

**T.C.**  
**AKDENİZ ÜNİVERSİTESİ**  
**EĞİTİM BİLİMLERİ ENSTİTÜSÜ**  
**EĞİTİM BİLİMLERİ ANABİLİM DALI**  
**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME**  
**TEZLİ YÜKSEK LİSANS PROGRAMI**

**PISA 2012 MATEMATİK OKUR-YAZARLIĞI TESTİNİN ÖLÇME**  
**DEĞİŞMEZLİĞİNİN İNCELENMESİ**

**YÜKSEK LİSANS TEZİ**  
**Merve AYVALLI**

**Antalya**  
**2016**

**T.C.  
AKDENİZ ÜNİVERSİTESİ  
EĞİTİM BİLİMLERİ ENSTİTÜSÜ  
EĞİTİM BİLİMLERİ ANABİLİM DALI  
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME  
TEZLİ YÜKSEK LİSANS PROGRAMI**

**PISA 2012 MATEMATİK OKUR-YAZARLIĞI TESTİNİN ÖLÇME  
DEĞİŞMEZLİĞİNİN İNCELENMESİ**

**YÜKSEK LİSANS TEZİ**

**Merve AYVALLI**

**Danışman:**

**Doç. Dr. Bayram BIÇAK**

**Antalya**

**2016**

T.C.  
AKDENİZ ÜNİVERSİTESİ  
EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Merve AYVALLI 'nın bu çalışması 10.08.2016 tarihinde jürimiz tarafından Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Lisans Programında **Yüksek Lisans Tezi** olarak **oy birliği/oy çokluğu** ile kabul edilmiştir

İMZA

**Başkan** : Doç. Dr. Hakan ATILGAN  
Ege Üniversitesi Eğitim Fakültesi  
Eğitim Bilimleri Bölümü



**Üye** : Yrd. Doç. Dr. Güçlü ŞEKERCİOĞLU  
Akdeniz Üniversitesi Eğitim Fakültesi  
Eğitimde Ölçme ve Değerlendirme



**Üye (Danışman)** : Doç. Dr. Bayram BIÇAK  
Akdeniz Üniversitesi Eğitim Fakültesi  
Eğitimde Ölçme ve Değerlendirme



**YÜKSEK LİSANS TEZİNİN ADI: PISA 2012 Matematik Okur-Yazarlığı Testinin Ölçme Değişmezliğinin İncelenmesi**

**ONAY:** Bu tez, Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulunun ..... tarihli ve ..... sayılı kararıyla kabul edilmiştir.

Prof. Dr. Yusuf TEPELİ  
Enstitü Müdürü

## DOĞRULUK BEYANI

Yüksek lisans tezi olarak sunduğum bu çalışmayı, bilimsel etik ve geleneklere aykırı düşecek bir yol ve yardıma başvurmaksızın yazdığımı, yararlandığım araştırmaların kaynakçalarda gösterilenlerden oluştuğunu ve bu kaynakları her kullanımında alıntı yaparak yararlandığımı belirtir; bunu onurumla doğrularım. Enstitü tarafından belli bir zamana bağlı olmaksızın, tezimle ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara katlanacağımı bildiririm.

10/08/2016

Merve AYVALLI

## TEŐEKKÜR

Tez alıŐmamı yürüttüğüm bu zorlu süreçte, hayatımda önemli yerlere sahip ve benim için çok deęerli olan kiŐilerin katkıları olmuŐtur.

Öncelikle alıŐmam süresince, her konuda iyi niyetini ve desteęini hiçbir zaman benden esirgemeyen ve beni her zaman yüreklendiren saygıdeęer danıŐman hocam Do. Dr. Bayram BIAK'a,

alıŐmalarım sırasında bıkmadan, usanmadan tüm sorularımı cevaplayan, bilgi ve deneyimleriyle gelişimime çok büyük katkısı olan ve her konuda bana destek olan deęerli hocam Güçlü ŐEKERCİOęLU'na, ayrıca bilgi ve önerileriyle alıŐmama katkı saęlayan hocalarım Prof. Dr. Hilmi DEMİRKAYA'ya, Do. Dr. Hakan ATILGAN'a, Yrd. Do. Dr. Hakan KOęAR'a,

Her zaman yanımda olan ve bu süreçte beni hiç yalnız bırakmayan sevgili arkadaşlarım AraŐ. Gör. Emine Ela KÖK'e ve Ceren KOCA'ya, yardımlarını benden esirgemeyen sevgili arkadaşlarım AraŐ. Gör. Funda ÖLMEZ'e, AraŐ. Gör. Gizem UYUMAZ'a,

Hayatımın her anında yanımda olan, bana her zaman güvenen ve arkamda duran, beni bugünlere getiren, en büyük destekçilerim canım annem Nebahat AYVALLI'ya, canım babam Mustafa AYVALLI'ya, canım kardeşlerim Mehmet Ali AYVALLI'ya ve Melike AYVALLI'ya çok teşekkür ederim.

## ÖZET

### PISA 2012 MATEMATİK OKUR-YAZARLIĞI TESTİNİN ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ

Ayvallı, Merve

Yüksek Lisans, Eğitim Bilimleri Anabilim Dalı

Tez Yöneticisi: Doç. Dr. Bayram Bıçak

Ağustos 2016, viii+61 Sayfa

Bu araştırmanın amacı PISA 2012 matematik okuryazarlığının OECD üyesi ülkeler ve Türkiye'deki cinsiyet ve bölge grupları arasında ölçme değişmezliğinin incelenmesidir.

İlişkisel tarama modelinde yürütülen araştırmaya PISA 2012 değerlendirmesinde uygulanan bilişsel test kitapçıklarından tüm ülkelerde ortak olarak uygulanmış olan 8 no'lu kitapçık seçilmiştir ve araştırma OECD üyesi olan ülkelerde 8 no'lu kitapçığı almış olan 23.311 öğrenci ve Türkiye'de bu kitapçığı alan 377 öğrenciye ait veri seti üzerinden yürütülmüştür. İlk olarak ülke, cinsiyet ve bölge grupları için ayrı ayrı doğrulayıcı faktör analizi yapılarak ölçme modelleri tüm gruplar için doğrulanmıştır. Daha sonra tüm gruplar için kovaryans matrislerinin eşitliği testi yapılarak ölçme değişmezliğinin test edilmesi aşamasına geçilmiştir. Ölçme değişmezliği çoklu-grup doğrulayıcı faktör analizi ile test edilmiştir.

Araştırma sonuçları Türkiye'deki cinsiyet ve bölge grupları için ölçme değişmezliğinin sağlandığını ancak OECD üyesi ülkeler için ölçme değişmezliğinin sağlanamadığını, ölçme değişmezliğine ilişkin modeller arasında en iyi çalışan modelin güçlü faktöriyel değişmezlik modeli olduğunu göstermiştir.

**Anahtar Sözcükler:** PISA 2012, Matematik okuryazarlığı, Ölçme değişmezliği, Çoklu-grup doğrulayıcı faktör analizi.

## ABSTRACT

### AN INVESTIGATION INTO THE MEASUREMENT INVARIANCE OF PISA 2012 MATHEMATICAL LITERACY TEST

Ayvallı, Merve

Master of Arts, Department of Educational Sciences

Supervisor: Assoc. Prof. Dr. Bayram Bıçak

August 2016, viii+61 pages

The aim of this study is to investigate the measurement invariance of PISA 2012 mathematical literacy among the OECD member countries, and gender and region groups in Turkey.

Among cognitive test booklets implemented in PISA 2012, booklet 8 which was used commonly by all countries was selected for this correlational survey study. The study was conducted using the dataset belonging to 23.311 students that took booklet 8 in the OECD member countries and 377 students that took this booklet in Turkey. Initially, measurement models were verified for all groups by performing a confirmatory factor analysis separately for OECD member countries, gender and region groups. The research then proceeded with the phase of testing the measurement invariance by testing the equivalence of covariance matrices for all groups. The measurement invariance was tested by means of a multi-group confirmatory factor analysis.

The results revealed that the measurement invariance held true for the gender and region groups in Turkey, but not for the OECD member countries, and that the strong factorial invariance model was the one that worked most successfully among the models of measurement invariance.

**Keywords:** PISA 2012, Mathematical literacy, Measurement invariance, Multi-group confirmatory factor analysis.

## İÇİNDEKİLER

İMZA SAYFASI.....	i
DOĞRULUK BEYANI .....	ii
TEŞEKKÜR.....	iii
ÖZET.....	iv
ABSTRACT .....	v
İÇİNDEKİLER .....	vi
TABLolar LİSTESİ.....	viii
BÖLÜM I.....	1
GİRİŞ .....	1
1.1.    Problem Durumu .....	1
1.2.    Araştırmanın Amacı .....	5
1.2.1.    Alt Problemler .....	5
1.3.    Araştırmanın Önemi .....	6
1.4.    Sınırlılıklar .....	7
1.5.    Tanımlar .....	7
1.6.    Kısaltmalar .....	7
BÖLÜM II.....	9
KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR .....	9
2.1. Geçerlilik .....	9
2.2. Faktör Analizi.....	12
2.2.1. Açıklayıcı Faktör Analizi .....	12
2.2.2. Doğrulayıcı Faktör Analizi .....	13
2.3. Ölçme Değişmezliği .....	18
2.3.1. Yapısal Değişmezlik (Baseline Model).....	20



2.3.2. Metrik Değişmezlik (Metric Invariance) .....	21
2.3.3. Güçlü Faktöriyel Değişmezlik (Strong Factorial Invariance) .....	21
2.3.4. Katı Faktöriyel Değişmezlik (Strict Factorial Invariance) .....	21
2.4. Ölçme Değişmezliği Modellerinin Karşılaştırılması.....	22
2.5. İlgili Araştırmalar .....	22
BÖLÜM III .....	26
YÖNTEM.....	26
3.1. Araştırmanın Modeli .....	26
3.2. Evren ve Örneklem.....	26
3.3. Araştırma Verileri.....	30
3.4. Verilerin Analizi.....	31
BÖLÜM IV .....	34
BULGULAR.....	34
4.1. Cinsiyet Grupları için Ölçme Değişmezliğine İlişkin Bulgular .....	34
4.2. Bölge 1, Bölge 2 ve Bölge 3 Grupları için Ölçme Değişmezliğine İlişkin Bulgular .....	37
4.2. OECD Üyesi Ülkeler İçin Ölçme Değişmezliğine İlişkin Bulgular.....	40
BÖLÜM V .....	47
SONUÇ, TARTIŞMA VE ÖNERİLER.....	47
5.1. Sonuç ve Tartışma .....	47
5.2. Öneriler.....	51
KAYNAKÇA.....	52
EKLER.....	60
Ek 1: Özgeçmiş .....	60
Ek 2: İntihal Raporu .....	61

## TABLolar LİSTESİ

Tablo 1. Cinsiyete göre örneklem dağılımı.....	27
Tablo 2. Bölgelere göre örneklem dağılımı .....	28
Tablo 3. OECD üyesi ülkelere göre örneklem dağılımı.....	28
Tablo 4. Analize dâhil edilen matematik okuryazarlığı maddeleri .....	30
Tablo 5. Kız ve Erkek Grupları için Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları.....	34
Tablo 6. Kız ve Erkek Grupları için Kovaryans Matrislerinin Eşitliği Testi Sonuçları .....	35
Tablo 7. Kız ve Erkek Grupları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları .....	35
Tablo 8. Bölgeler için Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları .....	37
Tablo 9. Bölgeler için Kovaryans Matrislerinin Eşitliği Testi Sonuçları.....	38
Tablo 10. Bölgeler için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları .....	38
Tablo 11. OECD Üyesi Ülkeler için Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları.....	41
Tablo 12. OECD Üyesi Ülkeler için Kovaryans Matrislerinin Eşitliği Testi Sonuçları .....	43
Tablo 13. OECD Üyesi Ülkeler Doğrulayıcı Faktör Analizi Sonuçları.....	43
Tablo 14. OECD Üyesi Ülkeler için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları .....	45

# BÖLÜM I

## GİRİŞ

Bu bölümde problem durumu, araştırmanın amacı, alt problemler, araştırmanın önemi, sınırlılıklar ve tanımlara ilişkin bilgilere yer verilmiştir.

### 1.1. Problem Durumu

Küreselleşen dünyada, eğitim politikalarının belirlenmesinde ve yeniden düzenlenmesinde etkili olan ulusal ölçme ve değerlendirme çalışmalarının yanı sıra uluslararası çalışmalar da önem kazanmaktadır. Uluslararası düzeyde ülkelerin eğitim sistemi açısından hangi konumda olduğunun, diğer ülkelerle olan farklılıklarının ve eğitim eksiklerinin belirlenebilmesi için birçok ülke tarafından referans olarak kabul edilen ve uluslararası düzeyde uygulanan sınavlardan biri de Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından yürütülen Uluslararası Öğrenci Değerlendirme Programıdır (PISA-Programme for International Student Assessment). Üç yılda bir yapılan ve 2000 yılından itibaren uygulanmaya başlanan bu değerlendirme çalışmasına, Türkiye 2003 yılından bu yana katılmaktadır. PISA'nın amacı, 15 yaşındaki bireylerin matematik okuryazarlığı, fen okuryazarlığı ve okuma becerileri olmak üzere üç temel alana ilişkin bilgi düzeylerini tespit ederek, okulda edindikleri bu bilgileri gerçek yaşam durumlarına uyarlayıp uyarlayamadıklarını belirleyebilmektir. Üç yılda bir yapılan PISA uygulamalarında, bu üç temel alandan birine ağırlık verilmektedir. PISA 2012 uygulamasında ise, 2003 uygulamasından sonra ilk kez matematik okuryazarlığı alanına odaklanılmıştır. Bu uygulamaya 34'ü OECD üyesi, 31'i OECD üyesi olmayan toplam 65 ülke katılmıştır (Organisation for Economic Co-operation and Development, 2013b). Bilişsel testlerle birlikte okul ve öğrenci anketleri de uygulamada kullanılan ölçme araçlarıdır.

A, B ve C olmak üzere üç farklı formdan oluşan öğrenci anketleri, öğrenciye ait kişisel bilgiler, evi ve ailesine ilişkin bilgiler, problem çözme deneyimleri, matematik öğrenimi ve bilişim teknolojilerine yönelik sorulardan oluşur. Tüm formlar farklı sayıda, farklı madde türlerinin olduğu bir dizi sorudan meydana gelir.

Okul anketi, farklı okul türleri arasındaki benzer ve farklı yönleri tespit etmek amacıyla okul yöneticilerinden, okulun kaynakları, iklimi, yapısı ve organizasyonu gibi okula ilişkin bilgilerin elde edildiği ankettir (Organisation for Economic Co-operation and Development, 2014).

Bilişsel testler ise her bir ülkede 13 ayrı kitapçık olarak uygulanmaktadır. Soruların birbirinden farklı olduğu bu kitapçıklar farklı türlerde test maddelerinden oluşur. Testlerde öğrencilerin genişletilmiş yanıtlar verdiği açık uçlu ve yapılandırılmamış, doğru ya da yanlış olarak yanıtladığı yapılandırılmış ve çoktan seçmeli madde olmak üzere üç farklı madde türü kullanılmıştır.

PISA 2012 uygulamasında matematik okuryazarlığı, bireylerin matematiği çeşitli durumlarda formüle etme, kullanma ve yorumlama kapasitelerini ifade eder. Matematiksel düşünmeyi ve matematiksel kavramları, yöntemleri, unsurları ve araçları kullanarak olayları açıklamayı, tanımlamayı ve tahmin etmeyi içerir. Matematik okuryazarlığı OECD (2013b) tarafından, bireylerin matematiğin dünyadaki rolünü fark etmelerine ve yapıcı, bağlı, yansıtıcı vatandaşların sağlam bir temele dayanan yargı ve kararlar vermelerine yardımcı olur, biçiminde tanımlanmıştır ve durumları matematiksel formüle etme, matematiksel kavram, yöntem, unsur ve araçları matematiksel düşünmeye uygulayabilme ve matematiksel çıktılarını yorumlayabilme süreçlerini içerir. Uygulamada, değişim ve ilişkiler (change and relationship), uzay ve şekil (space and shape), çokluk (quantity), belirsizlik ve veri (uncertainty and data) olmak üzere dört konu alanı ve bu alanlara ilişkin 15 konu başlığı bulunmaktadır. Matematik okuryazarlığı; (a) matematiksel süreçler, (b) temel matematik yetenekleri ve (c) matematiksel içerik olmak üzere üç bağlam çerçevesinde ele alınmaktadır (OECD, 2013b). Elde edilen ölçme sonuçları belirli işlemlerden geçirildikten sonra, her katılımcı ülkeden belirlenen örneklemeler kullanılarak, 6 yeterlilik düzeyi ve bu düzeylere karşılık gelen puan aralıkları belirlenmiştir. Matematik performanslarında ülkelerin puanları 368 ile 613 arasında değişmektedir. OECD üyesi ülkelerin puan ortalaması 494 iken, bütün ülkelerin puan ortalaması 487'dir. Türkiye'nin PISA 2012 uygulamasındaki matematik performansına ait puan ortalaması ise 448'dir ve ikinci düzeyde yer almaktadır. (OECD, 2014).

Türkiye örneklemindeki öğrencilerin matematik okuryazarlığı performanslarını açıklayan değişkenler incelendiğinde, ailenin sosyoekonomik durumu, matematiğe yönelik kaygı düzeyleri, anne-baba eğitim durumu, matematiğe yönelik duyuşsal özellikler (matematiğe yönelik ilgi, motivasyon, öz yeterlik algısı gibi), okula yönelik duyuşsal özellikler (okula yönelik tutum, aidiyet hissi gibi), öğrencilere göre öğretmen özellikleri, öğrenci devamsızlığı, öğrenci özellikleri gibi değişkenlerin olduğu görölmektedir.

Bununla birlikte, öğrencilerin matematik okuryazarlığı performanslarını etkileyen iki temel durum da göze çarpmaktadır. Bunlardan birincisi, kız ve erkek öğrencilerin performanslarının birbirinden farklı olması ve cinsiyet açığı (gender gap) olarak adlandırılan durumdur. Türkiye dâhil olmak üzere, hem OECD üyesi olan hem de olmayan ülkelerin büyük bir çoğunluğunda cinsiyet açığı tespit edilmiştir. Sonuçlar incelendiğinde erkek öğrencilerin performanslarının kız öğrencilerin performanslarından daha yüksek olduğu söylenebilir. İkinci durum ise performansların Türkiye'deki bölgelere göre farklılık göstermesidir. En düşük ve en yüksek ortalamaya sahip bölgeler arasında puan farklılıkları olduğu görölmüştür (Milli Eğitim Bakanlığı, 2015).

