

AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
YÜKSEK LİSANS PROGRAMI

PISA 2012 OKUMA BECERİLERİ ÖLÇEĞİNİN, UYGULAMA
DİLİ DOĞRULTUSUNDA BELİRLENEN ÜLKELER ARASINDA
ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

Eda CEYHAN

ANTALYA, 2019

AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
YÜKSEK LİSANS PROGRAMI

PISA 2012 OKUMA BECERİLERİ ÖLÇEĞİNİN, UYGULAMA
DİLİ DOĞRULTUSUNDA BELİRLENEN ÜLKELER ARASINDA
ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

Eda CEYHAN

Danışman:

Dr. Öğr. Üyesi Güçlü ŞEKERCİOĞLU

ANTALYA, 2019

DOĐRULUK BEYANI

Yüksek lisans tezi olarak sunduĐum bu alıřmayı, bilimsel ahlak ve geleneklere aykırı dűşecek bir yol ve yardıma bařvurmaksızın yazdıĐımı, yararlandıĐım eserlerin kaynakalardan gösterilenlerden oluřtuĐunu ve bu eserleri her kullanımında alıntı yaparak yararlandıĐımı belirtir; bunu onurumla doĐrularım. Enstitü tarafından belli bir zamana baĐlı olmaksızın, tezimle ilgili yaptıĐım bu beyana aykırı bir durumun saptanması durumunda, ortaya ıkacak tüm ahlaki ve hukuki sonuçlara katlanacaĐımı bildiririm.

02.08.2019

Eda CEYHAN

İMZA ONAY SAYFASI

T.C.

AKDENİZ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Eda Ceyhan'ın bu çalışması **05.07.2019** 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 : Dr. Öğr. Üyesi Hakan Koğar
(Akdeniz Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü)



Üye : Dr. Öğr. Üyesi Neşe Öztürk Gübeş
(Mehmet Akif Ersoy Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü)



Üye (Danışman) : Dr. Öğr. Üyesi Güçlü Şekercioğlu
(Akdeniz Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü)



YÜKSEK LİSANS TEZİNİN ADI: PISA 2012 Okuma Becerileri Ölçeğinin Uygulama Dili Doğrultusunda Belirlenen Ülkeler Arasında 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.

Doç. Dr. Ramazan KARATAŞ

Enstitü Müdürü

TEŐEKKÜR

İlk olarak bu alıőmanın ortaya ıkmasında büyük rol oynayan ve alıőma sürecinde sağladığı destek ve önerileriyle bu alıőmaya katkı sağlayan danışmanım Dr. Öğr. Üyesi Güçlü Şekerciođlu'na ok teőekkür ederim. İstatistik alanında bana kazandırdığı bilgi ve uygulamalar sayesinde bu alıőmayı tamamladığım için ve ileride yapacağım alıőmalar için kendisine minnettarım.

Bu süreçte bana olan inanlarından ve desteklerinden dolayı aileme de ok teőekkür ederim. Babam Mehmet Ceyhan'ın benden beklentilerini bir miktar da olsa yerine getirebilmenin mutluluđu içindeyim.

ÖZET

PISA 2012 OKUMA BECERİLERİ ÖLÇEĞİNİN, UYGULAMA DİLİ DOĞRULTUSUNDA BELİRLENEN ÜLKELER ARASINDA ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ

Ceyhan, Eda

Tez Yöneticisi: Dr. Öğr. Üyesi Güçlü Şekercioğlu
Eğitim Bilimleri Anabilim Dalı, Yüksek Lisans Tezi
Temmuz 2019, 92 sayfa

Araştırmanın amacı, PISA 2012 okuma becerileri testinin farklı ülkeler arasında ölçme değişmezliğinin incelenmesidir. Ülkeler belirlenirken, ölçek dili dikkate alınmıştır. Araştırmada, sırasıyla, aynı dili kullanan ülkeler ve farklı dili kullanan ülkeler arasındaki ölçme değişmezliği incelenmiştir.

PISA uygulaması yapılmadan önce ölçekler ilk olarak İngilizce ve Fransızca olarak geliştirilmekte olup, aynı dil formlarının birden fazla ülkede kullanılmasından dolayı, çalışmaya ilk olarak; İngilizce ve Fransızca kitapçıklarının kullanıldığı ülkeler dâhil edilmiştir. İngilizce ve Fransızca dillerinin kullanan ülkeler ayrı ayrı, kendi aralarında incelenmiş olup, İngilizce'nin kullanıldığı ülkeler; Amerika, Avustralya, İngiltere, İrlanda, Kanada, Singapur ve Yeni Zelanda olarak, Fransızca'nın kullanıldığı ülkeler de, Belçika, Fransa, İsviçre, Kanada, Lüksemburg olarak belirlenmiştir. Araştırmaya dâhil edilen bu gruplarla hedef; aynı dilin kullanıldığı ülkeler arasında ölçme değişmezliğinin sağlanıp sağlanmadığının tespit edilmesidir. Bir diğer hedef ise farklı dili kullanan kültürlerin incelenmesi olup; Türkçe kitapçık ile dünya çapında en yaygın konuşulan dillerin (Mandarin, İspanyolca, İngilizce) kitapçıkları karşılaştırılmıştır. Her bir dili temsilen, 2 numaralı kitapçığın uygulandığı, o dilin en çok konuşulduğu ülke seçilmiş ve neticede; çalışmada, Amerika, İspanya, Tayvan-Çin ve Türkiye örneklemi kullanılmıştır. Bu doğrultuda, çalışmanın genel amacı; PISA 2012 okuduğunu anlama becerisi ölçeğinin faktör yapısının, aynı ya da farklı dillerin konuşulduğu kültürler arası eşitliğinin test edilmesidir.

