüklüğünden kaynaklanan istatistiksel zayıflıklara karşın ki kare yerine CFI, GFI değerlerinin daha uygun olduğunu belirten çalışmalar bulunmaktadır. Wu ve diğ., (2007) TIMSS datası ile yapmış oldukları çalışmada ki kare farkı yerine CFI uyum katsayıları arasındaki farkın kullanılması gerektiğini şu biçimde açıklamışlardır. Ki kare farkı örneklem büyüklüğünün bir fonksiyonudur ve MI için verilecek kararda tek başına kullanılması doğru değildir. Yapılan bazı çalışmalarda da ölçmelerin değişmezliği için ki kare farkına güvenilmemesi gerektiği belirtilmiştir (Cheung ve Rensvold, 2002; Vandenberg ve Lance, 2000). Bu çalışmalarda bu biçimde sınırlandırılan modellerde uyum katsayılarında ki değişimin kullanılmasının daha doğru ve güvenilir sonuçlar

doğuracağı da desteklenmiştir. Bunun için kullanılması gereken uyum katsayıları hakkında farklı görüşler mevcuttur. (Cheung ve Rensvold, 2000). Bu çalışma belirlenen temel model çerçevesinde değişmezlik testi olarak sırasıyla; biçimsel değişmezlik sonucu elde edilen CFI ile metrik değişmezlik için bulunan CFI, skalar değişmezlik için bulunan CFI ve katı değişmezlik için bulunan CFI farkları incelenmiştir. CFI farklarının incelenmesinin nedeni ise uyum katsayılarının örtük skorlarla gözlenen skorlar arasındaki ilişkiye dair bilgi vermesinden kaynaklanır. Diğer uyum katsayıları da ölçmelerin derğişmezliği çalışmalarında kullanılabilir. Uyum katsayılarının eş olduğu durumda maddelere ait skorların gruplar arasında anlamlı biçimde karşılaştırılması yapılabilir. Bu çalışmada veri ile uyum gösteren MGCFI çalışma dosyaları için $0.01 \geq \Delta CFI \geq -0.01$ olan değişmezlik koşullarının sağlanıp sağlanmadığına bakılmıştır.

Ayrıca her bir değişmezlik testi sonunda model veri uyumu hakkında bilgi sahibi olabilmek adına CFI, χ^2/sd oranı ve kestirimin hata kareleri ortalaması (RMSEA) indeksleri rapor edilmiştir. Yapısal eşitlik modellemesi analizlerinde ki kare değerlerinin büyük örneklemelerden etkilenerek her koşulda anlamlı farklılık bildirme eğiliminde olduğu bilinmektedir (Brown, 2006). İyi bir model veri uyumu için örneklem büyüklüğünden kaynaklanan istatistiksel zayıflıklara karşı daha güçlü olan ve büyük örneklemelerde diğer uyum indekslerine oranla daha uygun olduğu bildirilen RMSEA değerleri ve CFI değerleri (Cheung ve Rensvold, 2002) özellikle dikkate alınmıştır. RMSEA değerlerinin 0.05'ten küçük olmasının, CFI değerinin ise 0,90'dan yüksek olmasının iyi bir model-veri uyumunu işaret ettiği kabul edilmektedir.