PISA sonuçları, birçok katılımcı ülkenin kendi eğitim sistemlerini değerlendirmesinde ve eğitim politikalarının belirlenmesinde oldukça etkili olduğu gibi, farklı grupların ve ülkelerin de karşılaştırıldığı bir uygulamadır. Fakat gruplar arası ölçme işlemlerinde kişilik, zekâ ya da diğer yapılarla ilgili psikometrik ölçme işlemleri yapılırken cinsiyet farklılıkları, bölgesel ya da kültürel farklılıklara sahip gruplar arasında eşit ve karşılaştırılabilir sonuçlar elde edebilmek için, belirli bir psikolojik yapıyı ölçen aracın hedef grup ya da kültürler için de anlamlı hale getirilmesi gerekir. Geçerliliği ve güvenilirliği kanıtlanmış araçlarla yapılan ölçme sonuçlarının farklı gruplarda uygulanabilmesi ve sonuçların karşılaştırılabilmesi, ancak ölçölmek istenen psikolojik özelliğın tüm gruplar tarafından aynı biçimde yorumlanabiliyor olması ile olanaklıdır. Günümüzde farklı özellikleri ölçmek için farklı ölçme araçlarına duyulan gereksinimin giderek artması, farklı kültürlere, bölgelere, şehirlere ya da çeşitli demografik özelliklere dayalı karşılaştırmalara dayanan çalışmaların oldukça fazla yapılıyor olması, ölçme aracının yeni bir evrende kullanılarak elde edilen sonuçların gruplar arası farklılıklardan kaynaklanacak etkiler göz önünde bulundurularak yorumlanabilir hale getirilmesi gerekliliğini de

beraberinde getirmektedir (Mushquash ve Bova, 2007). Bu durum ölçülen psikolojik yapının farklı gruplar arasındaki genellenebilirliğini sağlayabilmek için son derece önemlidir (Brown, 2006).

Farklı gruplarda aynı ölçme aracı kullanılarak yapılan ölçme uygulamalarında “Ölçme aracından kaynaklanan farklılıklar yoktur.” biçimindeki bir kabul, gruplar arası yapılan karşılaştırmaların ve yorumlamaların eksik kalmasına neden olabilir. Eğer bu kabul doğrulanabiliyorsa karşılaştırmalar ve yapılan analizler anlamlı hale gelmektedir. Aksi takdirde, elde edilen sonuçların güvenilirliği ve geçerliliği tehlikeye gireceğinden bulunan sonuçlar anlamlılığını yitirir (Başusta, 2010). Ölçmenin değişmezliği çalışması ölçme modellerinin gruplar arasında veya zamana karşı ne ölçüde genellenebileceğini belirlemek için yürütülür. Eğer bir ölçme aracı heterojen bir gruba uygulanmak üzere geliştirildiyse testin ölçme özelliklerinin alt gruplarda denk olduğunun kanıtlanması gerekmektedir. Ölçmelerin eşdeğerliği ölçme işlemi uygulanan evrendeki bireylere özgü olan özelliklerle ilgili olmaktan çok, ölçme aracının kendisi ile ilgilidir. Ölçme değişmezliği çalışmalarında amaç yeni bir ölçme aracı geliştirmek değildir. Ölçmelerin değişmezliği farklı gruplar arasında karşılaştırma yapabilmemiz için ortaya çıkar (Cheung ve Rensvold, 2000).

Klasik Test Kuramına göre geçerlilik ve güvenilirlik, ölçme araçlarından ziyade ölçme sonuçlarına göre değerlendirilir. Bu yüzden geçerlilik ve güvenilirlik bağlamında hesaplanan madde ve test istatistikleri gruba bağlı olarak hesaplandığından her grubun özelliğine göre farklı sonuçlar gösterebilir (Crocker ve Algina, 1986). Klasik Test Kuramındaki bu sınırlılık ölçme sonuçlarının hangi durumlarda, seviyelerde ve gruplarda geçerli ve genellenebilir olduğu sorunu ortaya çıkarmıştır.

Vandenberg ve Lance’a (2000) göre; bu sınırlılığa bağlı olarak araştırmacıların elde ettiği ölçme sonuçlarını yorumlayabilme ve karşılaştırabilmeleri ile ilgili aşağıdaki sorunlar ortaya çıkmaktadır.

- 1) Farklı kültürlere sahip bireylere ait bir ölçme sonucu kavramsal olarak benzer biçimde yorumlanabilir mi?

- 2) Derecelendirme kaynakları hedef grubun performansını aynı performans ölçütlerine göre derecelendirirken aynı performans tanımını mı kullanmaktadırlar?
- 3) Araçlardaki maddelere benzer biçimde yanıt vermenin önüne geçen cinsiyet, etnik yapı gibi farklılıklar var mıdır?
- 4) Zaman içerisinde önemli olan bir durum veya süreç belirli bir ölçme aracındaki maddelere verilen tepkiye ilişkin bakış açısını, kavramsal çerçevesini değiştirmekte midir?

Bu bağlamda, ölçmenin eşdeğerliği/değişmezliği çalışmalarında yukarıdaki varsayımlar dikkate alınarak araştırma süreci yürütülmektedir. Araştırmalarda kullanılan ölçme aracının farklı gruplarda belirli bir yapıyı ölçerken benzer biçimde çalışacağı ve elde edilen ölçümlerin eşit psikometrik özelliklere sahip olduğu varsayımı kullanılmaktadır. Farklı gruplar arasında karşılaştırma yapılabilmesi için, ön koşul yapılan ölçmelerin değişmezliğinin incelenmesidir denilebilir (Horn ve McArdle, 1992).

Bu doğrultuda araştırmanın problemini PISA 2012 uygulamasındaki matematik okuryazarlığının cinsiyet, bölge ve OECD üyesi ülkeler için ölçme değişmezliğinin test edilmesi oluşturmaktadır.

## **1.2. Araştırmanın Amacı**

Bu araştırmanın amacını PISA 2012 değerlendirmesinde matematik okuryazarlığının OECD üyesi ülkeler ve Türkiye'deki bölge ve cinsiyet grupları için ölçme değişmezliğinin test edilerek, sonuçların gruplar arası karşılaştırılabilir olup olmadığının tartışılmasıdır. Bu amaç doğrultusunda şu sorulara yanıt aranmıştır.

### **1.2.1. Alt Problemler**

- 1) PISA 2012 Türkiye uygulaması sonucunda elde edilen veri seti için matematik okuryazarlığının cinsiyete göre ölçme değişmezliği sağlanmakta mıdır?

- 2) PISA 2012 Türkiye uygulaması sonucunda elde edilen veri seti için matematik okuryazarlığının bölgelere göre ölçme değişmezliği sağlanmakta mıdır?
- 3) PISA 2012 uygulaması sonucunda elde edilen veri seti için matematik okuryazarlığının OECD üyesi olan ülkelere göre ölçme değişmezliği sağlanmakta mıdır?

### **1.3. Araştırmanın Önemi**

Gruplar arası karşılaştırmaların yapıldığı çalışmalarda araştırmacılar genellikle ölçme araçlarının ölçme değişmezliğini/eşitlili sağladığı varsayımıyla söz konusu karşılaştırmaları yaparlar. Fakat bilişsel ve duyuşsal özelliklerin ölçülmesinde, özellikle farklı dil, kültür ve cinsiyete sahip gruplar için ölçme aracının aynı anlamı ifade edip etmediği kanıtlanmadan bu karşılaştırmaları yaparak, sonuçlardan bazı çıkarımlar elde etmek yanlış yorumlamalara neden olabilir (Vandenberg ve Lance, 2000).

PISA ile ilgili yapılan çalışmalar incelendiğinde, çalışmalar daha çok duyuşsal alanların ölçülmesinde kullanılan öğrenci anketleri üzerinden yürütülmüştür, bilişsel testlere ilişkin değişmezlik durumunun incelendiği, ulaşılabilen çalışma sayısı oldukça azdır. Değişmezlik çalışmalarında cinsiyet, bölge ve ülke değişkenleri dâhil edilse bile tüm OECD ülkelerinin dâhil edildiği çalışmalara ulaşılamamıştır. Bilişsel test sonuçlarına göre özellikle cinsiyetler, bölgeler ve OECD ülkeleri arası matematik performanslarında farklılıklar tespit edilmiştir (MEB, 2015). Bu bağlamda araştırmada farklı bireysel özelliklere sahip farklı gruplar arasında ölçme değişmezliği çalışmasının yapılması, geniş ölçekli testler ve bu testlerin yapı geçerliliğine ilişkin kanıt oluşturması ve elde edilen sonuçların doğru yorumlanabilmesi için oldukça önemlidir.



#### 1.4. Sınırlılıklar

Bu araştırma kapsamında,

- 1) PISA 2012 değerlendirmesindeki bilişsel test uygulamalarında yer alan ve tüm ülkelerde uygulanmış olan 8 no'lu kitapçıktaki matematik okuryazarlığına ilişkin 11 çoktan seçmeli madde ele alınmıştır.
- 2) Araştırma 34 OECD üyesi ülke ve bu ülkelerden 8 no'lu kitapçığı alan 23.311 katılımcı ve bu kitapçıkta yer alan 11 çoktan seçmeli madde ile sınırlıdır.

#### 1.5. Tanımlar

**Doğrulayıcı Faktör Analizi:** Daha önceden tanımlanmış bir psikolojik yapıya ait modelin doğrulanıp doğrulanmadığını test eden çok değişkenli bir istatistiktir.

**Çoklu-Grup Doğrulayıcı Faktör Analizi:** Yapısal parametrelerin farklı gruplar için eşitliğini eşzamanlı olarak test eden ve gruplar arasındaki ölçme değişmezliğinin belirlenebildiği doğrulayıcı faktör analizinin özel bir türüdür.

**Ölçme Değişmezliği:** Bir ölçme aracının, ölçme işleminin uygulandığı tüm gruplar için aynı anlama gelmesini ifade eder.

**Matematik Okuryazarlığı:** Matematik bilgisinin gerçek yaşam durumlarında işlevsel olarak kullanılabilmesini ifade eder.

#### 1.6. Kısaltmalar

**OECD:** Organisation for Economic Co-operation and Development

**PISA:** Programme for International Student Assessment

**DFA:** Doğrulayıcı faktör analizi

**ÇG DFA:** Çoklu-grup doğrulayıcı faktör analizi

**YEM:** Yapısal eşitlik modellemesi

**sd:** Serbestlik derecesi

**RMSEA:** Root mean square error of approximation (yaklaşık hataların ortalama karekökü)

**CFI:** Comparative fit index (karşılaştırmalı uyum indeksi)

**NFI:** Normed fit index (normleştirilmiş uyum indeksi)

**NNFI:** Non-normed fit indeks (normleştirilmemiş uyum indeksi)

**GFI:** Goodness of fit (uyum iyiliği indeksi)

**AGFI:** Adjusted goodness of fit (düzeltilmiş uyum iyiliği indeksi)

**SRMR:** Standardized root mean square residual (standardize edilmiş artık ortalamaların karekökü)

## BÖLÜM II

### KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde geçerlilik, faktör analizi, ölçme değişmezliği ve ölçme değişmezliği ile ilgili araştırmalar alan yazın taraması ışığında açıklanacaktır.

#### 2.1. Geçerlilik

Geçerlilik terimi ölçme aracının bilimsel yararını ifade etmektedir. Geniş anlamda ise ölçmeyi amaçladığı şeyi ne kadar iyi ölçebildiği biçiminde belirtilebilir. (Anastasi, 1968; Nunnally ve Bernstein, 1994; Murphy ve Davidshofer, 2005). Kelley (1927); bir test ancak ölçmeyi iddia ettiği özelliği ölçebiliyorsa geçerlidir demiştir. Cronbach (1984) geçerliliği; ölçme aracını geliştiren kişinin, elde ettiği test puanlarının yorumlamasında ve anlamlandırmasında kanıt topladığı bir süreç olarak görmüştür. Messick (1989) ise geçerliliği, farklı değerlendirme biçimleri ya da test sonuçlarını temel alan uygulama ve eylemlerin, yeterliliği ve uygunluğunu destekleyen deneysel kanıtların ve kuramsal gerekçelerin derecesini belirleyen genel bir değerlendirme olarak tanımlamıştır. Sireci'ye (2007) göre ise geçerlilik,

- testin bir özelliğinden çok, belli bir amaç için kullanımını ifade eder,
- bir testin belirli bir amaca uygunluğunu değerlendirebilmek için farklı kaynaklardan edinilen kanıtların olması gerekir,
- bir testin belirli bir amaç için kullanımının savunulur olabilmesi bekleniyorsa, bu amaç için testin kullanımını savunacak yeterli kanıt ortaya koyulmalıdır ve
- bir testin geçerliliğinin değerlendirilmesi durağan bir işlem değil, devam eden bir süreçtir.

Genel olarak geçerlilik, bir ölçme aracının belirli bir yapıya ilişkin ölçülmesi amaçlanan özelliğin, başka etkenler dâhil olmadan ölçülebilme derecesidir. Bu anlamda test geçerliliği ise test sonuçlarının anlamlılık, kullanışlılık ve uygunluk derecesi olarak ifade edilebilir.

1966 yılında Amerikan Psikoloji Birliği tarafından geçerlilik, kapsam geçerliliği, ölçüt dayanaklı geçerlilik ve yapı geçerliliği olmak üzere üç farklı kategori altında toplanmıştır. Messick (1995) ise, ilk kez Cronbach ve Meehl (1955) tarafından ortaya

atılmış olan, tüm geçerlilik türlerinin yapı geçerliliğinin altında birleştirilmiş bir kavram olduğunu ve diğer geçerlilik türlerinin yapı geçerliliğine kanıt oluşturduğu iddiasını desteklemiştir. Borsboom Mellenbergh ve Heerden (2004), birleştirilmiş geçerlilik yaklaşımının, uygulanan testin amaca uygunluğunu tespit etmekte eksik yönleri olduğunu düşündükleri için bu görüşe karşı çıkmışlardır. Geçerliliğin yalnızca yapı geçerliliği çatısı altında kavramsallaştırılmasının, özellikle eğitim ortamlarında kullanılan testlerin amaca uygunluğunu değerlendirmede, kapsam geçerliliğinin önemini zayıflattığı düşünülmektedir (Sireci, 1998; 2007). Murphy ve Davidshofer (2005) günümüzde araştırmacıların genel olarak geçerlilik türlerinin, geçerliliği kanıtlamak için, test sonuçlarını anlamlandırma ve çıkarımda bulunma amacı çerçevesinde farklı geçerlilik stratejileri olduğu konusunda hem fikir olduğunu söylemişlerdir.

Bu bağlamda, oluşturulan bir testin kapsam geçerliliği, maddelerin belirli bir yapı ya da performans alanını yeterince temsil edebilme durumudur (Crocker ve Algina, 1986). Aynı zamanda yordama geçerliliği olarak da bilinen ölçüt dayanaklı geçerlilik, kestirimde bulunmadan önce, bulunma esnasında ve bulunduktan sonra, yordayıcı ve ölçütler arasındaki işlevsel ilişkileri ifade etmektedir (Nunnally ve Bernstein, 1994). Yapı, var olduğunu düşündüğümüz davranışların bazı yönlerini açıklayan bireysel karakterlerdir. Yapı geçerliliği ise, bir veya daha fazla yapıya ilişkin elde edilen değerlendirmede yer alan hangi performansın, ne ölçüde yorumlanabildiğine karar verme sürecidir. Cronbach ve Meehl'e (1955) göre yapı geçerliliği daha çok bir teoriyi inşa ederken ya da varolan teoriyi test ederken kullanılmasına rağmen, yapı geçerliliği aynı zamanda belirli bir yapıya ait ölçme sonuçlarının yorumlanması ve pratik kullanımına ilişkin çıkarımda bulunma durumudur. Linn ve Gronlund'e (1995) göre ise yapı geçerliliği, birçok farklı kaynaktan elde edilen kanıtların birikimine dayalı olmasıyla birlikte testin geliştirme aşamasında da çok önemli bir rol oynar ve, (1) ölçülecek yapının teorik çerçevesine uygun biçimde tanımlanması, (2) yapının teorik temeline uygun olarak değerlendirme hipotezinin oluşturulması, (3) deneysel ve mantıksal yollarla hipotezin doğrulanması süreçlerini içinde barındırır.

Ölçülecek yapılara ilişkin geliştirilen ölçme araçlarının yapı geçerliliğine ilişkin kanıt ortaya koymak için genellikle şu yöntemler kullanılmaktadır.

- a) Gruplar Arası Farklılaşma: Farklı gruplardan elde edilen puan ortalamalarının test edilen hipotez doğrultusunda farklılaşmasının beklenmesi durumudur. Beklenen farkın bulunması durumunda ölçülen yapıya ait geçerlilik kanıtları elde edilmiş olur (Crocker ve Algina, 1986).
- b) Var Olan Testlerle Ölçülen Yapı Arasındaki Korelasyonlar: Yeni oluşturulan bir testin daha önceden mevcut ve geçerliliği kanıtlanmış benzer bir testle olan yüksek korelasyonu, yine yapı geçerliliğine kanıt oluşturan durumlardandır (Anastasi, 1968).
- c) İç Tutarlılık: Bu yöntemin temel özelliği ölçme aracının toplam puanının testin kendisine kriter oluşturmasıdır. Üst kriter grubu ile alt kriter grubunun toplam puanları temel alınarak yapılan karşılaştırmada, üst gruptakiler alt gruptakileri anlamlı bir oranda geçtiyse maddeler geçerlidir biçiminde yorumlanabilir. Bu kanıtlama yöntemiyle ilgili diğer bir uygulama ise alt testlere ait puanların toplam puanla olan ilişkisinin belirlenerek geçerliliğe kanıt oluşturmasıdır (Anastasi, 1968).
- d) Test Puanları Üzerinde Deneysel Değişkenlerin Etkisi: Seçilen değişkenlerin test puanları üzerindeki etkisini araştıran deneylerle yapı geçerliliğinin sağlanmasına ilişkin kanıtlar elde edilir. Ölçülen alana ilişkin belirli bir testin uygulamaya elverişli olup olmamasıyla ilgilidir (Anastasi, 1968).
- e) Yakınsak ve Ayırt Edici Geçerlilik: Campbell ve Fiske (1959), bu yöntemin yapı geçerliliğini değerlendirmede sık kullanılan yöntemlerden biri olduğunu ve birden fazla özellik veya yapıyı ölçmek için birden fazla yöntem kullanılırsa bu ölçümler arasındaki korelasyonların çoklu özellik-çoklu yöntem matrisi halini aldığını söylemişlerdir. Örneğin bir grup öğrencinin dürüstlük ve saldırganlık düzeyleri öğretmen değerlendirmesi, kağıt-kalem testi ve dış gözlemci ile ölçüldüğünde bu özellikler arasındaki ilişkinin her üç yöntemle de ölçülmesi çoklu özellik- çoklu yöntem (multitrait-multimethod matrix) matrisini oluşturur. Bu yöntemle ilgili olarak belirli bir yapıyı ölçen iyi bir testin (i) test sonuçlarının aynı yapıya ait diğer ölçümleri kullanarak elde edilen sonuçlarla tutarlı olması, (ii) testin, ölçülen yapıyla kuramsal açıdan ilgisi olmayan ölçümlerle ilişkisiz sonuçlar vermesi, (iii) testle

kullanılan ölçme yöntemlerinin yanlılık için kanıt oluşturması çoklu özellik-çoklu yöntem matrisi yönteminin ileri sürdüğü karakterlerdir.