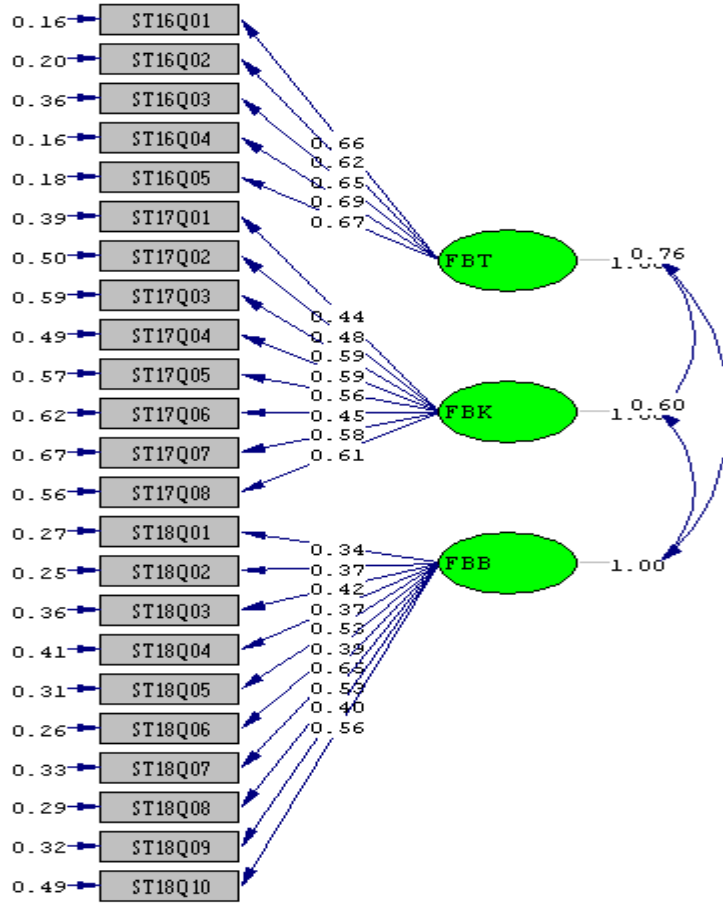
3. BULGULAR

PISA öğrenci anketinin 3. Bölümünde yer alan fen bilimleri ile ilgili görüşleriniz kısmında yer alan 23 maddeye açımlayıcı faktör analizi yapılarak madde kümeleşmelerinin oluşturulan teoriye ve modele uygun olup olmadığına bakılmıştır. Faktörlerin özdeğerlerine dayalı olarak çizilen çizgi grafiği aşağıda verilmiştir.



Şekil 1: Faktörlerin özdeğerlerine dayalı olarak çizilen çizgi grafiği

Şekil 1 de faktörlerin özdeğerlerine ilişkin çizilen çizgi grafiğinde 3 faktörün belirgin olarak öne çıktığı görülmektedir. Açımlayıcı faktör analizi uygulanan 23 maddeden 5 madde fen bilimine yönelik Tutum (FBT), 8 madde fen bilimine yönelik bilgilerini kullanmaya ilişkin inanç (FBK), 10 madde de fen bilimine yönelik bakış (FBB) faktörleri olmak üzere 3 faktör altında toplanmıştır. Üç faktörün açıkladığı toplam varyans %52,6'dır. FBT faktörünün özdeğeri 8,3; FBK faktörünün özdeğeri 2,1 ve FBB faktörünün özdeğeri ise 1,8 bulunmuştur. Açımlayıcı faktör analizi sonuçlarına dayanarak yol (path) diagramı yardımı ile çizilen ölçme modeli şekil 2'deki gibidir.



Şekil 2: Tüm veri için path diagramı yardımı ile çizilen model

Şekil 2’de yer alan modele ilişkin uyum istatistikleri kabul edilebilir uyum düzeyinde bulunmuştur ($d= 227$, $\chi^2= 4425.62$, $CFI=0,97$, $GFI=0,91$, $RMSEA= 0,069$). Kullanılan veri model ile uyumlu çıkmıştır. Bundan sonraki aşamada şekil 2’deki gibi belirlenen bu ölçme modelinin kız ve erkek öğrenciler için değişmez olup olmadığı test edilmiştir. Test edilen değişmezlik aşamalarına ilişkin bulgular Tablo 1’de mevcuttur. Tabloda yer alan Model A, faktör yükleri, faktör korelasyonları ve hata varyansları serbest olan modeli; Model B, faktör yükleri sabit, faktör korelasyonları ve hata varyansları serbest modeli; Model C, faktör yükleri ve faktör korelasyonları sabit, hata varyansları serbest modeli ve Model D; faktör yükleri, faktör korelasyonları ve hata varyansları sabit modeli temsil etmektedir.

Şekil 2’de yer alan model üzerine yukarıda anlatılan sınırlandırmalar getirildiğinde elde edilen uyum istatistikleri ve değişmezlik testine ilişkin (ΔCFI) uyum katsayılarına ait fark değerleri şu biçimdedir.