- f) Faktör Analizi: Ölçme araçlarının yapı geçerliliğine ilişkin güçlü kanıtlar elde etmemize yardımcı olan ve en sık kullanılan yöntemlerden biri olan faktör analizine, yapılan çalışmanın temelini oluşturduğu için aşağıda ayrıntılı biçimde yer verilmiştir.

## **2.2. Faktör Analizi**

Psikolojik özelliklerin belirlenmesi için kullanılan çok değişkenli bir istatistik olan faktör analizinin asıl amacı gözlenen değişkenler arasındaki kovaryans ya da korelasyonlara dayalı olarak örtük değişken ya da değişkenlerin sayısı ve türlerini belirlemektir (Brown, 2006). Büyüköztürk (2002) ise faktör analizini aynı yapıya ait ve birbirleriyle ilişkili değişkenlerin bir araya gelerek, kavramsal açıdan daha az sayıda değişkenler olarak tanımlanmasına yardımcı olan çok değişkenli istatistiksel bir yöntem olarak tanımlanmıştır. Crocker ve Algina'ya (1986) göre faktör analizinin üç temel amacı bulunmaktadır. Bunlar (1) bir modele ait bir dizi testin tüm çiftleri arasındaki korelasyonları için hesaplanması gereken ortak faktörlerin sayısına karar vermek, (2) test iç korelasyonları için hesaplanan ortak faktörlerin niteliğine karar vermek ve (3) ortak faktör varyansı ile ilişkili bir gözlenen değişken için varyans oranını belirlemektir. Faktör analizinin açıklayıcı ve doğrulayıcı olmak üzere iki farklı türü vardır. Hem açıklayıcı hem de doğrulayıcı faktör analizi bir grup gizil değişken ile göstergeler arasındaki gözlenen ilişkiyi yeniden ortaya koymayı amaçlar. Fakat bu analizler temelde, faktör modeli üzerindeki kısıtlamalar ve öncül şartların sayısı ve niteliği konusunda farklılık gösterir (Brown, 2006).

### **2.2.1. Açıklayıcı Faktör Analizi**

Açıklayıcı faktör analizi, belirli bir yapıya ait gözlenen değişkenlerle gizil değişkenlerin arasındaki ilişkiyi ortaya koymayı amaçlayan istatistiksel bir tekniktir (Brown, 2006). Açıklayıcı faktör analizinin psikolojik yapıların ölçülmesinde, açıklama ve veri indirgeme olmak üzere iki genel kullanım alanı vardır. Bunlardan birinci kullanım alanı olan açıklama, belirli bir ölçme aracı ile belirli bir alana ilişkin boyutların belirlenmesini ve gizil değişkenlerin keşfedilmesini sağlar. Bu amaca ulaşmak için, ölçülen değişkenler arasındaki korelasyon ya da kovaryanslar matrisini

kullanır. İkinci kullanım alanı olan veri indirgemedede ise ölçülen değişkenler kümesini özetlenmiş indeksler olarak birleştirir ve temel bileşenler analizi kullanılarak yapılır (Floyd ve Widaman, 1995).

Açımlayıcı faktör analizi,

- a) bir değişken setini seçme ve bu seti ölçme,
- b) seçilen veri setine ait korelasyon matrisini hazırlama,
- c) hazırlanan korelasyon matrisinden belirli faktörler çıkarma,
- d) faktör sayısını tespit etme,
- e) analiz sonuçlarının yorumlanmasını kolaylaştırmak için faktörlere döndürme işlemi uygulama ve
- f) analiz sonucunda elde edilen verileri yorumlama adımlarını içerir (Tabachnick ve Fidell, 2013).

### **2.2.2. Doğrulayıcı Faktör Analizi**

Yapısal eşitlik modelinin özel bir türü olan doğrulayıcı faktör analizi, gözlenen değişkenler ile gizil değişkenler arasındaki ilişkileri test eden kapsamlı, nedensel bir analizdir (Yılmaz, 2004). Doğrulayıcı faktör analizinde, güçlü deneysel ya da kavramsal alt yapıya sahip olan, önceden yapılandırılmış bir modeldeki faktör yükleri, faktör sayıları vb. gibi özelliklere bakılır. Bu özelliklerden yola çıkılarak kurgulanan modelin yapı geçerliliği test edilmiş olur (Maruyama, 1998). Bir modele ait doğrulayıcı faktör analizi yapılmadan önce, (1) değişkenlere ait faktörlerin sayısı, (2) gözlenen değişkenlere ait yüklerin, belirlenmiş olan faktörlerden hangilerinin altında toplanacağı, (3) hangi faktör çiftlerinin birbirleriyle ilişkili olduğu bellidir (Brown, 2006).

Kline'a (2005) göre doğrulayıcı faktör analizi,

- a) modelin betimlenmesi,
- b) modelin tanımlanması için koşulların uygunluğunun belirlenmesi,
- c) model içinde gösterilen değişkenler doğrultusunda verilerin toplanması,
- d) modelin analizi,
- e) veri seti ile modelin uyumunun değerlendirilmesi,
- f) modelin iyi uyum göstermediği durumlarda modelin yeniden tanımlanması ve
- g) tartışma ve yorumlama adımlarını içerir.

Betimlenen modele ilişkin uyumun değerlendirilmesinde ise bazı uyum indekslerinden faydalanılır. Bunlar mutlak uyum indeksleri ve artan (fazlalık) uyum indeksleri olmak üzere ikiye ayrılır.

Mutlak uyum indeksleri, önsel modelin, örneklem modele ne kadar iyi biçimde uyduğunu belirleyerek (McDonald ve Ho, 2002) önerilen modellerden en üst düzey uyuma sahip olanı gösterir. Bu ölçümler önerilen teorinin veriye ne kadar iyi uyduğunun en temel göstergesini elde etmeyi sağlar. Artan uyum indekslerinden farklı olarak, mutlak uyum indekslerinin hesaplanması temel (baseline) bir modelle karşılaştırmaya dayandırılmaz, bunun yerine, hiçbir modelle karşılaştırılmadan modelin ne kadar uyumlu olduğu ölçülür (Jöreskog ve Sörbom, 1993).

Aynı zamanda karşılaştırmalı (Miles ve Shevlin, 2007) veya görelî (bağıl/nispi) uyum indeksleri olarak da bilinen artan uyum indeksleri ise temel formunda ki-kare modelini kullanmayan, ancak ki-kare değerini temel model ile kıyaslayan bir grup indekstir (McDonald ve Ho, 2002).

Bu çerçevede araştırmada Ki-Kare uyum iyiliği testi, GFI, AGFI, RMR, SRMR, NFI, CFI ve RMSEA uyum indeksleri açıklanacaktır.

Ki-kare değeri ( $\chi^2$ ): Model uyumunun tamamını değerlendirmek için kullanılan geleneksel bir ölçümdür ve örneklem ile uyumlu kovaryans matrisleri arasındaki uyumsuzluk boyutunu inceler (Hu ve Bentler, 1999: 2). İyi bir model uyumu, .05 düzeyinde manidar olmayan bir sonuç verir (Barrett, 2007), böylece ki-kare istatistik testi çoğu zaman “kötülük uyumu” (Kline, 2005) veya “uyum eksikliği” (Mulaik ve diğerleri, 1989) olarak tanımlanır. Ki-kare testi uyum istatistiğinin kullanımına ilişkin bazı kısıtlamalar mevcuttur. Bu kısıtlamalardan ilki çok değişkenli normallik varsayımıdır ve normallikten doğan ciddi sapmalar, model uygun biçimde belirlense bile modelin reddedilmesine yol açabilir (McIntosh, 2006). Bir diğer sebep de ki-kare istatistiği, temelde istatistiksel manidarlık testi olduğundan, örneklem büyüklüğüne karşı duyarlıdır. Bu bağlamda ki-kare istatistiği büyük örneklem kullanıldığı takdirde çoğunlukla modeli reddeder (Bentler ve Bonnet, 1980; Jöreskog ve Sörbom, 1993).

Diğer yandan, küçük örneklem kullanıldığında ki-kare istatistiği uyumu düşüren bir etki yarattığından, iyi uyumlu modeller ile zayıf uyumlu modeller arasında ayrımı



yapamayabilir (Kenny ve McCoach, 2003). Ki-kare istatistiğinin kısıtlayıcı olması nedeniyle, model uyumlarını değerlendirmek için araştırmacılar tarafından alternatif indeksler önerilmiştir. Wheaton, Muthen Alwin ve Summers (1977) tarafından örneklem büyüklüğünün etkisini azaltan normlaştırılmış ki-kare istatistiği ( $\chi^2 /sd$ ) önerilmiş ve bu değer kabul edilebilir düzeyinin en fazla 5, en az 2 (Tabachnick ve Fidell, 2013) olabileceği belirtilmiştir.

RMSEA: Modelin bilinmeyen ancak en ideali olan parametre tahminleriyle birlikte evrenlerin kovaryans matrisine ne kadar uyduğunu gösterir (Byrne, 1998). RMSEA istatistiği için .06'ya yakın bir kesme noktası iyi uyum olarak değerlendirilirken (Hu ve Bentler, 1999) .05 ile 0 arasındaki değerler modelin mükemmel uyum gösterdiği anlamına gelir (Jöreskog ve Sörbom, 1993; Brown, 2006; Raykov ve Marcoulides, 2008).

GFI: Jöreskog ve Sörbom tarafından geliştirilmiştir, tahmini evren kovaryansı ile açıklanan varyans oranını hesaplar (Tabachnick ve Fidell, 2013). GFI 0 ile 1 arasında değerler alır ve örneklem büyüdükçe değeri artar (Bollen, 1990; Miles ve Shevlin, 1998). Genel olarak GFI için kesme noktası olarak .90 önerilse de, simülasyon çalışmaları, faktör yükleri ve örneklem büyüklüğü küçük olduğunda, .95 değerindeki kesme noktasının daha uygun olduğunu göstermiştir (Miles ve Shevlin, 1998). Bununla birlikte AGFI da örneklem büyüklüğünün artmasıyla birlikte artış gösterir ve 0 ile 1 arasında değerler alır. AGFI, .90 ve üzerinde değer alıyorsa bu modelin mükemmel uyum gösterdiğini işaret eder.

RMR ve SRMR: Örneklem kovaryans matrisi ve varsayılan kovaryans modelinin artık değerleri arasındaki farkın kareköküdür. RMR aralığı her bir göstergenin değer ölçeğine bağlıdır. Örneğin bir ölçek, değişik düzeyde maddeler içeriyorsa (bazı maddeler 1 ile 5 arasında değişirken, diğerleri 1 ile 7 arasında değişiklik gösteriyorsa), RMR'nin yorumlanması zorlaşır (Kline, 2005). Standardize edilmiş RMR (SRMR) bu sorunu çözdüğü için yorumlanması daha kolaydır. SRMR değerleri 0 ile 1 arasında değişir ve .05'ten küçük değerler model uyumunun iyi olduğu anlamına gelir (Byrne, 2013). Bununla birlikte .08'e kadar olan değerler de kabul edilebilir düzeydedir (Hu ve Bentler, 1999). SRMR'nin 0 değerini alması mükemmel uyumu işaret eder ancak SRMR modelde çok sayıda parametre

olduğunda ve büyük örneklem büyüklüklerine sahip modellerde düşük değer göstermeye eğilimlidir.

NFI: Modelin  $\chi^2$  değerini, yokluk hipoteziyle kurulmuş modelin  $\chi^2$  değeriyle karşılaştırarak ölçer. Yokluk hipotezine dayalı oluşturulmuş model, ölçülmüş tüm değişkenlerin korelasyonsuz olduğunu belirttiği için en kötü durum senaryosu olduğu söylenebilir. Bu istatistik için değerler 0 ile 1 bir arasında yer alır. Bentler ve Bonnet (1980) .90'dan büyük değere sahip modellerin iyi uyum sergilediğini söylemişlerdir. Fakat bazı araştırmacılar kesme noktası kriterinin .95'den büyük olması gerektiğini ileri sürmektedirler (Hu ve Bentler, 1999). Bu uyum indeksi örneklem büyüklüğüne karşı duyarlılık gösterir (Mulaik, James, Van Alstine, Bennet, Lind ve Stilwell, 1989; Bentler, 1990) ve bu nedenle bu indekse tek başına güvenilmemesi gerekir (Kline, 2005). Bu sorun, daha basit modelleri tercih eden Normlaştırılmamış Uyum İndeksi (NNFI, aynı zamanda Tucker-Lewis indeksi olarak da bilinir) ile iyileştirilmiştir. Bununla birlikte, küçük örneklemelerin kullanıldığı durumlarda, diğer istatistikler iyi uyum değerleri gösterirken, NNFI zayıf uyum değeri gösterebilir (Bentler, 1990; Kline, 2005; Tabachnick ve Fidell, 2007). Bentler ve Hu (1999), NNFI için kesme noktasının .95 ve üzerinde olması gerektiği önerisinde bulunmuşlardır.

CFI: Örneklem boyutu küçükken dahi iyi sonuç veren, örneklem boyutunu göz önünde bulunduran NFI'nın (Byrne, 1998) revize edilmiş halidir (Tabachnick ve Fidell, 2007). NFI gibi, bu istatistik de tüm örtük değişkenlerin korelasyonsuz olduğunu varsayar ve örneklem kovaryansını, yokluk hipotezine dayalı olarak oluşturulan model ile kıyaslar. NFI'da olduğu gibi, bu istatistik için de değerler 0 ve 1 arasında değişkenlik gösterir ve değerler 1'e yaklaştıkça model uyumu artar. Yapılan çalışmalara göre .95'den büyük değer mükemmel uyum değeri olarak kabul edilmektedir (Hu ve Bentler, 1999). Günümüzde bu indeks tüm YEM programlarına dâhil edilmiş olup, örneklem boyutundan en az etkilenen uyum indekslerinden biri olduğundan en çok rapor edilen uyum indekslerinden biri olmuştur (Fan, Thompson, ve Wang, 1999).

Uyum İndekslerini Raporlaştırma: McDonald ve Ho (2000) tarafından yapılan bir araştırmada, en çok rapor edilen uyum indekslerinin CFI, GFI, NFI ve NNFI olduğu görülmüştür. Ki-kare istatistiğine bakılarak yapılan değerlendirmeler örneklem büyüklüğünden etkilendiği için, bu istatistiğin, serbestlik derecesi ve ona bağlı p

değeriyle birlikte her zaman kullanılması gerektiği söylenmektedir (Kline, 2005; Hayduk, Cummings, Boadu, Pazderka-Robinson ve Boulianne, 2007). Bunun yanında, örneklem büyüklüğünün çok olduğu ve normal dağılımdan sapma gösteren durumlarda, normallik gösteren ve küçük örnekleme sahip durumlarda hesaplanan  $\chi^2$  değerine yakın değerler üreten Satorra-Bentler istatistiği ( $S-B_{\chi^2}$ ) hesaplanır.  $S-B_{\chi^2}$  düzeltmesi farklı örneklem büyüklüğüne ve puan dağılımına sahip veri setlerinde, örneklem büyüklüğünün  $\chi^2$  üzerindeki etkisini azaltmak için kullanılan güvenilir bir istatistiktir (Byrne, 2013).

Hu ve Bentler (1999) ise her zaman için SRMR kullanımını, NNFI (TLI), RMSEA veya CFI ile birlikte olması gerektiğini ifade etmişlerdir. Kline (2005), raporlaştırmada kesinlikle ki-kare istatistiğinin kullanılması gerektiğini söylemiştir. Genel olarak incelendiğinde, yapılan araştırmalara dayanarak çalışmalarda ki-kare istatistiğini ki-kare serbestlik derecesini ve p değerini, RMSEA, SRMR, CFI ve NNFI gibi uyum indeksleri örneklem boyutu, modelin yanlış belirlenmesi ve parametre tahminleri açısından en az duyarlılığa sahip oldukları için modellerin değerlendirilmesinde kullanılabilir.

### **2.2.2.1. Çoklu Grup Doğrulayıcı Faktör Analizi**

Doğrulayıcı faktör analizi belirli bir teoriye dayalı olarak oluşturulan bir modelin ya da psikolojik bir yapının geçerliliğini kanıtlamak için yapılan bir analizdir. Aynı zamanda bir ölçme aracının, alt grupları olan bir evrene uygulanması durumunda, söz konusu psikolojik yapının bu alt gruplarda da değişmez olduğunu kanıtlamak için yine doğrulayıcı faktör analizinin özel bir türü olan çoklu grup doğrulayıcı faktör analizi yapılır. Çoklu grup doğrulayıcı faktör analizi en az iki grupta eş zamanlı olarak yapılır ve bu analiz ölçme değişmezliğinin incelenerek, adı söz konusu psikolojik yapıların farklı gruplar arasında genellenebilirliğini belirleyebilmek için de oldukça önem taşır. Çoklu-grup doğrulayıcı faktör analizinde oluşturulan ölçme modeline ait parametreler analize dâhil edilen bütün gruplarda eş zamanlı olarak

kestirilerek, bu parametrelerin birbirlerinden anlamlı bir biçimde farklılaşım farklılaşmadığı test edilir (Brown, 2006).

### 2.3. Ölçme Değişmezliği

Herdman'a (1998) göre ölçme değişmezliği bir psikolojik ölçme aracının güvenilirlik ve yapı geçerliliği gibi psikometrik niteliklerinin farklı gruplardaki eşitliğinin biçimsel olarak değerlendirilmesi şeklinde tanımlanır. Byrne ve Watkins (2003) ise ölçme değişmezliğini tüm gruplarda, ölçme aracındaki maddelerin tamamıyla aynı biçimde algılanıp yorumlanması şeklinde tanımlamaktadırlar. Ölçme değişmezliğinin temelinde, karşılaştırma yapılan gruplar için yapılan ölçümlerin geçerliliğinin sağlanması yatar (Tyson, 2004).

Araştırmalar incelendiğinde ölçme değişmezliği çalışmalarının özellikle psikolojik yapıların kültürler arası genellenebilirliğini incelemek ve bu yapıların karşılaştırılabilirliğini belirlemek için yapıldığı görülür. Bunlara ek olarak, aynı kültüre sahip gruplarda cinsiyet, yaş, etnik köken gibi alt gruplar için yapılan karşılaştırmalarda da ölçme değişmezliğinin test edildiği görülmektedir.

Kültürler arası karşılaştırma çalışmalarının hem psikometrik hem de kavramsal bir tarafı vardır. Bu bağlamda ölçme değişmezliği incelenirken test yanlılığı ve test eşitliği gibi kavramsal ve yöntemsel durumlardan da bahsedilebilir.

**Test Yanlılığı:** Test yanlılığı farklı kültürlerde yapılan uygulamalarda geçerliliği tehdit eden bir unsurdur. Yanlılık bir özellik ya da yeteneğin altında yatan farklılıklardan değil, belli bir yapının göstergeleri üzerindeki puan farklılıklarından meydana gelir (Van de Vijver ve Tanzer, 2004). Yanlılık yapı yanlılığı, yöntem yanlılığı ve madde yanlılığı (değişen madde fonksiyonu) olmak üzere üç çeşittir.

**Yapı yanlılığı** ölçülen özelliğin psikolojik yapısının gruplar arasındaki farklılığıyla ilgilidir. Psikolojik yapıların farklı gruplarda farklı anlamlandırılmasından kaynaklanır.

**Yöntem yanlılığı** örneklem arası karşılaştırma yapılamamasından kaynaklanan, ölçme aracının yapısal özellikleri, aracının karakteristik özellikleri ve aracı cevaplayanların cevaplama esnasında kültür kaynaklı nedenlerden dolayı farklı

eğilimler göstermesinden kaynaklanan ve uygulama sürecinin, uygulama koşullarının farklılık göstermesinden meydana gelen sorunların genel halidir. Yöntemden kaynaklı yanlılıklar test sonuçlarını ve sonuçların yorumlanmasını oldukça etkiler.

Madde yanlılığı (değişen madde fonksiyonu) testte yer alan maddelerin farklı gruplar arasında farklı anlamlara gelmesinden ve farklı biçimde yorumlanmasından kaynaklanır (Matsumoto ve Van de Vijver, 2012).

#### Yanlılık ve Değişmezlik

Ölçme yanlılığı ve ölçme değişmezliği arasındaki ayrımı yapabilmek önemlidir. Bu iki terim birbirleriyle eş anlamlı değildir. Yanlılık gerçek parametrelerle tahmin edilen parametre arasındaki farktır. Yanlılık, yapı, yöntem ya da madde yanlılığı nedeniyle istenmeyen bir faktöre ait üretilen istenmeyen bir ölçme varyansı kaynağıdır (Van de Vijver ve Poortinga, 2005, pp. 41-49). Ölçme değişmezliği ise elde edilen puanların gruplar arası karşılaştırılabilirliğini ve değişkenler arasında maddelerin eşit bir biçimde katkısının olup olmadığını (özellikle faktör yük değerleri üzerinde) açıklamaya çalışır (Meredith, 1993).

#### Test Eşitliği

Test eşitliği farklı ölçme işlemleri sonucu elde edilen puanların farklı gruplarda farklı zamanlarda karşılaştırılabilir olmasını sağlayan bir kavramdır. Yani test sonuçlarının farklı zamanlarda ve farklı kültürlerde olsa dahi, aynı anlama gelmesi demektir. Van de Vijver ve Leung (1997), yapısal eşitlik, ölçme birimi eşitliği ve tam puan eşdeğerliği olmak üzere test eşitliği kavramını üç gruba ayırmışlardır.

Yapısal eşitlik, ölçülen özelliğin kültürler arasındaki teorik eşitliğinin sağlanabilmesiyle ilgilidir.

Ölçme birimi eşitliği, ölçülen özelliğin eşit aralıklı ölçek ya da oran ölçeği gibi aynı ölçek düzeyinde ölçülmesi demektir.

Tam puan eşdeğerliğinde ise ölçme aracı aynı ölçme birimine ve başlangıç noktasına sahiptir. Bu eşitliğe en üst düzey eşitlikte denebilir. Bu eşitlik sağlandığında kültürler arasında karşılaştırma ve birçok istatistiksel analiz yapılabilir (Matsumoto ve Van de Vijver, 2012).

Tüm bunlar dikkate alınarak ölçmelerin kusursuz bir biçimde sağlanabilmesi yönündeki çalışmaların yapılması gerekir. Ölçme değişmezliğinde deneysel olarak, gizil değişken puanları ile gözlenen puanları ilişkilendirerek bir ölçme modeli oluşturulur ve bu model üzerinden incelenebilir (Vandenberg ve Lance, 2000). Kurulan bir ölçme modeline ait yapının farklı gruplar için aynı olması, kullanılan ölçme aracındaki maddelerin faktör yüklerinin, faktör korelasyonlarının, hata varyanslarının aynı olması anlamına gelir (Bollen, 1989; Byrne, 1998; Jöreskog ve Sörbom, 1993).

Ölçme değişmezliğini incelemek için kurulan modellerin test edilmesinde uygulanan adımlar ise şöyledir (Kline, 2005; Jöreskog ve Sörbom, 1993),

- a) Faktör yapılarının ve maddelere ait parametrelerin bütün gruplardaki eşitliğinin test edilmesi,
- b) Bütün gruplara ait hata varyanslarının eşit, faktör yüklerinin farklı olduğu durumun test edilmesi,
- c) Bütün gruplarda hata varyanslarının ve faktör yüklerinin farklı olduğu durumun test edilmesi,
- d) Bütün gruplara ait faktör yüklerinin eşit, hata varyanslarının farklı olduğu durumun test edilmesi.

Bu çerçevede ölçme değişmezliğinin test edilmesi için çoğunlukla kullanılan modeller yapısal değişmezlik (Model 1), zayıf faktöriyel değişmezlik (Model 2), güçlü faktöriyel değişmezlik (Model 3) ve katı faktöriyel değişmezlik (Model 4) (Cheung ve Rensvold, 2002) aşağıdaki gibi özetlenmiştir.

### **2.3.1. Yapısal Değişmezlik (Baseline Model)**

Temel model olarak adlandırılan bu modelde “grupların faktör yapıları arasında farklılık yoktur” hipotezi test edilir. Tüm gruplar için faktör korelasyonları, faktör yükleri ve hata varyansları eşittir. Yapısal değişmezliğin sağlanması durumunda ölçülen yapının ve bu yapıya ait maddelerin gruplar arasında aynı anlama geldiği ifade edilir (Gregorich, 2006). Bu durum aracının yapı geçerliliğinin sağlanması demektir (Salzberger, Sinkovics, ve Schlgelmich, 1999; Spini, 2003; Vandenberg ve Lance, 2000).

### **2.3.2. Metrik Değişmezlik (Metric Invariance)**

Meredith'e (1993) göre zayıf faktöriyel değişmezlik (weak factorial invariance) olarak da adlandırılan bu değişmezlik türünde, gözlenen puanları ilgili gizil değişkene bağlayan regresyon eğilimleri olarak adlandırılan faktör yüklerinin gruplar arasında değişmez olup olmadığı test edilir (Gregorich, 2006). Bu durumda karşılaştırılan gruplar arasındaki ölçme birimlerinin eşit olduğu söylenebilir (Salzberger ve diğerleri, 1999). Değişmezlik test edilirken faktör korelasyonları ve hata varyansları sabit, faktör yükleri serbesttir. Eğer metrik değişmezliğe ait kurulan hipotez kabul ediliyorsa tahmin edilen faktör varyans ve kovaryansları nicel grup karşılaştırılmalarında savunulabilir (Gregorich, 2006). Aynı zamanda faktör yüklerinin eşit olması maddelerin tüm gruplarda aynı biçimde yorumlanması ve herhangi bir yanlılık oluşturmaması biçiminde de yorumlanmaktadır (Knight ve Hill, 1998).

### **2.3.3. Güçlü Faktöriyel Değişmezlik (Strong Factorial Invariance)**

Ölçek değişmezliği olarak da adlandırılan bu değişmezlik türü ölçme değişmezliğinin güçlü bir düzeyidir. Ölçek değişmezliğinde faktör yükleri ve hata varyansları serbest, faktör korelasyonları sabittir. Bu değişmezlik türünde faktör yüklerinin eşit olmasının yanı sıra madde sabitlerinin de eşit olması gerekmektedir. Madde sabiti, örtük değişkenin sıfır olduğu durumlarda gözlenen değişkenin aldığı değerdir. Bu değişmezlik türü sağlanırsa "gözlenen değişkenlerdeki ortalama farklılıkları, örtük yapıların ortalamalarından kaynaklanmaktadır" biçiminde yorum yapılabilir (Tucker, Ozer, Lyubomirsk ve Boehm, 2006). Gruplar arasında madde sabiti farklarının olup olmadığının tespit edilmiş olması çok güçlü bir geçerlilik kanıtı olarak kabul edilir (Vandenberg ve Lance, 2000). Gregorich'e (2006) göre güçlü faktöriyel değişmezliğin sağlanması durumunda, faktör ve gözlenen değişkenlerin ortalamaları karşılaştırılabilir.

### **2.3.4. Katı Faktöriyel Değişmezlik (Strict Factorial Invariance)**

Katı faktöriyel değişmezliğin sağlanması durumunda faktör ve gözlenen değişken ortalamalarıyla birlikte gözlenen varyans ve kovaryansların karşılaştırılması da mümkündür (Gregorich, 2006). Bu değişmezlik türünde faktör korelasyonları ve faktör yükleri sabit, hata varyansları serbest durumdadır. Katı faktöriyel

değişmezlikte faktör yükleri, faktör yapısı ve madde sabitlerinin eşit olmasıyla birlikte madde artık varyanslarının da aynı zamanda eşit olması beklenir (Widaman ve Rice, 1997).

#### **2.4. Ölçme Değişmezliği Modellerinin Karşılaştırılması**

Yapılan çalışmalar incelendiğinde, ölçme değişmezliği çalışmalarında kurulan ölçme modellerinin karşılaştırılmasında farklı yaklaşımlar kullanılarak karşılaştırmalar yapıldığı görülmektedir. Araştırmada ÇGDFA sonuçları modellerin aşamalı olarak karşılaştırıldığı yuvalanmış (nested) yöntem (Brown, 2006) kullanılarak karşılaştırılmıştır. Modeller sırasıyla önce Model 1 (yapısal değişmezlik) ve Model 2 (metrik değişmezlik), Model 2 ve Model 3 (güçlü faktöriyel değişmezlik) ve son olarak Model 3 ve Model 4 (katı faktöriyel değişmezlik) karşılaştırılarak sonuçlar değerlendirilmiş ve ölçme değişmezliğine ilişkin karar alınmıştır. Model uyumu karşılaştırılan modelin uyumuna eşit uyum göstermişse ya da daha kötü uyum göstermişse daha çok sınırlandırılmış ölçme modelinin doğrulandığı kabul edilmiştir (Van de Vijver ve Leung, 1997).

#### **2.5. İlgili Araştırmalar**

Bu bölümde ilk olarak uluslararası düzeyde uygulanan geniş ölçekli sınavların ölçme değişmezliği ile ilgili çalışmalara daha sonra da ulusal düzeyde yapılmış ölçme değişmezliği çalışmalarına yer verilmiştir.

Ercikan ve Koh (2005) 1995 TIMSS Kanada uygulamasına katılan öğrencilerin dâhil olduğu, Fransızca ve İngilizce formlarının değişmezliğini araştırdıkları çalışmalarında, ölçme değişmezliğini ÇGDFA ile incelemiş ve maddelerin gruplara göre değişen madde fonksiyonu gösterip göstermediklerini belirlemeye çalışmışlardır. Araştırma sonucunda, TIMSS matematik ve fen bilimleri alanlarının her ikisinde ölçme değişmezliğinin sağlanmadığını tespit etmişlerdir. Ayrıca maddelerin, yine matematik ve fen bilimleri alanlarında gruplar arasında farklı fonksiyon gösterdiği belirlenmiştir. Elde edilen bu sonuçlar, farklı ülkeler, kültürler ve dillerde yapılan uygulamalarda gruplar arası karşılaştırmalar yapılmadan önce ölçme değişmezliği çalışmalarının yapılması gerekliliğini ortaya koymaktadır.



Marsh ve arkadaşları (2006), 2000 PISA uygulamasında kullanılan öğrencilerin öğrenme yaklaşımları ölçme aracının 25 ülkedeki ölçme değişmezliğini inceledikleri çalışmalarında bu ölçeğe ait faktör yapısının doğrulandığını ve söz konusu 25 ülke için ölçme değişmezliğinin sağlandığını tespit etmişlerdir.

Wu, Li ve Zumbo (2007), 1999 TIMSS uygulamasında matematik alanına ait testin benzer ve farklı kültürlere sahip ülkeler arasında ölçme değişmezliğini inceledikleri çalışmalarında, benzer kültürlere sahip olan ABD, Kanada, Avustralya ve Yeni Zelanda ile yine benzer kültürlere sahip Kore, Japonya, Tayland ülkelerini seçerek ülkeleri ikili olarak karşılaştırmışlardır. Sonuçlarda benzer kültürlere sahip ülkeler arasında ölçme değişmezliğinin sağlandığı fakat farklı kültürlere sahip ülkeler arasında ölçme değişmezliğinin sağlanamadığı, sadece yapısal ve metrik ölçme değişmezliği modellerinin doğrulandığı görülmüştür.

Marsh, Abduljabbar, Ebu-Hilal ve diğerleri (2013), 2007 TIMSS uygulamasında, Suudi Arabistan, Ürdün, Mısır, Umman ile ABD, Avustralya, İngiltere ve İskoçya olmak üzere İngiliz kökenli ülkeler ve Arap ülkelerinde matematik ve fen bilimleri alanlarında kullanılan motivasyon ölçeklerinin cinsiyete ve ülkelere göre değişmezliğini incelemişlerdir. Çalışmalarında, cinsiyete göre değişmezliğin sağlandığını, fakat ülkelere göre ölçme değişmezliğinin sağlanamadığını tespit etmişlerdir.

Öğretmen (2006), yaptığı araştırmada PIRLS 2001 sonuçlarının okuma parçalarına ilişkin testlerin psikometrik özelliklerini Amerika Birleşik Devletleri ve Türkiye örneklemelerinde karşılaştırmalı olarak incelemiştir. Araştırma sonuçları, okuma parçalarına ait testlerin kültürler arasında ölçme değişmezliğinin olmadığını göstermiştir. Ayrıca araştırmada yapılan değişen madde fonksiyonu analizleri sonucunda, maddelerin birçoğunun kültürler arası değişen madde fonksiyonu içerdiği görülmüştür.

Akyıldız (2009), araştırmasında PIRLS 2001 uygulamasındaki sıfır, üç, altı ve sekiz numaralı kitapçıkları seçerek, bu kitapçıklar üzerinden 35 ülke için PIRLS 2001 uygulamasının ölçme değişmezliğini incelemiştir. Elde ettiği sonuçlara göre, PIRLS 2001 testinin tüm ülkeler için yapısal olarak eşit olduğunu yani ölçme değişmezliğinin sağlandığını tespit etmiştir.

Uzun ve Öğretmen (2010), 1999 TIMSS-R uygulamasına katılan öğrencilerin fen bilimleri alanındaki başarılarını açıkladığı düşünülen modelin, cinsiyete göre değişmezliğini araştırdıkları çalışmalarında, modele alınan bütün değişkenlerin metrik değişmezliği sağladığını fakat katı değişmezliği sağlamadığını belirlemişlerdir.

Bahadır (2012), çalışmasında PISA 2009 uygulaması okuma becerisi modelinin coğrafi bölgeler arasında ölçme değişmezliğinin sağlanıp sağlanmadığını araştırmıştır. Elde ettiği sonuçlara göre, okuma becerisine ait modelin bölgelere göre ölçme değişmezliğinin sağlandığını ve modele alınan değişkenlerin bölgelere göre farklılıklarının incelenerek bu farklılıklara ilişkin yorum yapılabileceğini tespit etmiştir.

Uyar ve Doğan (2014), PISA 2009 Türkiye uygulamasında öğrenme stratejilerine ait modelin cinsiyet, okul türü ve istatistiksel bölgelere göre ölçme değişmezliğini belirlemeye çalıştıkları araştırmalarında, öğrenme stratejileri modelinin cinsiyet grupları ve okul türüne göre yapısal ve metrik ölçme değişmezliği koşulunu sağladığını tespit etmişlerdir. 12 istatistiksel bölge arasında yapılan değişmezlik çalışmasında ise, Doğu Karadeniz ve Kuzeydoğu Anadolu bölgelerine ait ölçme modellerinin uygunluğu kabul edilebilir düzeyde olmadıkları için analiz dışı bırakılmış, analize dâhil edilen bölgeler arasında ise ölçme değişmezliğinin sağlandığını belirlemişlerdir.

Kıbrıslıoğlu (2015), PISA 2012 Uygulaması matematik öğrenme modelinin Türkiye, Çin-Şangay ve Endonezya ülkeleri ve cinsiyete göre ölçme değişmezliğini belirlemeye çalışmıştır. Ükelere göre yapılan değişmezlik testlerinin sonucunda, matematik öğrenme algısı modelinde yalnızca şekil değişmezliğine ait ölçme modelinin doğrulandığını tespit etmiştir. Cinsiyete göre yaptığı analizler sonucunda ise kız ve erkek grupları arasında ölçme değişmezliğinin sağlandığını belirlemiştir.

Karakoç Alatl (2016), PISA 2012 uygulamasında Avustralya, Fransa, Şangay-Çin ve Türkiye örneklemelerinde 3 numaralı kitapçığı alan bireylere ait veriler üzerinden yürüttüğü çalışmasında matematik okuryazarlığı, fen okuryazarlığı ve okuma becerileri testlerinin dile göre ölçme değişmezliğini incelemiştir. Araştırma

sonularında matematik ve fen okuryazarlıęının yapısal deęişmezlik gösterirken metrik deęişmezlik göstermedięini tespit etmiştir.



## BÖLÜM III

### YÖNTEM

Bu bölümde araştırmanın modeli, evren ve örneklem, verilerin elde edilmesi ve verilerin analizine ilişkin bilgiler yer almaktadır.