Tablo 1: Ölçme değişmezliği aşamalarına ilişkin uyum istatistikleri

Aşamalar	χ^2	d	CFI	GFI	RMSEA	ΔCFI
Model A (Biçimsel Değişmezlik)	2320.32	457	0.97	0.92	0.068	-
Model B (Metrik değişmezlik)	2360.01	480	0.97	0.91	0.067	0,01
Model C (Ölçek değişmezliği)	2968.56	523	0.96	0.91	0.070	0,01
Model D (Katı Değişmezlik)	3081.23	546	0.96	0.91	0.069	0,01

Tabloda yer alan aşamalarda gruplar arası ölçme değişmezliğine karar verebilmek için daha sınırlı olan modeller ile biçimsel model kıyaslanarak, (ΔCFI) uyum katsayılarına ait fark değerleri incelenmiştir. Tablodaki değerler her bir aşama için şu biçimde yorumlanabilir. Biçimsel değişmezliği test etmek üzere yapılan çok gruplu DFA sonucunda uyum indeksleri bu aşamanın sağlandığını göstermektedir. Biçimsel değişmezliğin sağlanmış olması, ölçülen yapıların gruplar arası aynı ve kız ile erkek öğrencilerin ölçek maddelerine yanıt vermede aynı kavramsal bakış açılarını kullandıklarına işaret etmektedir. Metrik değişmezliği test etmek üzere yapılan çok gruplu DFA sonucunda uyum indeksleri ve CFI fark testi sonucunda elde edilen ΔCFI değeri yorumlanmıştır. Bu bulgu, anket maddeleri ile ölçülen özellikler ile epistemolojik inanç boyutları arasındaki ilişkilerin, kız ve erkek öğrenci grupları için benzer olduklarına işaret etmektedir. Ölçek değişmezliğini test etmek üzere yapılan çok gruplu DFA sonucunda uyum indeksleri ile hesaplanan ΔCFI değeri göz önüne alındığında gruplar için maddeler bazında herhangi bir yanlılık bulunmadığı sonucuna ulaşılmıştır. Diğer bir ifadeyle; maddeler için oluşturulan regresyon denklemlerindeki sabit sayının gruplar arası eşit/değişmez olduğu şeklindeki bir hipotez doğrulanmıştır. Son aşamada ise ölçme aracını oluşturan maddelere ilişkin hata terimlerinin karşılaştırma grupları arasında eşit/değişmez olduğu şeklindeki hipotez, yapılan çok gruplu DFA sonucunda uyum indeksleri ile hesaplanan ΔCFI değeri göz önüne alındığında kabul edilmiştir. Elde edilen bu bulgulardan yola çıkılarak AFA sonuçları temel alınarak kurulan modellemeye ilişkin cinsiyet çerçevesinde yapılan tüm karşılaştırmalar anlamlı olacaktır.

4. TARTIŞMA ve SONUÇ

Ölçme değişmezliğinin incelenmesine dayalı olarak bu araştırmada; fen bilimlerine yönelik tasarlanan ölçme modeli çerçevesinde; biçimsel değişmezliğin, metrik değişmezliğin, ölçek değişmezliğinin ve katı değişmezliğinin sağlandığı gözlenmiştir. Bu araştırmada gerçekleştirilen ölçme değişmezliği testlerinin sonuçları, PISA öğrenci anketinde yer alan ve öğrencilerin fen bilimlerine yönelik tutumlarını, bakışlarını, ilgilerin, vb., ortaya koymaya yönelik olan maddelerden oluşan ölçme modelinden elde edilen ölçümlerin psikometrik niteliklerin, cinsiyet grupları arası genellenebileceğini göstermektedir. Bu doğrultuda araştırma sonuçları PISA formunun; cinsiyete göre yanlı davranmaksızın Türk öğrencilerinin bu yapıya ilişkin özelliklerini belirlemede geçerli ve güvenilir ölçümler sağlayabildiğine kanıt sağlar niteliktedir. Ölçmelerin değişmezliğinin test edildiği ölçme modelinin kız ve erkek öğrenci grupları için aynı şekilde ölçüm yaptığını ve dolayısıyla da ölçme aracından elde edilen puanlara dayalı olarak kız ve erkek öğrencilerin belirtilen ölçme modeli çerçevesinde karşılaştırılmasının uygun olacağını göstermektedir.

Bilindiği üzere YEM çalışmaları teori geliştirmek amacı ile de kullanılmaktadır. Teori geliştirilirken ilgilenilen değişkenler arasındaki olası ilişkilerin açıklanmasında benzer veri setleri ile tekrarlanan uygulamalar yapılmaktadır. Ölçme değişmezliği çalışmaları yapılan YEM uygulamaları, teori hakkında önemli ve genellenebilir bilgiler içerecektir. Bu bağlamda tanımlanmaya çalışılan yapıların ortaya konulmasında ölçme değişmezliği çalışmalarının çok önemli bir yerinin olduğunu söylemek yanlış olmayacaktır. Çalışmanın giriş kısmında belirtildiği üzere değişmezlik ölçme araçlarının geçerliği ile yakından ilgilidir. Özellikle büyük ölçekli yürütülen sınavlara ilişkin yapılacak değişmezlik çalışmalarının her biri karşılaştırmaların geçerliği ile ilgili önemli ipuçları vermektedir. Bu bağlamda farklı ölçme araçları için farklı farklı değişkenler işe koşularak değişmezlik çalışmalarının artırılmasının alanyazına katkı sağlayacağı düşünülmektedir.

5. KAYNAKLAR

Akyıldız M. (2009) PIRLS 2001 Testinin yapı geçerliliğinin ülkelerarası karşılaştırılması. *Yüzüncü Yıl Üniversitesi, Eğitim Fakültesi Dergisi*, 6(1), 18-47.

- Bahadır, E. (2012). Uluslararası Öğrenci Değerlendirme Programı'na (PISA 2009) göre Türkiye'deki öğrencilerin okuma becerilerini etkileyen değişkenlerin bölgelere göre incelenmesi. Ankara. Hacettepe Üniversitesi. Yüksek Lisans Tezi.
- Cheung, G. W. & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal Of Cross-Cultural Psychology*, 31(2), 187-212.
- Cheung, G., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 233-255.
- Cheung, G.W. & Rensvold, R.B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, 25, 1-27.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 281-302.
- Kline, R. B. (1998). Principles and Practice of Structural Equation Modeling. New York: The Guilford Press.
- Horn, L. & Mcardle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3), 117 - 144
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32, 53-76.
- Mark, B. A. & Wan, T.T.H (2005). Testing measurement equivalence in a patient satisfaction instrument. *Western Journal of Nursing Research*, 27(6), 772-787.
- Meredith, W. (1993), Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525-543.
- Ployhart, R. E., & Oswald, F. L. (2004). Applications of mean and covariance structure analysis: Integrating correlational and experimental approaches. *Organizational Research Methods*, 7, 27-65.
- Sehee H. & 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). Assessing Measurement Invariance In Cross-National Consumer Research
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington: American Psychological Association.
- Uzun, B & Öğretmen, T. (2010). Fen başarısı ile ilgili bazı değişkenlerin TIMSS-r türkiye örnekleminde cinsiyete göre ölçme değişmezliğinin değerlendirilmesi. *Eğitim ve Bilim* 35(155), 26-35.
- Wicherts, J. M. (2007). Group differences in intelligence test performance. Unpublished dissertation, University of Amsterdam.
- Wu, Amery D. Li, Zhen & Bruno D. Zumbo. (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, and Evaluation*, 12(3).
- PISA (2007). Ulusal ön rapor. [www.earged.meb.gov.tr adresinden indirilmiştir.]

Extended Abstract

Researches are often concerned with comparisons across groups (e.g., genders, cultural groups, racial groups, age groups, occupational groups, etc.) in social science. The purpose of this study is to investigate the factors of PISA student questionnaire's items and assessing the measurement invariance (MI) of these factors across gender in Turkey.