#### 3.1. Araştırmanın Modeli

Bu araştırmada, PISA 2012 uygulamasında bilişsel testte yer alan matematik okuryazarlığına ilişkin maddelerin OECD üyesi ülkeler arası ve Türkiye’de cinsiyet ve bölgeler açısından eşitlik durumu incelenmiştir. Bu bağlamda çalışma, PISA 2012 değerlendirmesine ait bilişsel testin geçerlilik düzeyini belirlemeye yönelik olduğu için ilişkisel tarama modelinde yürütülmüştür. Tarama modelinde yürütülen çalışmalarda, evreni temsil eden belirli bir örneklem grubu üzerinde yapılan çalışmalarla evren hakkında genel bir yargıya varabilme amaçlanır (Bailey, 1982).

#### 3.2. Evren ve Örneklem

PISA 2012 uygulamasında toplam 65 ülkeden, 15 yaş grubundaki yaklaşık 510.000 öğrenci yer almıştır. Uygulamada örneklem iki aşamalı tabakalı örnekleme yöntemi ile belirlenir. Tabakalı örnekleme evrendeki bütün alt grupların, evrendeki oranlarına uygun olacak biçimde örnekleme temsil edilme durumlarının garanti altına alındığı örnekleme türüdür (Balcı, 2013). Türkiye’den ise PISA uluslararası merkezce, 15 yaş grubundaki ulaşılabilir 955.349 öğrenciyi temsilen, seçkisiz yöntemle belirlenmiş 170 okuldan toplam 4848 öğrenci katılmıştır. Araştırmanın evrenini OECD üyesi ülkelerde PISA 2012 uygulamasına katılan 15 yaş grubundaki öğrenciler oluşturmaktadır.

Her bir ülkede PISA 2012 uygulamasında kullanılan bilişsel test maddeleri, 13 ayrı kitapçığa ayrılarak uygulanır (MEB, 2015). Her bir kitapçıktaki sorular birbirinden farklıdır. Bu nedenle araştırmada tüm ülkelerde ortak olarak uygulanan 8 no’lu kitapçık tercih edilerek analizler bu kitapçıktaki matematik okuryazarlığını ölçmeyi amaçlayan maddeler üzerinden gerçekleştirilmiştir.

Bu arařtırmaya OECD üyesi olan ÷lkelerde 8. Kitapçıđı alan tüm bireyler dâhil edildiđi için arařtırmanın amaçları dođrultusunda iki ayrı evren tanımlanmıştır. Arařtırmanın alt amaçları dođrultusunda tanımlanan ilk evren, OECD üyesi ÷lkelerde 8. kitapçıđın uygulandıđı 23.311 öđrenci, ikinci evren ise Türkiye’de 8. Kitapçıđı alan 377 öđrencidir.

Deđişmezlik çalıřmasının Türkiye’deki cinsiyet ve bölgeler üzerinde yapılmasını da amaçlayan bu çalıřmada; İstatistikî Bölge Birimleri Sınıflaması (İBBS) Düzey 1’e göre ayrılmıř 12 bölgeye göre katılımcı sayılarının yapılacak olan analizin varsayımlarını karřılamaması üzerine bölge sayısı birleřtirilerek üçe düşür÷lmüřtür. Bölgeler beřeri ve fiziki cođrafya alanlarında uzman görüřü alınarak birleřtirilmiştir. Bölgeler birleřtirilirken, bölgelerin sosyo-ekonomik durumu, eđitim durumları, kültürel özellikleri bakımından benzerlikleri dikkate alınmıştır. Bu bağlamda, Ege, Akdeniz, Batı Anadolu bölgeleri birleřtirilerek “Bölge 1”; Orta Anadolu, Güneydođu Anadolu, Ortadođu Anadolu, Kuzeydođu Anadolu, Dođu Karadeniz bölgeleri birleřtirilerek “Bölge 2”; İstanbul Batı Marmara, Dođu Marmara ve Batı Karadeniz bölgeleri birleřtirilerek “Bölge 3” olarak tanımlanmıştır. Türkiye uygulamasında 8. Kitapçıđı alan bireylerin cinsiyet göre dađılımı Tablo 1’deki řekliyledir.

**Tablo 1. Cinsiyete göre örneklem dađılımı**

Grup	N	%
Kız	183	48.5
Erkek	194	51.5
Toplam	377	100

Tablo 3.1’de gör÷ldüđu üzere çalıřmaya dâhil edilen 377 bireyden %48.5’i kız (183 kiři) ve %51.5’i (194 kiři) erkektir. Buna göre, cinsiyete göre öđrenci oranlarının birbirine yakın olduđu söylenebilir.

Türkiye uygulamasında 8. Kitapçıđı alan bireylerin bölgelere göre dađılımı Tablo 2’deki řekliyledir.

**Tablo 2. Bölgelere göre örneklem dağılımı**

Grup	N	%
<b>Bölge 1</b>	141	37.4
<b>Bölge 2</b>	109	29
<b>Bölge 3</b>	127	33.6
<b>Toplam</b>	377	100

Tablo 3.2 incelendiğinde çalışmaya dâhil edilen bireylerin %37.4'ü (141 kişi) Bölge 1'de, %29'u (109 kişi) Bölge 2'de, %33.6'sı (127 kişi) Bölge 3'te yer almaktadır. Katılımcılara ait bu değerlere bakıldığında Bölge 1, Bölge 2 ve Bölge 3'e göre öğrenci oranlarının birbirlerine yakın olduğu söylenebilir.

Örnekleme OECD üyesi ülkelerde 8. Kitapçığı alan bireylerin ülkelere göre dağılımı ise Tablo 3'teki gibidir.

**Tablo 3. OECD üyesi ülkelere göre evren dağılımı**

Ülke	N	%
<b>Avustralya</b>	1198	5,13
<b>Avusturya</b>	373	1,60
<b>Belçika</b>	661	2,83
<b>Kanada</b>	1679	7,20
<b>Şili</b>	532	2,28
<b>Çek Cumhuriyeti</b>	429	1,84
<b>Danimarka</b>	558	2,39
<b>Estonya</b>	404	1,73
<b>Finlandiya</b>	665	2,85
<b>Fransa</b>	353	1,51
<b>Almanya</b>	387	1,66
<b>Yunanistan</b>	391	1,67
<b>Macaristan</b>	361	1,54

**Tablo 3'ün devamı: OECD üyesi ülkelere göre örneklem dağılımı**

Ülke	N	%
İzlanda	258	1,10
İrlanda	376	1,61
İsrail	561	2,40
İtalya	2643	11,33
Japonya	477	2,04
Kore	394	1,69
Lüksemburg	403	1,72
Meksika	2591	11,11
Hollanda	323	1,38
Yeni Zelanda	346	1,48
Norveç	351	1,50
Polonya	380	1,63
Portekiz	442	1,89
Slovakya	396	1,69
Slovenya	453	1,94
İspanya	1963	8,42
İsveç	354	1,51
İsviçre	864	3,70
Türkiye	377	1,61
Birleşik Krallık	954	4,09
Amerika Birleşik Devletleri	414	1,77

Tablo 3 incelendiğinde 34 OECD üyesi ülkeden çalışmaya dâhil edilen bireylerin dağılımı incelendiğinde en fazla katılımcıya sahip ülke % 11,33 (2643 kişi) ile İtalya, en az katılımcıya sahip ülke %1,10 (258 kişi) ile İzlanda'dır. Diğer ülkelerin katılımcı oranları ise %11,11 ve %1,38 (2591 ve 323 kişi) arasında değişmektedir.

### 3.3. Araştırma Verileri

PISA 2012 uygulaması, öğrencilerin akademik performanslarının ölçüldüğü bilişsel testler ile öğrencilerin bütün olarak değerlendirilmesinin amaçlandığı okul ve öğrenci anketlerinden meydana gelir. Bu uygulamada bilişsel testler, öğrencilerin matematik okuryazarlığı ağırlıklı olmak üzere, fen okuryazarlığı ve okuma becerilerine ait performanslarını ölçmeyi amaçlar. Kâğıt-kalem testi olarak yapılan sınav, 30’ar dakikadan oluşan dört küme halinde toplam iki saat süren bir uygulamadır.

PISA 2012 araştırmasına katılan ülkelerin tümünde, test maddeleri 13 kitapçığa ayrılarak uygulanmıştır (OECD, 2013a). Kitapçıklarda açık uçlu ve çoktan seçmeli maddeler bulunmaktadır. Bu çalışmada, tüm ülkelerde uygulanmış olan 8 no’lu kitapçık seçilmiş ve maddelere ait puanlamalar madde türüne göre farklılık gösterdiği için sadece matematik okuryazarlığına ait, puan değerleri doğru cevap “1” puan, yanlış cevap “0” puan olan çoktan seçmeli 11 madde analize dâhil edilmiştir. Araştırma kullanılan maddeler Tablo 4’te gösterilmiştir.

Verilere OECD’ye ait resmi web sitesinin PISA ile ilgili bölümünden ulaşılmış ve <https://www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm> adresinden indirilmiştir.

**Tablo 4. Analize dâhil edilen matematik okuryazarlığı maddeleri**

<b>Maddeler</b>	
<b>Matematik okuryazarlığı</b>	PM00KQ02- Wheelchair Basketball Q2
	PM906Q01- Crazy Ants Q1
	PM915Q01- Carbon Tax Q1
	PM915Q02- Carbon Tax Q2
	PM982Q01- Employment Data Q1
	PM982Q02- Employment Data Q2
	PM982Q03T- Employment Data Q3
	PM982Q04- Employment Data Q4
	PM992Q01- Spacers Q1
	PM992Q02- Spacers Q2
	PM992Q03- Spacers Q3



### 3.4. Verilerin Analizi

Veriler analiz edilmeden önce tüm sayıtlar incelenmiş, yapılacak olan analizlere uygun biçimde düzenlenmiştir. Veri setinin psikometrik özellikleri incelenmiş, merkezi eğilim ölçüleri, basıklık ve çarpıklık katsayıları, KR-20 güvenilirlik katsayıları hesaplanmıştır. Yapı geçerliliğine ilişkin kanıt elde edebilmek amacıyla DFA yapılmıştır. Daha sonra kovaryans matrislerinin eşitliği test edilerek, maddelere ait ölçme değişmezliği OECD üyesi ülkelere ve Türkiye'deki bölgelere ve cinsiyete göre ÇGDFA yöntemi ile test edilmiştir. Analizler LISREL 8.7 programı kullanılarak yapılmış, asimtotik kovaryanslar matrisi üzerinden model parametrelerinin kestirilmesinde en çok olabilirlik yöntemi (maximum likelihood estimation) kullanılmıştır.

Analizlerin yapılabilmesi için veri setinin temel varsayımları karşılayıp karşılamadığı test edilmiştir. Bu varsayımlar ait sonuçlar aşağıda verilmiştir.

**Kayıp Değerler:** Analiz sonuçlarını önemli derecede etkileyen kayıp değerlerle baş etmede kullanılan farklı yaklaşımlar vardır. Bu yaklaşımlardan ilki; çok az sayıda değişkenin kayıp değere sahip olması durumunda probleme neden olan değişkenlerin silinerek veri setinden çıkarılmasıdır. İkincisi ise; veri setine ait ortalama hesaplayarak kayıp değerlere elde edilen ortalamaları atamaktır. Üçüncü yaklaşım ise; regresyon yaklaşımını kullanarak kayıp değerlere ait kestirim yapmaktır. Bu araştırmada kayıp değerlerin örneklem büyüklüğünü önemli derecede etkilemeyeceği düşünüldüğünden, bu değişkenlerin veri setinden çıkarılmasına karar verilmiştir.

**Uç Değerler:** Bir veri setindeki deneklerin aldıkları puanların aşırı değerlere sahip olması, dağılımın uç noktalarında yer alması durumunda, bu denekler uç değer olarak adlandırılır. Uç değerler istatistiksel sonuçları önemli düzeyde etkileyerek, analizde hataların çıkmasına neden olur. (Tabachnick ve Fidell, 2013). Uç değerler, tek yönlü ve çok yönlü olabilirler. Tek yönlü uç değerleri belirlemede sıklıkla kullanılan yöntemlerden biri puanların Z puanlarına dönüştürülerek belirlenen kesme noktasından büyük ya da küçük olan verilerin örneklemden çıkarılmasıdır (Çokluk, Şekercioğlu ve Büyüköztürk, 2012). Çok yönlü uç değerler ise Mahalanobis Uzaklığı, Kaldıraç Noktası ve Cook Uzaklığı yöntemleri kullanılarak tespit edilebilir (Koğar, 2010). Araştırmada tek yönlü uç değer taraması yapılmış, kesme noktası olarak  $\pm 3$  belirlenerek (Raykov ve Marcoulides, 2008), bu kesme noktasının dışında

kalan deęerler analiz dıřı bırakılmıř ve analiz OECD üyesi ÷lkeler için 23.311 kiři, Türkiye’de ise 377 kiři üzerinden yürüt÷lmüřtür.

Normallik: Tek deęiřkenli ve çok deęiřkenli normallik olmak üzere iki çeřit olan normallik veri setindeki deęiřkenlere ait gözlemlerin normal daęılım göstermesi anlamına gelir. Yapısal Eřitlik Modellemesi çerçevesinde çok deęiřkenli normallięin hesaplanması güç olduęundan, tek deęiřkenli normallik hesaplama yöntemlerinden olan basıklık ve çarpıklık katsayılarına bakılarak verilerin normallięine karar verilmiřtir (Weston ve Gore, 2006). Elde edilen sonuçlara bulgular bölümünde yer verilmiřtir.

Çoklu Baęlantılılık Problemi: Baęımsız deęiřkenlerin birbirleriyle güçlü iliřkilerinin olduęu durumlarda çoklu baęlantı problemi vardır ve deęiřkenlerin arasındaki korelasyonun genellikle .90 ve üzerinde olduęu durumlarda gözlenir (Çokluk ve dięerleri, 2013). Baęımsız deęiřkenler arasındaki korelasyonlara bakılmıř ve deęiřkenler arasındaki korelasyonların .046 ile .31 arasındaki deęiřtięi gör÷lmüřtür. Tolerans deęerleri .01’den büyük olduęu durumlarda çoklu baęlantı probleminin olmadığı söylenebilir (Tabachnick ve Fidell, 2013). Tolerans deęerleri .78 ile .94 arasında deęiřmektedir. Elde edilen deęerler dikkate alındığında çoklu baęlantı problemi olmadığını söylemek mümkündür.

Bu çerçevede, veri setine iliřkin tüm sayıltıların karřılandığı söylenebilir.

Yapılan arařtırmada veri setine iliřkin yapı geçerlilięi kanıtlarını ortaya koymak amacıyla her bir grup için ayrı ayrı DFA yapılmıřtır. DFA sonuçları bulgular bölümünde verilmiřtir.

Arařtırmada, S-B<sub>χ<sup>2</sup></sub> düzeltmesi farklı örneklem büyüklüęüne ve puan daęılımına sahip veri setlerinde, örneklem büyüklüęünün  $\chi^2$  üzerindeki etkisini azalttıęı için,  $\chi^2$  deęeri olarak S-B<sub>χ<sup>2</sup></sub> kullanılmıř ve model uyumuna karar verilirken  $\chi^2$  ,  $\chi^2/sd$ , CFI ve SRMR uyum indeksleri dikkate alınmıřtır.

Ölçme modelleri karřılařtırılırken ise uyum indekslerindeki deęiřimler incelenerek, bu indeksler arasındaki farka bakılmıřtır. İlk olarak karřılařtırılan modellerin,  $\chi^2$  istatistięi ve serbestlik dereceleri arasındaki (sd) fark ( $\Delta\chi^2$  ve  $\Delta sd$ ) incelenir. Karřılařtırma sonucunda elde edilen farkın anlamlı olup olmadığı  $\chi^2$  tablosundaki

değerler  $p < .01$  ya da  $p < .05$  düzeyinde karşılaştırılarak tespit edilir. Eğer  $\chi^2$  tablosundaki değer elde edilen sonuçtan büyükse modeller arasında anlamlı fark olduğu kabul edilir (Kline, 2005). Bu araştırmada  $\chi^2$  olarak  $S-B_{\chi^2}$  değeri temel alındığı için fark derecesi belirlenirken  $T_s$  değerleri hesaplanmış ve modeller arasındaki farkın manidarlığı elde edilen  $T_s$  değeri ile  $\chi^2$  tablosundaki kritik değerler  $p < .05$  düzeyinde karşılaştırılarak belirlenmiştir. Bununla birlikte model karşılaştırmalarında  $\Delta CFI$ ,  $\Delta SRMR$ ,  $\Delta NFI$ ,  $\Delta GFI$ ,  $\Delta RMSEA$ ,  $\Delta TLI$  gibi uyum indekslerinin değerlendirilmesi önerilmiştir (Cheng ve Rensvold, 2002). Araştırmada  $\Delta CFI$  ve  $\Delta SRMR$  değerleri temel alınarak karşılaştırmalar yapılmıştır. Bu karşılaştırmalar için kabul düzeyleri,  $n > 300$  olduğunda metrik değişmezlik için  $\Delta CFI \geq -.010$  ve  $\Delta SRMR \geq .015$ , güçlü ve katı faktöriyel değişmezliğin karşılaştırılmasında  $\Delta CFI \geq -.010$  ve  $\Delta SRMR \geq 0,010$ .  $n < 300$  olduğunda metrik değişmezlik için  $\Delta CFI \leq -.005$  ve  $\Delta SRMR \geq .025$ , güçlü ve katı faktöriyel değişmezliğe ait karşılaştırmalarda  $\Delta CFI \geq -.005$ ,  $\Delta SRMR \geq 0.005$  olarak belirlenmiştir (Chen, 2007).

## BÖLÜM IV

### BULGULAR

Bu bölümde araştırmanın amaçları doğrultusunda kullanılan veri seti için PISA 2012 Matematik okuryazarlığının, Türkiye’de cinsiyet ve bölgelere göre, sonrasında da OECD üyesi ülkelere göre değişmezliğine ilişkin bulgulara ve söz konusu gruplara ait test istatistiklerine, normallik testlerine ve güvenilirlik katsayılarına yer verilmiştir.

#### 4.1. Cinsiyet Grupları için Ölçme Değişmezliğine İlişkin Bulgular

Bu başlık altında araştırmanın birinci alt problemi olan PISA 2012 Türkiye uygulaması matematik okuryazarlığının cinsiyet grupları için ölçme değişmezliğinin sağlanıp sağlanmadığı sorusuna cevap aranmıştır. Cinsiyetler arası ölçme değişmezliğine ilişkin bulgular verilmeden önce, Türkiye uygulamasında yer alan kız ve erkek grupları için merkezi eğilim ölçüleri, basıklık ve çarpıklık katsayıları, KR-20 güvenilirlik katsayısı hesaplanarak sonuçlar Tablo 5’de verilmiştir.