Invariance testing has been developed to address the question of how we can be sure that a test or survey measures the same trait dimension, in the same way, when administered to two or more qualitatively distinct groups. This means that; group comparisons must measure identical concepts across groups to be valid. Through invariance testing we are able to determine the structure of PISA student questionnaire's items were measuring the same trait in the groups. If the intercept of a particular subtest is different across groups, this implies that between-group differences on this subtest cannot be solely due to between-group differences in the construct that the subtest is supposed to measure. In other words, an intercept difference indicates measurement bias in the sense there are one or more construct-irrelevant

variables causing group differences in test scores. This study investigated measurement invariance using confirmatory factor analysis. Most studies stress the importance of modeling the measurement intercepts when assessing measurement invariance across groups (Little, 1997; Ployhart & Oswald, 2004).

This study examined the PISA data for Turkish students with the sample size of 4942 through the analysis of Structural Equation Modeling (SEM). Measurement invariance was tested using Multigroup Confirmatory Factor Analysis (MG-CFA), which examines the change in the comparative-of-fit index (CFI) when cross-group constraints are imposed on a measurement model.

First, it is designed a model over the factors from student questionnaire with Structural equation modeling. We determine 3 factors after the analysis of AFA. It was understood from the lisrel outputs that the designed model's goodness fit index output was good. Then four forms of invariance was assessed for each latent variables progressively and assessed measurement invariance (MI). The analyses basically begin by fitting a proposed model to the data for each sample considered separately with none of the parameters constrained to be equal across groups (configural invariance). Configural invariance is defined as the same pattern of fixed and free factor loadings (and other parameters) across groups, but no equality constraints. This unconstrained model serves as the baseline model. If configural variance is not supported, this is a serious problem. In that case, the evidence argues against even similar factor patterns across groups. Subsequently, in a hierarchical fashion, more stringent constraints are placed on the model by specifying the parameters of interest to be constrained across the groups(e.g., factor loadings, factor intercorrelations, error variances). Then metric (factorial) invariance was assessed. At this time the factor loadings are constrained to be equal across groups, but no other equality constraints are imposed. It widely believed that equal factor loadings are sufficient to establish measurement invariance in confirmatory factor analysis. In scalar invariance tested by specifying factor loadings and intercepts invariant across groups. lastly factor loadings, factor intercorrelations, error variances are constrained to be equal across groups. This condition is also called " Strict factorial invariance." Strict factorial invariance is a highly constrained model and may often not hold in practice. The variables were examined using comparative fit index (CFI) difference test between the more restrictive invariance form and the basic form to determine whether the model and the individual parameter estimates are invariant across the samples. The MI results revealed that there were no invariance problems across gender in PISA for Turkish students.

An assessment of MI is necessary for meaningful between group comparisons in social science research. For any use of scale scores, there is a critical assumption that the scale is measuring the same trait in all of the groups. If that assumption holds, then comparisons and analyses of those scores are acceptable and yield meaningful interpretations. But if that assumption is not true, then such comparisons and analyses do not yield meaningful results. We designed a model over the factors from student questionnaire and tested invariance of these factors of measurement model. So all of the comparisons are meaningful for gender.

The designed model had good fit index. And for gender groups it works same. So the comparisons that we are making and analyses of PISA scores are acceptable and yield meaningful interpretations. Relationships between the indicators and the trait are the same across gender groups for PISA data for Turkish students.

This should be noted that Testing for MI is necessary for meaningful comparisons between groups. Without this essential test, measurement bias may be overlooked. In the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999, henceforth the Standards), test fairness means that "examinees of equal standing with respect to the construct the test is intended to measure should on average earn the same test score, irrespective of group membership"(p.74). Measurement invariance with respect to groups is an essential aspect of the fair use of scores of tests and other psychological measurements.

Kaynakça Bilgisi

Başusta, N. B., & Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: pisa öğrenci anketi örneği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi [Hacettepe University Journal of Education]*, 30(4), 80-90.

Citation Information

Başusta, N. B., & Gelbal, S. (2015). Examination of measurement invariance at groups' comparisons: a study on pisa student questionnaire [in Turkish]. *Hacettepe University Journal of Education [Hacettepe Üniversitesi Eğitim Fakültesi Dergisi]*, 30(4), 80-90.