**Tablo 5. Kız ve Erkek Grupları için Matematik Okur-Yazarlığı Testi Puanlarının Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları**

Cinsiyet	N	$\bar{X}$	Mod	S	Ranj	$K_y$	$B_y$	KR-20
Kız	183	4.17	2	2.23	10	.493	-.358	.664
Erkek	194	4.23	4	2.37	11	.628	.011	.695

$K_y$ : Çarpıklık katsayısı

$B_y$ : Basıklık katsayısı

KR-20: Güvenilirlik katsayısı

Tablo 5’deki sonuçlara göre, gruplara ait merkezi eğilim ölçülerinin birbirine yakın olduğu söylenebilir. Normallik için basıklık ve çarpıklık katsayıları incelendiğinde değerlerin  $\pm 1$  arasında yer alıyor olması veri setinin normal dağılıma yakın bir dağılım gösterdiği anlamına gelir (Rosenthal ve Rosnow, 2008). Cinsiyet grupları için güvenilirlik katsayıları incelendiğinde ise güvenilirliğin kabul edilebilir düzey olarak belirlenen .70-.80 aralığında olmadığı (Nunnaly ve Bernstein, 1994), ancak .70 kabul düzeyine yakın olduğu görülmüştür. Madde sayısının az olması

güvenilirliği oldukça etkileyen bir durumdur. Bu durumun madde sayısının az olması ile ilişkili olduğu düşünülebilir.

ÇGDFA yapılmadan önce cinsiyet gruplarına ait kovaryans matrislerinin eşitliği test edilmiştir. Test sonuçları Tablo 6’da verilmiştir.

**Tablo 6. Kız ve Erkek Grupları için Matematik Okur-Yazarlığı Testi Puanlarının Kovaryans Matrislerinin Eşitliği Testi Sonuçları**

Grup	S-B <sub>xx</sub> (sd)	p	$\chi^2/sd$	RMSEA	GFI	CFI	SRMR
Kız-Erkek	41.55 (66)	.99	.62	.00	.98	1.00	.045

Tablo 6’da  $\chi^2/sd$  oranı 2’nin altında, RMSEA ve SRMR değerlerinin .05’in altında ve CFI ve GFI değerlerinin .95’in üzerinde olduğu görülmektedir. Bu durumda kız ve erkek gruplarına ait iki kovaryans matrisi arasında yüksek bir uyumun olduğu söylenebilir.

Cinsiyet gruplarına ait DFA sonuçları ve PISA 2012 matematik okuryazarlığının, kız ve erkek grupları için ölçme değişmezliğinin sağlanıp sağlanmadığını tespit etmek amacıyla yapılan ÇGDFA sonuçları Tablo 7’de verilmiştir.

**Tablo 7. Kız ve Erkek Grupları için Matematik Okur-Yazarlığı Testi Puanlarının Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları**

	S-B <sub>xx</sub> (sd)*	M.K.**	$\Delta\chi^2(\Delta sd)$	$\chi^2/sd$	$\Delta\chi^2/\Delta sd$	CFI	$\Delta CFI$	SRMR	$\Delta SRMR$
Kız	34.17 (44)	-	-	.77	-	1.00	-	.044	-
Erkek	37.89 (44)	-	-	.86	-	1.00	-	.044	-
Model 1 <sup>A</sup>	85.79 (110)	-	-	.78	-	1.00	-	.057	-
Model 2 <sup>B</sup>	77.75 (99)	M1-M2	8.04 (11)	.78	0	1.00	0	.047	.01
Model 3 <sup>C</sup>	77.75 (99)	M2-M3	0	.78	0	1.00	0	.047	0
Model 4 <sup>D</sup>	41.55 (66)	M3-M4	36.2 (33)	.62	.16	1.00	0	.045	.002

\*p<.05

\*\*Model karşılaştırmaları (M=Model)

*A* Yapısal Değişmezlik (Faktör yükleri, faktör korelasyonları ve hata varyansları sabit)

*B* Zayıf Faktöriyel Değişmezlik (Faktör yükleri serbest, faktör korelasyonları ve hata varyansları sabit)

*C* Güçlü Faktöriyel Değişmezlik (Faktör yükleri ve hata varyansları serbest, faktör korelasyonları sabit)

*D* Katı Faktöriyel Değişmezlik (Hata varyansları serbest, faktör yükleri ve faktör korelasyonları sabit)

Tablo 7'deki kız ve erkek grupları için DFA sonuçlarına ait uyum indekslerini incelediğimizde her iki grup için de  $\chi^2/sd$  oranının 2'nin altında olduğu, CFI değerinin .95'in üzerinde olduğu, SRMR değerinin ise .05'ten küçük olduğu görülmektedir. Bu durumda, testin faktör yapısının kız ve erkek grupları için ayrı ayrı doğrulandığı söylenebilir.

Tablo 7'de görüldüğü üzere cinsiyet gruplarına ait yapısal modelin (Model 1) kovaryans matrislerinde faktör yüklerinin, faktör korelasyonlarının ve hata varyanslarının tüm gruplar için eşit olup olmadığının test edilmesi sonucunda  $S-B_{\chi^2}$  ve serbestlik derecesi oranının 2'nin altında olduğu, CFI değerinin 1'e eşit olduğu ve SRMR değerinin .05 civarında olduğu görülmektedir. Bu doğrultuda uyum indekslerine ait değerlerin kabul edilebilir düzeylerde olduğu ve model uyumunun iyi olduğu söylenebilir.

Model 1 (yapısal değişmezlik) ve faktör yüklerinin serbest bırakıldığı Model 2 (zayıf faktöriyel değişmezlik) karşılaştırıldığında  $\Delta\chi^2/\Delta sd$  oranının ve CFI'nın değişmediği,  $\Delta SRMR$ 'nin ise manidar olmadığı ( $<.025$ ) söylenebilir.

Model 2 (zayıf faktöriyel değişmezlik) ile faktör yükleri ve hata varyanslarının serbest, faktör korelasyonlarının sabit olduğu Model 3 (güçlü faktöriyel değişmezlik) karşılaştırıldığında,  $\Delta\chi^2/\Delta sd$  oranının, CFI ve SRMR değerlerinin herhangi bir değişim göstermediği görülmektedir.

Model 3 (güçlü faktöriyel değişmezlik) ve hata varyanslarının serbest, faktör yükleri ve faktör korelasyonlarının sabit olduğu Model 4 (katı faktöriyel değişmezlik) karşılaştırıldığında  $\Delta\chi^2/\Delta sd$  oranının bir miktar iyileştiği görülmektedir. İyileşmenin manidar olup olmadığını belirleyebilmek için  $S-B_{\chi^2}$  fark derecesi için  $T_s$  değeri hesaplanmış ve bu değer 37.63 bulunmuştur. Hesaplanan  $T_s$  değerinin  $\chi^2$  dağılımı tablosundaki kritik değerden küçük olduğu belirlenmiştir,  $\chi^2_{fark}(33)=47.4$ ,  $p>.05$ . Bu durumda bu değer manidar olmadığını söylemek olanaklıdır. CFI değerinin değişmediği ve  $\Delta SRMR$  değerinin de manidar olmadığı tespit edilmiştir ( $<.025$ ). Bu durumda Model 3 ve Model 4 arasındaki farkın manidar olmadığı ifade edilebilir.

Tüm model karşılaştırmalarına ait sonuçlar değerlendirildiğinde karşılaştırılan modeller içerisinde en iyi çalışan modelin Model 1 (yapısal değişmezlik modeli) olduğu söylenebilir. Bu durumda cinsiyet gruplarına ait kovaryans matrisleri temelinde faktör yapılarının eşit olduğu ifade edilebilir. Bu doğrultuda PISA 2012 Türkiye uygulaması matematik okuryazarlığı testinin kız ve erkek grupları için ölçme değişmezliğinin sağlandığı kararına varılmıştır.

#### 4.2. Bölge 1, Bölge 2 ve Bölge 3 Grupları için Ölçme Değişmezliğine İlişkin Bulgular

Bu başlık altında araştırmanın ikinci alt problemi olan PISA 2012 Türkiye uygulaması matematik okuryazarlığının bölgelere göre ölçme değişmezliğinin sağlanıp sağlanmadığı sorusuna cevap aranmıştır. Bölgeler arası ölçme değişmezliğine ilişkin bulgulara yer verilmeden önce, Türkiye uygulamasında yer alan Bölge 1, Bölge 2 ve Bölge 3 grupları için merkezi eğilim ölçüleri, basıklık ve çarpıklık katsayıları, KR-20 iç tutarlılık güvenilirlik katsayısı hesaplanmıştır. Sonuçlar Tablo 8’de verilmiştir.

**Tablo 8. Bölgeler için Matematik Okur-Yazarlığı Testi Puanlarının Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları**

Bölgeler	N	$\bar{X}$	Mod	S	Ranj	$K_y$	$B_s$	KR-20
Bölge 1	141	4.34	3	2.32	11	0,574	-0.34	.68
Bölge 2	109	4.02	4	2.27	10	0,549	-.350	.66
Bölge 3	127	4.19	2	2.30	11	0,592	-0.43	.69

$K_y$ : Çarpıklık katsayısı

$B_s$ : Basıklık katsayısı

KR-20: Güvenilirlik katsayısı

Tablo 8’de yer alan sonuçlara göre, gruplara ait merkezi eğilim ölçülerinin birbirine yakın olduğunu söylemek olanaklıdır. Normallik için basıklık ve çarpıklık katsayıları incelendiğinde tüm gruplara ait değerlerin  $\pm 1$  arasında yer alıyor olması veri setinin normale yakın bir dağılım gösterdiği anlamına gelir. Söz konusu gruplar için güvenilirlik katsayıları incelendiğinde ise güvenilirliğin 0,70’e yakın olduğu görülmüştür.

ÇGDFA öncesinde Bölge 1, Bölge 2 ve Bölge 3'e ait kovaryans matrislerinin eşitliği test edilmiştir. Test sonuçları Tablo 9'da verilmiştir.

**Tablo 9. Bölgeler için Matematik Okur-Yazarlığı Testi Puanlarının Kovaryans Matrislerinin Eşitliği Testi Sonuçları**

Grup	S-B <sub>χ<sup>2</sup></sub> (sd)	p	χ <sup>2</sup> /sd	RMSEA	GFI	CFI	SRMR
Bölge 1-Bölge 2- Bölge 3	123.98 (132)	.68	.93	.00	.95	1.00	.060

Tablo 9'da görüldüğü gibi χ<sup>2</sup>/sd oranı 2'nin altında, RMSEA değeri 0.0, SRMR değeri .08'in altında ve CFI ve GFI değerleri .95 ve üzerindedir. Uyum indeksleri değerlendirildiğinde Bölge 1, Bölge 2 ve Bölge 3 gruplarına ait üç kovaryans matrisi arasında uyumun yüksek olduğu söylenebilir.

Bölgelere ait DFA sonuçları ve PISA 2012 matematik okuryazarlığının, bölgeler arasında ölçme değişmezliğinin sağlanıp sağlanmadığını tespit etmek amacıyla yapılan ÇGDFA sonuçları Tablo 10'da verilmiştir.

**Tablo 10. Bölgeler için Matematik Okur-Yazarlığı Testi Puanlarının Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları**

	S-B <sub>χ<sup>2</sup></sub> (sd)*	M.K.**	Δχ <sup>2</sup> (Δsd)	χ <sup>2</sup> /sd	Δχ <sup>2</sup> /Δsd	CFI	ΔCFI	SRMR	Δ SRMR
Bölge 1	30.57 (43)	-	-	.71	-	1.00	-	.046	-
Bölge 2	44.97 (44)	-	-	1.02	-	0.93	-	.067	-
Bölge 3	32 (43)	-	-	.74	-	1.00	-	.05	-
Model 1 <sup>A</sup>	173.26 (176)	-	-	.98	-	1.00	-	.069	-
Model 2 <sup>B</sup>	144.95 (154)	M1-M2	28.31(22)	.94	.04	1.00	0	.060	.009
Model 3 <sup>C</sup>	129.15 (132)	M2-M3	15.8 (22)	.97	.03	1.00	0	.056	.004
Model 4 <sup>D</sup>	161.80 (154)	M3-M4	-32.65 (-22)	1.05	-.08	.99	.01	.067	-.011

\*p<.05

\*\*Model karşılaştırmaları (M=Model)

*A* Yapısal Değişmezlik (Faktör yükleri, faktör korelasyonları ve hata varyansları sabit)

*B* Zayıf Faktöriyel Değişmezlik (Faktör yükleri serbest, faktör korelasyonları ve hata varyansları sabit)

*C* Güçlü Faktöriyel Değişmezlik (Faktör yükleri ve hata varyansları serbest, faktör korelasyonları sabit)

*D* Katı Faktöriyel Değişmezlik (Hata varyansları serbest, faktör yükleri ve faktör korelasyonları sabit)



Tablo 10'da bölgelere ait DFA sonuçlarının uyum indeksleri incelendiğinde her üç gruba ait  $\chi^2/sd$  oranının 2'nin altında olduğu, CFI değerinin Bölge 1 ve Bölge 3 için .95'in üzerinde, Bölge 2 için .93 olduğu ve tüm bölgeler için SRMR değerinin ise .08'den düşük olduğu görülmektedir. Bu durumda, testin faktör yapısının bölgeler için ayrı ayrı doğrulandığını söylemek mümkündür.

Tablo 10'da görüldüğü üzere bölgelere ait yapısal modelin (Model 1) kovaryans matrislerinde faktör yüklerinin, faktör korelasyonlarının ve hata varyanslarının tüm gruplar için eşit olup olmadığının test edilmesi sonucunda  $S-B_{\chi^2}$  ve serbestlik derecesi oranının 2'nin altında, CFI değerinin .95'in üzerinde ve SRMR değerinin .08'den küçük olduğu görülmektedir. Uyum indekslerinin değerlerinin kabul edilebilir düzeylerde olduğu ve Model 1'e ait uyumunun iyi olduğu söylenebilir.

Model 1 (yapısal değişmezlik) ve faktör yüklerinin serbest bırakıldığı Model 2 (zayıf faktöriyel değişmezlik) karşılaştırıldığında  $\Delta\chi^2/\Delta sd$  oranının iyileştiği görülmektedir. İyileşmenin manidar olup olmadığını belirleyebilmek için hesaplanan  $T_s$  değeri 28.126 bulunmuştur. Bu değer  $\chi^2$  dağılımı tablosundaki kritik değerden küçük olduğu ve değer manidar olmadığı tespit edilmiştir,  $\chi^2_{fark}(22)=33.924$ ,  $p>.05$ . Ayrıca CFI değerinin değişmediği,  $\Delta SRMR$  değerinin ise .025'ten küçük olduğu belirlenmiştir. Bu durumda yapılan test sonucu bir miktar iyileşmenin olduğu fakat Model 1 ve Model 2 arasındaki farkın manidar olmadığı ifade edilebilir.

Model 2 (zayıf faktöriyel değişmezlik) ile faktör yükleri ve hata varyanslarının serbest, faktör korelasyonları sabit olduğu Model 3 (güçlü faktöriyel değişmezlik) karşılaştırıldığında,  $\Delta\chi^2/\Delta sd$  oranının iyileştiği tespit edilmiştir.  $S-B_{\chi^2}$  fark derecesinin manidar olup olmadığını tespit etmek için  $T_s$  değeri 0.772 olarak hesaplanmıştır.  $T_s$  değerinin  $\chi^2$  dağılımı tablosundaki kritik değerden küçük olduğu belirlenmiştir,  $\chi^2_{fark}(22)=33.924$ ,  $p>.05$ . Bu farkın manidar olmadığı söylenebilir. CFI değerinde bir değişim gözlemlenmezken ve  $\Delta SRMR$  değerinin manidar olmadığı belirlenmiştir. Bu durumda zayıf faktöriyel değişmezlik ile güçlü faktöriyel değişmezlik arasında manidar bir farkın olmadığını söylemek mümkündür.

Model 3 (güçlü faktöriyel değişmezlik) ve hata varyanslarının serbest, faktör yükleri ve faktör korelasyonlarının sabit olduğu Model 4 (katı faktöriyel değişmezlik) karşılaştırıldığında  $\Delta\chi^2/\Delta sd$  oranı,  $\Delta CFI$  ve  $\Delta SRMR$  değerleri doğrultusunda uyumun kötüleştiğini söylemek mümkündür.

Tüm model karşılaştırmalarına ait sonuçlar değerlendirildiğinde karşılaştırılan modeller içerisinde en iyi çalışan modelin bölgeler için de Model 1 (yapısal değişmezlik modeli) olduğu söylenebilir. Bu durumda bölgeler için kovaryans matrisleri temelinde faktör yapılarının eşit olduğu yani PISA 2012 matematik okuryazarlığının Bölge 1, Bölge 2 ve Bölge 3 arasında ölçme değişmezliğinin sağlandığı ifade edilebilir.

#### **4.2. OECD Üyesi Ülkeler İçin Ölçme Değişmezliğine İlişkin Bulgular**

Bu başlık altında araştırmanın üçüncü alt problemi olan PISA 2012 uygulaması matematik okuryazarlığının OECD üyesi ülkelere göre ölçme değişmezliğinin sağlanıp sağlanmadığı sorusuna cevap aranmıştır.

OECD üyesi ülkeler arası ölçme değişmezliğine ilişkin bulgulara yer verilmeden önce, OECD üyesi ülkelerin matematik okur-yazarlığı testi puanları dağılımı için merkezi eğilim ölçüleri, basıklık ve çarpıklık katsayıları, KR-20 iç tutarlılık güvenilirlik katsayısı hesaplanarak sonuçlar Tablo 11’de verilmiştir.

**Tablo 11. OECD Üyesi Ülkelerin Matematik Okur-Yazarlığı Testi Puanları için Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları**

Ülke	N	$\bar{X}$	Mod	S	Ranj	$K_y$	$B_y$	KR-20
Avustralya	1198	4.99	4	2.35	11	.206	-.506	.70
Avusturya	373	5.18	4	2.22	11	.138	.252	.68
Belçika	661	5.65	6	2.44	11	-.061	-.434	.72
Kanada	1679	5.20	5	2.26	11	.097	-.388	.67
Şili	532	4.2	4	2.15	10	.217	-.548	.64
Çek Cumhuriyeti	429	5.85	5	2.27	11	-.025	.235	.68
Danimarka	558	5.07	6	2.40	10	-.038	-.686	.73
Estonya	404	5.41	5	2.11	11	.174	-.074	.63
Finlandiya	665	5.26	5	2.27	11	.075	-.334	.69
Fransa	353	5.13	4	2.36	11	.203	-.415	.71
Almanya	387	5.39	5	2.32	11	.068	-.517	.71
Yunanistan	391	4.03	3	2.25	11	.520	-.007	.70
Macaristan	361	4.91	4	2.33	11	.227	-.331	.70
İzlanda	258	5.28	5	2.39	10	-.134	-.558	.71
İrlanda	376	5.22	6	2.11	11	-.057	-.158	.63
İsrail	561	4.46	4	2.49	10	.209	-.648	.74
İtalya	2643	5.10	5	2.40	11	.155	-.500	.71
Japonya	477	5.70	7	2.43	11	-.217	-.666	.70
Kore	394	6.04	5	2.57	11	-.037	-.769	.73
Lüksemburg	403	4.73	4	2.48	11	.206	-.484	.75
Meksika	2591	3.58	4	1.91	11	.369	.082	.55
Hollanda	323	5.56	6	2.38	11	.034	-.672	.71

$K_y$ : Çarpıklık katsayısı

$B_y$ : Basıklık katsayısı

KR-20: Güvenilirlik katsayısı

**Tablo 11'in devamı OECD Üyesi Ülkelerin Matematik Okur-Yazarlığı Testi Puanları için Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları**

Ülke	N	$\bar{X}$	Mod	S	Ranj	$K_y$	$B_y$	KR-20
Yeni Zelanda	346	5.08	4	2.32	11	.231	-.450	.68
Norveç	351	4.68	5	2.29	11	.105	-.440	.70
Polonya	380	5.61	5	2.51	11	.076	-.586	.72
Portekiz	442	5.03	4	2.37	11	.123	-.681	.70
Slovakya	396	5.13	4	2.68	11	.164	-.759	.76
Slovenya	453	5.07	4	2.44	11	.256	-.389	.72
İspanya	1963	5.02	5	2.26	11	.111	-.396	.68
İsveç	354	4.65	4	2.41	11	.173	-.733	.72
İsviçre	864	5.48	5	2.34	11	-.040	-.604	.69
Türkiye	377	4.20	4	2.30	11	.569	-.144	.68
Birleşik Krallık	954	5.02	6	2.26	11	.003	-.588	.67
Amerika Birleşik Devletleri	414	4.53	4	2.33	11	.298	-.494	.69

$K_y$ : Çarpıklık katsayısı

$B_y$ : Basıklık katsayısı

KR-20: Güvenilirlik katsayısı

Tablo 11'de yer alan sonuçlara göre, matematik okur-yazarlığı test puanlarının ülkelere ait merkezi eğilim ölçülerinin birbirine yakın olduğu söylenebilir. Normallik için basıklık ve çarpıklık katsayıları incelendiğinde tüm gruplara ait değerlerin  $\pm 1$  arasında yer alıyor olması veri setinin normale yakın bir dağılım gösterdiği anlamına gelir. Ülkelere ait güvenilirlik katsayıları incelendiğinde ise güvenilirliğin .55 ile .76 arasında değiştiği görülmüştür. Güvenilirlik katsayılarının bazı ülkeler için kabul edilebilir düzeyin altında olması, yapılan araştırma bir ölçme aracı geliştirme çalışması olmadığı için sadece raporlaştırılmış ve herhangi bir işlem yapılmamıştır.

ÇGDFA yapılmadan önce ülkelere ait kovaryans matrislerinin eşitliği test edilmiştir. Test sonuçları Tablo 12'de verilmiştir.

**Tablo 12. OECD Üyesi Ülkelerin Matematik Okur-Yazarlığı Testi Puanları için Kovaryans Matrislerinin Eşitliği Testi Sonuçları**

Grup	S-B <sub>χ<sup>2</sup></sub> (sd)	p	χ <sup>2</sup> /sd	RMSEA	GFI	CFI	SRMR
OECD Üyesi Ülkeler	8317.72 (2178)	.00	3.81	.064	.94	.85	.092

Tablo 12’de  $\chi^2/sd$  oranı 5’nin altında, RMSEA değeri .07’nin altında, SRMR değerinin .08’in üzerinde, CFI değerinin .90’ın altında ve GFI değerinin .90’ın üzerinde olduğu görülmektedir. Uyum indeksleri değerlendirildiğinde ülkelere ait kovaryans matrislerinin orta düzeyde uyum gösterdiği ifade edilebilir.

Ülkelere ait DFA sonuçlarının çok sayıda ülke olması sebebiyle ÇGDFA sonuçlarından ayrı tablo halinde sunulması uygun görülmüş ve sonuçlar Tablo 13’de verilmiştir.

**Tablo 13. OECD Üyesi Ülkelerin Matematik Okur-Yazarlığı Testi Puanları için Doğrulayıcı Faktör Analizi Sonuçları**

Ülkeler	S-B <sub>χ<sup>2</sup></sub> (sd)	χ <sup>2</sup> /sd	CFI	SRMR
Avustralya	175.34 (44)	3,98	.95	.037
Avusturya	77.42 (44)	1,75	.95	.047
Belçika	116.13 (44)	2,63	.96	.041
Kanada	129.03 (44)	2,93	.97	.028
Şili	66.81 (44)	1,51	.96	.039
Çek Cumhuriyeti	63.11 (44)	1,43	.97	.038
Danimarka	82.81 (44)	1,88	.97	.038
Estonya	74.75 (44)	1,69	.93	.046
Finlandiya	157.97 (44)	3,59	.92	.048
Fransa	81.44 (44)	1,85	.96	.045
Almanya	72.39 (44)	1,64	.96	.042
Yunanistan	95.02 (44)	2,15	.94	.048

**Tablo 13'ün devamı OECD Üyesi Ülkelerin Matematik Okur-Yazarlığı Testi Puanları için Doğrulayıcı Faktör Analizi Sonuçları**

Ülkeler	S-B <sub>χ²</sub> (sd)	χ²/sd	CFI	SRMR
Macaristan	107.32 (44)	2,43	.93	.053
İzlanda	50.77 (44)	1,15	.99	.044
İrlanda	61.03 (44)	1,38	.96	.042
İsrail	103.80 (44)	2,35	.96	.041
İtalya	288.72 (44)	6,56	.96	.034
Japonya	62.08 (44)	1,41	.98	.036
Kore	77.97 (44)	1,77	.97	.042
Lüksemburg	101.01 (44)	2,29	.96	.048
Meksika	211.52 (44)	4,80	.91	.032
Hollanda	72.38 (44)	1,64	.96	.046
Yeni Zelanda	73.35 (44)	1,66	.95	.045
Norveç	63.34 (44)	1,43	.97	.042
Polonya	66.21 (44)	1,50	.98	.041
Portekiz	90.31 (44)	2,05	.95	.046
Slovakya	68.93 (44)	1,56	.98	.039
Slovenya	97.27 (44)	2,21	.95	.044
İspanya	235.04 (44)	5,34	.95	.035
İsveç	38.70 (44)	0,87	1.00	.034
İsviçre	93.33 (44)	2,12	.97	.033
Türkiye	41.60 (44)	0,94	1.00	.033
Birleşik Krallık	130.18 (44)	2,95	.95	.038
Amerika Birleşik Devletleri	73.37 (44)	1,66	.96	.042

Tablo 13'deki sonuçlara göre 29 ülkenin  $\chi^2/sd$  oranınının 3'ün altında, 3 ülkenin  $\chi^2/sd$  oranınının 5'in altında olduğu görülmektedir. İspanya ve İtalya'ya ait  $\chi^2/sd$  oranlarınının 5'ten büyük olduğu tespit edilmiştir. Fakat söz konusu iki ülkenin örneklem boyutlarının büyük olmasının bu oranı etkilediği düşünüldüğünden diğer uyum indeksleri incelenmiştir. Tüm ülkelere ait CFI değerleri .90'nın üzerindedir. Macaristan hariç diğer ülkelere ait SRMR değerlerinin .05'in altında olduğu, Macaristan'a ait SRMR değerinin ise 0.053 olduğu görülmektedir. Uyum indeksleri genel olarak değerlendirildiğinde tüm ülkelere ait modellerin 34 OECD üyesi ülke için ayrı ayrı doğrulandığını kararı alınmıştır.

PISA 2012 matematik okuryazarlığının, ülkeler arasında ölçme değişmezliğinin sağlanıp sağlanmadığını tespit etmek amacıyla yapılan ÇG DFA sonuçları Tablo 14'te verilmiştir.

**Tablo 14. OECD Üyesi Ülkelerin Matematik Okur-Yazarlığı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları**

	S- $B_{\chi^2}(sd)$	M.K.	$\Delta\chi^2(\Delta sd)$	$\chi^2/sd$	$\Delta\chi^2/\Delta sd$	CFI	$\Delta CFI$	SRMR	$\Delta SRMR$
<b>Model 1<sup>A</sup></b>	10684.821 (2222)	-	-	4.81	-	.79	-	.095	-
<b>Model 2<sup>B</sup></b>	10692.06 (2222)	M1-M2	-7.24 (0)	4.8119	0.00035	.70	0	.095	0
<b>Model 3<sup>C</sup></b>	5630.03 (1859)	M2-M3	5062.03 (363)	3.0285	1.78338	.91	-0.12	.070	.025
<b>Model 4<sup>D</sup></b>	5630.03 (1859)	M3-M4	-	3.0285	0	.91	0	.070	0

\*p<.05

\*\*Model karşılaştırmaları (M=Model)

*A*Yapısal Değişmezlik (Faktör yükleri, faktör korelasyonları ve hata varyansları sabit)

*B*Zayıf Faktöriyel Değişmezlik (Faktör yükleri serbest, faktör korelasyonları ve hata varyansları sabit)

*C*Güçlü Faktöriyel Değişmezlik (Faktör yükleri ve hata varyansları serbest, faktör korelasyonları sabit)

*D*Katı Faktöriyel Değişmezlik (Hata varyansları serbest, faktör yükleri ve faktör korelasyonları sabit)

Tablo 14'de görüldüğü üzere ülkelere ait yapısal modelin (Model 1) kovaryans matrislerinde faktör yüklerinin, faktör korelasyonlarının ve hata varyanslarının tüm gruplar için eşit olup olmadığının test edilmesi sonucunda S- $B_{\chi^2}$  ve serbestlik derecesi oranınının 5'in altında, CFI değerinin .90'ın altında ve SRMR değerinin .1'in altında görülmektedir. Uyum indekslerinin değerlerinin kabul edilebilir düzeylerde

olmadığı ve modelin doğrulanmadığı söylenebilir. Bu durumda yapısal değişmezliğin sağlanamadığını söylemek mümkündür.

Model 1 (yapısal değişmezlik) ve faktör yüklerinin serbest bırakıldığı Model 2 (zayıf faktöriyel değişmezlik) karşılaştırıldığında  $\Delta\chi^2/\Delta sd$  oranının kötüleştiği görülmektedir. Ayrıca  $\Delta CFI$  ve  $\Delta SRMR$  değerlerinde değişme meydana gelmemiştir. Bu durumda Model 1 ve Model 2 arasındaki farkın manidar olmadığı söylenebilir.

Model 2 (zayıf faktöriyel değişmezlik) ile faktör yükleri ve hata varyanslarının serbest, faktör korelasyonları sabit olduğu Model 3 (güçlü faktöriyel değişmezlik) karşılaştırıldığında,  $\Delta\chi^2/\Delta sd$  oranının oldukça iyileştiği tespit edilmiştir.  $S-B_{\chi^2}$  fark derecesinin manidar olup olmadığını tespit etmek için  $T_s$  değeri hesaplanmış ve bu değer 4910.97 olarak bulunmuştur.  $T_s$  değerinin  $\chi^2$  dağılımı tablosundaki kritik değerden büyük olduğu tespit edilmiştir,  $\chi^2_{fark}(363)=394.626$ ,  $p>.05$ . Bu durumda modeller arasındaki farkın manidar olduğu ifade etmek mümkündür.  $\Delta CFI$  değeri -0.12 ( $<-.01$ ) ve  $\Delta SRMR$  ( $\geq .025$ ) değerinin de manidar olduğunu belirlenmiştir. Bu durumda zayıf faktöriyel değişmezlik ile güçlü faktöriyel değişmezlik arasındaki farkın manidar olduğu ifade edilebilir. Güçlü faktöriyel değişmezlik modelinin iyi uyum göstermesi nedeniyle ölçme değişmezliğinin sağlanamadığı ifade edilebilir.

Model 3 (güçlü faktöriyel değişmezlik) ve hata varyanslarının serbest, faktör yükleri ve faktör korelasyonlarının sabit olduğu Model 4 (katı faktöriyel değişmezlik) karşılaştırıldığında  $\Delta\chi^2/\Delta sd$  oranı,  $\Delta CFI$  ve  $\Delta SRMR$  değerlerinin değişiklik göstermediği tespit edilmiştir. Bu durumda Model 3 ve Model 4 arasında manidar bir farkın olmadığını söylemek mümkündür.

Tüm model karşılaştırmalarına ait bulgular değerlendirildiğinde karşılaştırılan dört model içerisinde en iyi çalışan model güçlü faktöriyel değişmezlik (Model 4) modelidir. Bu doğrultuda PISA 2012 matematik okuryazarlığının faktör yapısının OECD üyesi ülkeler için eşit olmadığı ve ölçme değişmezliğinin sağlanmadığı kararı alınmıştır.



## BÖLÜM V

### SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde araştırmanın bulgularına ilişkin sonuçlara ve bu sonuçlara dayalı olarak öneriler yer almaktadır.

#### 5.1. Sonuç ve Tartışma

PISA 2012 matematik okuryazarlığının OECD üyesi ülkeler ve Türkiye'deki cinsiyet ve bölgelere ölçme değişmezliğine ilişkin sonuçlar maddeler halinde aşağıda verilmiştir.

1. PISA 2012 uygulaması matematik okuryazarlığının cinsiyet gruplarına göre ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemek için öncelikle test istatistikleri, normallik ve güvenilirlik hesaplamaları yapılmıştır. Bu hesaplamalar sonucunda veri setinin normale yakın bir dağılım gösterdiği sonucuna ulaşılmıştır. Güvenilirlik hesaplamaları sonucunda ise maddelere ilişkin KR-20 iç tutarlılık güvenilirlik katsayısı, madde sayısının az olması sebebiyle yüksek değerlere sahip olmamakla birlikte kabul edilebilir düzeye yakın olduğu sonucuna varılmıştır.

Cinsiyet grupları için kovaryans matrislerinin eşitliği testi sonuçlarına göre, gruplara ait kovaryans matrisleri arasında orta düzeyde bir uyumun olduğu belirlenmiştir.

PISA 2012 matematik okuryazarlığının cinsiyet gruplarına göre ölçme değişmezliği çoklu-grup doğrulayıcı faktör analizi ile dört model temel alınarak test edilmiştir. Temel model olarak ele alınan Model 1 faktör yapılarının eşit olduğu hipotezi ile kurulmuştur. Model 1'e ait uyum indeksleri kabul edilebilir düzeyde çıkmıştır. Model 1'e alternatif olarak kurulan Model 2, Model 3 ve Model 4'ün yuvalanmış model (nested) temel alınarak karşılaştırılması sonucunda, uyumun Model 1'e göre kötüleştiği belirlenmiştir. Cinsiyet gruplarına ait faktör yapılarının eşit olduğu hipotezi ile kurulan Model 1 en iyi çalışan model olarak belirlenmiştir. Elde edilen bu bulgular çerçevesinde, PISA 2012 matematik okuryazarlığının Türkiye'deki

cinsiyet grupları arasında ölçme değişmezliğinin sağlandığı sonucuna ulaşılmıştır.

2. PISA 2012 uygulaması matematik okuryazarlığının Türkiye'deki bölgelere göre (Bölge 1, Bölge 2, Bölge 3) ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemek için test istatistikleri, normallik ve güvenilirlik hesaplamaları yapılmıştır. Bu hesaplamalar sonucunda veri setinin normale yakın bir dağılım gösterdiği belirlenmiştir. Güvenilirlik hesaplamaları sonucunda ise maddelere ilişkin KR-20 iç tutarlılık güvenilirlik katsayısının, kabul edilebilir düzeye yakın olduğu sonucuna varılmıştır.

Bölge 1, Bölge 2 ve Bölge 3'e göre kovaryans matrislerinin eşitliği testi sonuçlarına göre, gruplara ait kovaryans matrisleri arasında yüksek bir uyumun olduğu tespit edilmiştir.

PISA 2012 matematik okuryazarlığının bölgelere göre ölçme değişmezliği çoklu-grup doğrulayıcı faktör analizi ile dört model temelinde test edilmiştir. Temel model olarak ele alınan Model 1 faktör yapılarının eşit olduğu hipotezi ile kurulmuştur. Model 1'e ait uyum indeksleri kabul edilebilir düzeyde çıkmıştır. Model 1'e alternatif olarak kurulan Model 2, Model 3 ve Model 4'ün yuvalanmış model (nested) temel alınarak karşılaştırılması sonucunda, uyumun Model 1'e göre kötüleştiği belirlenmiştir. Bölge 1, Bölge 2 ve Bölge 3'e ait faktör yapılarının eşit olduğu hipotezi ile kurulan Model 1 en iyi çalışan model olarak belirlenmiştir. Elde edilen bu bulgular çerçevesinde, PISA 2012 matematik okuryazarlığının Türkiye'deki bölgelere göre ölçme değişmezliğinin sağlandığı sonucuna ulaşılmıştır.

3. PISA 2012 matematik okuryazarlığının OECD üyesi ülkeler arasında ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemek için test istatistikleri, normallik ve güvenilirlik hesaplamaları yapılmıştır. Bu hesaplamalar sonucunda veri setinin normale yakın bir dağılım gösterdiği kabul edilmiştir. Güvenilirlik hesaplamaları sonucunda ise ülkeler ilişkin KR-20 iç tutarlılık güvenilirlik katsayısının bazı ülkelerin kabul düzeyinin altında sonuç gösterdiği, bazı ülkelerin kabul düzeyinde değer gösterdiği tespit edilmiştir.

Ülke grupları arasında kovaryans matrislerinin uyum düzeyini belirleyebilmek için yapılan kovaryans matrislerinin eşitliği testi sonucunda, ülkeler arasında kovaryans matrislerine ait modelin iyi uyum göstermediği sonucuna varılmıştır.

PISA 2012 matematik okuryazarlığının OECD üyesi ülkelere göre ölçme değişmezliğinin çoklu-grup doğrulayıcı faktör analizi ile dört model temel alınarak test edilmesi sonucunda temel model olarak kurulan Model 1'e ait uyum indekslerinin kabul edilebilir düzeyde olmadığı ve modelin doğrulanmadığı tespit edilmiştir. Bu durumda OECD ülkelerine göre yapısal değişmezliğin sağlanamadığı ifade edilebilir. Model 1'e alternatif olarak kurulan Model 2, Model 3 ve Model 4'ün yuvalanmış model (nested) temel alınarak karşılaştırılması sonucunda, Model 3'e ait uyumun oldukça iyileştiği belirlenmiş ve güçlü faktöriyel değişmezliğin sağlandığı tespit edilmiştir.

PISA 2012 matematik okuryazarlığında OECD üyesi ülkelere göre en iyi çalışan modelin Model 3 (güçlü faktöriyel değişmezlik) olduğu ve ölçme değişmezliğinin yalnızca güçlü faktöriyel değişmezlik düzeyinde sağlandığı sonucuna ulaşılmıştır.

PISA 2012 Matematik okuryazarlığının değişmezliğinin 8 no'lu kitapçık üzerinden değerlendirildiği bu araştırmanın bulguları, Türkiye'deki cinsiyet ve bölgeler arasında ölçme değişmezliğinin sağlandığını ancak OECD üyesi ülkeler arasında ölçme değişmezliğinin sağlanmadığını göstermektedir.

Cinsiyet değişkeni ele alındığında elde edilen bulgular, Kıbrıslıoğlu'nun (2015), PISA 2012 Uygulaması matematik öğrenme modelinin Türkiye, Çin-Şangay ve Endonezya ülkelerinde cinsiyete göre ölçme değişmezliğinin sağlandığı çalışmasıyla aynı doğrultuda olduğunu göstermektedir. Bununla birlikte, Öğretmen'in (2010) 1999 TIMSS-R uygulamasına katılan öğrencilerin fen bilimleri alanındaki başarılarını açıkladığı düşünülen çalışmada, cinsiyete göre sadece metrik değişmezliğe ilişkin modelin doğrulandığını, Uyar ve Doğan'ın (2014), PISA 2009 uygulaması üzerinden yürüttükleri çalışmaları, Türkiye örneğinde öğrenme stratejileri modelinin cinsiyete göre yapısal ve metrik değişmezliğe ilişkin modellerin doğrulandığını göstermektedir.

Bulgular bölge değişkeni ele alınarak değerlendirildiğinde, Bahadır'ın (2012), PISA 2009 uygulaması okuma becerisi modelinin coğrafi bölgeler arasında ölçme değişmezliğinin sağlandığını belirlediği araştırması ile Uyar ve Doğan'ın (2014), PISA 2009 Türkiye uygulaması öğrenme stratejileri modelinin istatistiksel bölgelere göre ölçme değişmezliğinin sağlandığını belirledikleri araştırma sonuçları ile paralellik göstermektedir.

Ülke değişkenine ilişkin sonuçlar değerlendirildiğinde, geniş ölçekli sınavlarda kullanılan ölçme araçlarının farklı dillere ve kültürlere uyarlamasının sınırlılık oluşturduğu söylenebilir. Bazı çalışmalar farklı dil ve kültürlere sahip bireyler arasında kullanılan ölçme araçlarında ölçme değişmezliğinin sağlanamadığını raporlamıştır (Ercikan ve Koh 2005; Öğretmen, 2006; Wu ve diğerleri 2007; Marsh ve diğerleri, 2013; Kıbrıslıoğlu, 2015; Karakoç Alatlı, 2016). Bununla birlikte, bazı araştırma sonuçları farklı ülkelerde uygulanan geniş ölçekli çalışmalarda farklı ülkeler arasında ölçme değişmezliğinin sağlandığını raporlamışlardır (Marsh ve diğerleri, 2006; Akyıldız, 2009). Bununla birlikte, geniş ölçekli sınav uygulamalarının sonuçları değerlendirilirken ölçme değişmezliğinin sağlanamıyor olması, özellikle ülkeler arası karşılaştırmalar ve yorumlamalar yapılırken göz önünde bulundurulmalıdır.

Tüm sonuçlar değerlendirildiğinde, PISA 2012 Türkiye uygulamasında matematik okuryazarlığına ait elde edilen sonuçların cinsiyet ve bölgeler için karşılaştırılabilir ve yorumlanabilir olduğu belirlenmiştir. Fakat farklı kültür ve dillere sahip katılımcılardan elde edilen sonuçların karşılaştırılabilirliği ve buna bağlı olarak yapılan yorumlar ölçme değişmezliğinin sağlanamaması yani uygulanan ölçme aracının tüm katılımcılar için aynı anlamı ifade etmemesi nedeniyle tartışmaya açık hale gelmektedir. Geniş ölçekli olarak yürütülen ölçme ve değerlendirme uygulamalarında, ölçülen özelliğin ya da yapılan uygulamanın türüne göre ölçme değişmezliğinin farklılık gösterdiğini, elde edilen sonuçlara göre karşılaştırma ve yorumlama yapılmadan önce ölçme değişmezliği çalışmalarının yapılması gerekliliği göz ardı edilmemelidir.

## 5.2. Öneriler

Araştırma süreci ve elde edilen sonuçlar değerlendirildiğinde, araştırma ve uygulamaya dönük olarak şu önerilere yer verilebilir.

- 1) Araştırmada özellikle ülkeler arasında ölçme değişmezliğinin sağlanamadığının belirlenmiş olması, farklı gruplar arasında karşılaştırmalar yapılmadan önce yapılan yorumların anlamlı ve daha sağlıklı olabilmesi için uygulanan ölçme aracına ilişkin ölçme değişmezliği çalışmalarının yapılmasının oldukça önem taşıdığını göstermiştir.
- 2) Bu araştırma yalnızca PISA 2012 uygulaması 8 no'lu kitapçıkta bulunan matematik okuryazarlığına ilişkin maddeler üzerinden yürütülmüştür. Sonuçların genellenebilmesi için gelecek araştırmalarda farklı kitapçıklarda ya da farklı başarı alanlarına (okuma becerileri ve fen okuryazarlığı) ilişkin ölçme değişmezliği incelenebilir.
- 3) Araştırma sadece OECD ülkelerine ait veri seti üzerinden yürütülmüştür. Gelecek çalışmalarda tüm ülkeler ve diğer demografik değişkenlerin dâhil edildiği araştırmalar yapılabilir.
- 4) Bu araştırmada ölçme değişmezliği çoklu-grup doğrulayıcı faktör analizi kullanılarak test edilmiştir. Gelecek araştırmalarda farklı yöntemler kullanılarak ölçme değişmezliği test edilirse, yöntemler arası karşılaştırmalar yapılması ve hangi yöntemin daha etkili olduğuna karar verilmesi mümkün olabilir.

## KAYNAKÇA

- Akyıldız, M. (2009). PIRLS 2001 Testinin Yapı Geçerliğinin Ülkelerarası Karşılaştırılması *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 6(1), 18-47.
- American Psychological Association. (1966). *Standarts for Educational and psychological testing*. Washington, DC: Author.
- Anastasi, A. (1968). *Psychological testing* (Third Edition). Oxford: Macmillan
- Bahadır, E. (2012). *Uluslararası öğrenci değerlendirme programına (PISA 2009) göre Türkiye'deki öğrencilerin okuma becerilerini etkileyen değişkenlerin bölgelere göre incelenmesi* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Bailey, K.D. (1982). *Methods of social research*. (Second Edition). New York: The Free Press.
- Balcı, A. (2013). *Sosyal bilimlerde araştırma yöntem, teknik ve ilkeler* (10. baskı). Ankara: Pegem Akademi
- Barrett, P. (2007). Structural equation modelling: adjudging model fit. *Personality and individual differences*, 42 (5), 815-24.
- Başusta, N. B. U. (2010). Ölçme Eşdeğerliği. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2).
- Bentler, P. M. and Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3), 588.
- Bollen, K. A. (1990). Overall fit in covariance structure models: two types of sample size effects. *Psychological bulletin*, 107(2), 256.
- Borsboom, D., Mellenbergh, G., & Van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061–1071.

- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Newyork: Guilford Publications.
- Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliřtirmede kullanımı. *Kuram ve uygulamada eğitim yönetimi*, 32(32), 470-483.
- Byrne, B. M. and Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, 34(2), 155-175.
- Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press.
- Campbell, D. T. and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), 81.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling*, 14(3), 464-504.
- Cheung, G. W. and 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. W. and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2), 233-255.
- Crocker, L. and Algina, J. (1986). *Introduction to classical and modern test theory*. Toronto: Holt, RineHart, and Winston, Inc.
- Cronbach, L. J. and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281.
- Çokluk, Ö., Şekerciođlu, G., ve Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok deđişkenli istatistik: SPSS ve LISREL uygulamaları*. Pegem Akademi.
- Ercikan, K. and Koh, K. (2005). Examining the construct comparability of the English and French versionf of TIMSS. *International Journal of Testing*, 5, 23-35.

- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural equation modeling: a multidisciplinary journal*, 6(1), 56-83.
- Floyd, F. J. and Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment*, 7(3), 286.
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical care*, 44(11 Suppl 3), S78-S94.
- Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! Testing! One, two, three—testing the theory in structural equation models!. *Personality and Individual Differences*, 42(5), 841-850.
- Horn, J. L. and McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3), 117-144.
- Hu, L.T. and Bentler, P.M. (1999). Cut off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6 (1), 1-55.
- Jöreskog, K. and Sörbom, D. (1993). Lisrel 8. Chicago: Scientific Software International Inc.
- Karakoç Alatlı, B. (2016). *Uluslararası Öğrenci Değerlendirme Programı (PISA - 2012) okuryazarlık testlerinin ölçme değişmezliğinin incelenmesi*. (Yayımlanmamış Yüksek Lisans Tezi). Ankara Üniversitesi, Ankara.
- Kelley, T. L. (1927). *Interpretation of educational measurements*. New York: Macmillan.
- Kenny, D. A. and McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural equation modeling*, 10(3), 333-351.



- Kıbrıslıođlu, N. (2015). *PISA 2012 Matematik Öğrenme Modelinin Kültürlere ve Cinsiyete Göre Ölçme Deđişmezliğinin İncelenmesi: Türkiye-Çin (Şangay)-Endonezya Örneđi* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (Second Edition) New York:Guilford Press.
- Knight, G. P. and Hill, N. E. (1998). Measurement equivalence in research involving minority adolescents. *Studying Minority Adolescents: Conceptual, Methodological, And Theoretical Issues*, 183-210.
- Kođar, H. (2010). *Farklı örneklem büyüklüklerinde uçdeđerlerle başetme yöntemlerinin puanların geçerlik ve güvenirliliđine etkisi*. (Yayımlanmamış Yüksek Lisans Tezi). Ankara Üniversitesi, Ankara.
- Linn, R.L. and Gronlund, N.E. (1995). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Merrill.
- Marsh, H. W., Hau, K. T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, 6(4), 311-360.
- Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J., Abdelfattah, F., Leung, K. C., ... & Parker, P. (2013). Factorial, convergent, and discriminant validity of timss math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology*, 105(1), 108.
- Maruyama, G. M. (1998). *Basics of structural equation modeling* (First Edition). CA: Sage Publications. Inc.
- Matsumoto, D. and Van de Vijver, F. J. (Eds.). (2011). *Cross-cultural research methods in psychology*. NY:Cambridge University Press.
- McDonald, R.P. and Ho, M.-H.R. (2002). Principles and practice in reporting statistical equation analyses. *Psychological Methods*, 7 (1), 64-82.

- McIntosh, C. (2006), "Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007). *Personality and Individual Differences*, 42 (5), 859-67.
- MEB. (2015). *PISA 2012 araştırması ulusal nihai raporu*. Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). Washington, DC: American Council on Education and National Council on Measurement in Education.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Miles, J. and Shevlin, M. (2007). A time and a place for incremental fit indices. *Personality and individual differences*, 42(5), 869-874.
- Mulaik, S.A., James, L.R., Van Alstine, J., Bennet, N., Lind, S., and Stilwell, C.D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105 (3), 430-45.
- Murphy, K. R., and Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th Edition). Upper Saddle River, NJ: Prentice Hall
- Mushquash, C. J. and Bova, D. L. (2007). Cross-cultural assessment and measurement issues. *Journal on Developmental Disabilities*, 13(1), 53-65.
- Nunnally, J. C. and Bernstein, I. H. (1994). *Psychometric theory*. (Third Edition). New York: McGraw-Hill, Inc.
- OECD (2013a). PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I), PISA, OECD Publishing. <http://dx.doi.org/10.1787/9789264201118-en>

- OECD (2013b). PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, PISA, OECD Publishing.<http://dx.doi.org/10.1787/9789264190511-en>
- OECD. (2014). PISA 2012 Results in Focus: What 15-year-old Students Know and What They can Do with What They Know.
- Öğretmen, T. (2006). *Uluslararası Okuma Becerilerinde Gelişim Projesi (PIRLS) 2001 Testinin psikometrik özelliklerinin incelenmesi: Türkiye-Amerika Birleşik Devletleri örneği*. (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Raykov, T. and Marcoulides, G. A. (2008). *An Introduction to Applied Multivariate Analysis (First Edition)*. NY: Taylor & Francis Group.
- Rosenthal, R. and Rosnow, R. L. (2008). *Essentials of behavioral research: Methods and data analysis (3rd ed.)*. New York:McGraw-Hill.
- Salzberger, T.,Sinkovics, R. R., & Schlgelmich, B. B. (1999).Data equivalence in cross-cultural research: a comparison of classical test theory and latent trait theory based approaches. *Australasian Marketing Journal*, 7 (2), 23-38.
- Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, 45, 83-117.
- Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher*, 36(8), 477-481.
- Spini, D. (2003). Measurement equivalence of 10 value types from the schwartz value survey across 21 countries. *Journal of Cross-Cultural Psychology*, 34 (1),3-23.
- Sümer, N. (2000). Yapısal eşitlik modelleri: Temel kavramlar ve örnek uygulamalar. *Türk psikoloji yazıları*, 3(6), 49-74.
- Tabachnick, B.G. and Fidell, L.S. (2013). *Using multivariate statistics (6th ed.)*. New Jersey: Pearson.

- Tucker, K. L., Ozer, D. J., Lyubomirsk, S. & Boehm, J. K. (2006). Testing for measurement invariance in the satisfaction with life scale: A comparison of Russians and North Americans. *Social Indicators Research*, 78: 341–360.
- Tyson, H.E. (2004). Ethnic differences using behavior rating scales to assess the mental health of children: A conceptual and psychometric critique. *Child Psychiatry and Human Development*, 34 (3), 167-201. 11.
- Uyar, Ş. ve Doğan, N. (2014). PISA 2009 Türkiye örnekleminde öğrenme stratejileri modelinin farklı gruplarda ölçme değişmezliğinin incelenmesi. *Uluslararası Türk Eğitim Bilimleri Dergisi*, 2, 30-43.
- Uzun, B. ve 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.
- Van de Vijver, F. J. R., and Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage
- 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 (1), 4-70.
- Van de Vijver, F. J. R. and Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R. K. Hambleton, P. F. Merenda, ve C. D. Spielberger (Eds.), *Adapting Educational and Psychological Tests For Cross-Cultural Assessment* (pp. 39-63). Mahwah, NJ: Lawrence Erlbaum.
- Weston, R. and Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, 34(5), 719-751.
- Wheaton, B., Muthen, B., Alwin, D. F. & Summers, G. F. (1977). Assessing reliability and stability in panel models. *Sociological methodology*, 8(1), 84-136.
- Widaman, K. F. and Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The*

*science of prevention: Methodological advances from alcohol and substance abuse research*, 281-324.

Wu, D. A., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, 12(3), 1-26.

Yılmaz, V. (2004). Lisrel ile yapısal eşitlik modelleri: Tüketici şikâyetlerine uygulanması. *Sosyal Bilimler Dergisi*, 1.



## EKLER

### Ek 1: Özgeçmiş

<b>Kişisel Bilgiler</b>	
<b>Ad-Soyad</b>	Merve AYVALLI
<b>Doğum Tarihi</b>	27.10.1989
<b>e-mail</b>	merveayvalli@gmail.com
<b>Eğitim Bilgileri</b>	
<b>Lisans</b>	Gazi Üniversitesi, Gazi Eğitim Fakültesi, İlköğretim Matematik Öğretmenliği, 2009-2013
<b>İş Tecrübesi</b>	
<b>Çalıştığı Kurumlar</b>	Akdeniz Üniversitesi Eğitim Fakültesi, Araştırma Görevlisi, 2014-Devam ediyor
<b>Akademik Çalışmalar</b>	
<b>Yayınlar</b>	Bıçak B., Karataş S., Ayvalli M., "High School Teachers' Scientific Research Literacy Habits In Terms Of Their Professional Development", 3. Uluslararası Eğitim Programları ve Öğretim Kongresi, ADANA, TÜRKİYE, 22-24 Ekim 2015, vol.1, no.1, pp.36-41
	Güzeller C.O., Ayvalli M., Gök N., "An Investigation Of Psycho-Social Variables In Predicting Smartphone Addiction Among University Students", Intellectual Capital & Education, VARŞOVA, POLONYA, 4-6 Aralık 2014, pp.---
	Gök N., Ayvalli M., "İngilizce Öğretmen Adaylarının Web 2.0 Araçlarını Kullanma Eğilimlerinin Çeşitli Değişkenler Açısından İncelenmesi", 3rd World Conference on Educational and Instructional Studies- WCEIS, ANTALYA, TÜRKİYE, 6-8 Kasım 2014, pp.39-39
	Şekercioğlu G., Çetin M., Ayvalli M., "Okul Öncesi Öğretmenlerinin Formal Ve Informal Ölçme Yöntem Ve Tekniklerini Tanıma Ve Uygulama Durumları", Uluslararası Erken Çocuklukta Müdahale Kongresi, ANTALYA, TÜRKİYE, 3-6 Nisan 2014, pp.231-231
<b>Projeler</b>	
<b>AB Çerçeve Programı Projesi</b>	Resources for Inclusion, Diversity and Equality (RIDE), 2015-2-UK01-KA205-014061, Yönetici, 2015- Devam Ediyor

## Ek 2: İntihal Raporu



Uploaded 1 document successfully

Search

Trash

Move selected to...

Move

My Folders

My Folders

1234

eee1234

9C

mtdadav

My Documentsaaa

svabaci

zerbay

Trash

1234

Title

PISA 2012 MATEMATİK OKUR-YAZARLIĞI TESTİNİN ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ

Documents

Report Author Processed Actions

4%

Merve Sep 20, 2016  
ANVALLI 10:45:17 AM

page 1 of 1

Sharing Settings

Submit a document

81,412 Pages remaining

[Upload a File](#)

[Zip File Upload](#)

[Cut & Paste](#)

View: [Recent Uploads](#)