Araştırma, kayıp verilen atılmasıyla, İngilizce konuşan ülkeler grubu için 4390, Fransızca konuşan ülkeler grubu için 1210 ve farklı dilleri konuşan grup için 2668 öğrenci verisiyle ve 2 numaralı kitapçıkta bulunan okuma becerisi ölçeğinde

kullanılan 28 maddeyle gerçekleştirilmiştir. İlk aşamada, 28 maddeden oluşan bu modelin yapısı, bütün gruplarda DFA (Doğrulayıcı Faktör Analizi) ile doğrulanmış olup, devamında gruplar arası ölçme değişmezliğinin sağlanıp sağlanmadığını ölçmeye yönelik olarak ÇGDFA (Çok Gruplu Doğrulayıcı Faktör Analizi) yapılmıştır. Analiz sonucunda aşamalı olarak oluşturulan değişmezlik modelleri incelenerek, farklı ülke grupları için farklı sonuçlara ulaşılmıştır. İngilizcenin kullanıldığı gruplar arasında, yapısal faktöriyel değişmezlik sağlanmış, fakat model 2 ve model 3 model daha iyi çalıştığı için ölçme değişmezliğinin sağlanmadığı tespit edilmiştir. Fransızca formun kullanıldığı gruplar da sadece faktör yapıları açısından benzemekte olup model 2 daha iyi çalıştığı için ölçme değişmezliği sağlanamamıştır. Son olarak, farklı dilleri kullanan grupların karşılaştırıldığı analizde de, zayıf faktöriyel değişmezliğin sağlandığı, fakat model 3'ün daha iyi çalışması dolayısıyla, söz konusu gruplar arasında ölçme değişmezliğinin var olmadığı belirlenmiş olup, bu gruplar arasında yapılan karşılaştırmaların anlamlı olmayacağı bulgusuna ulaşılmıştır. PISA 2012 Okuma becerisi ölçeği başarı puanlarının sözü edilen ülkeler arasında karşılaştırılmaz olduğu ve farklılıkların sebebinin öğrenci performansı değil, çeşitli yanlılıklar olabileceği ortaya koyulmuştur.

Anahtar Sözcükler: Ölçme değişmezliği, doğrulayıcı faktör analizi, çok gruplu doğrulayıcı faktör analizi

ABSTRACT

ASSESSING MEASUREMENT INVARIANCE OF PISA 2012 READING LITERACY SCALE AMONG THE COUNTRIES DETERMINED IN ACCORDANCE WITH THE LANGUAGE OF APPLICATION

Ceyhan, Eda

Supervisor: Assoc. Prof. Dr. Güçlü Şekercioğlu
Department of Educational Sciences, Postgraduate Thesis
July 2019, 92 pages

The aim of the study is to examine the measurement invariance of PISA 2012 reading skills test among different countries. When determining countries, the language of scale was taken into consideration. In this study, the measurement invariance between the countries using the same language and the countries using different languages were examined respectively.

Before the implementation of the PISA, the scales were first developed in English and French and the same language forms are used in different countries, therefore, firstly, countries where booklets in English and French are used are included in this study. The countries that use the English and French languages have been examined separately among themselves and the countries where English is used are determined as America, Australia, England, Ireland, Canada, Singapore and New Zealand, and the countries where French is used are determined as Belgium, France, Switzerland, Canada and Luxembourg. With these groups included in the research, the target is to determine whether there is measurement invariance between countries where the same language is used. Another objective is to examine cultures using different languages; within this scope the Turkish booklet and the booklets of the most widely spoken languages worldwide (Mandarin, Spanish, English) were compared. Representing each language, the country where the booklet number 2 is applied and where the particular language is used at most is chosen and as a result; in the study, America, Spain, Taiwan-China and Turkey sample were used. In this respect, the general purpose of the study is testing the equality of factor structure of PISA 2012 reading comprehension scale between the countries using the same or different languages.

The study was carried out with 4390 student data for the group of English speaking countries, 1210 student group for the French speaking group and 2668

student data for the group speaking different languages and 28 items used in the reading skill scale in booklet 2. In the first stage, the structure of this model consisting of 28 items was confirmed by CFA (Confirmatory Factor Analysis) in all groups, and then, in order to measure whether or not inter-group measurement invariance was achieved, Multi-Group Confirmatory Factor Analysis was performed. As a result of the analysis, the invariance models created in stages were examined and different results were obtained for different country groups. It was found that structural factorial invariance was provided among the groups where English was used, but measurement invariance was not established because model 2 and model 3 models performed better. The groups using the French form are similar only in terms of factor structures and since the model 2 performed better, the measurement invariance could not be confirmed. Finally, in the analysis comparing the groups using different languages, it was found that weak factorial invariance was established, but due to the better performance of model 3, there was no measurement invariance between these groups and the comparisons between these groups were not significant. It has been demonstrated that PISA 2012 achievement scores of the reading skill scale are incomparable among the countries mentioned and that the differences may be caused by various biases rather than student performance.

Keywords: *Measurement invariance, confirmatory factor analysis, multi-group confirmatory factor analysis*

İÇİNDEKİLER

TEŞEKKÜR	i
ÖZET.....	ii
ABSTRACT	iv
İÇİNDEKİLER	vi
TABLolar LİSTESİ.....	viii

BÖLÜM I

GİRİŞ

1.1. Problem Durumu	1
1.2. Araştırmanın Amacı	2
1.3. Araştırmanın Önemi.....	2
1.4. Sayıltı ve Sınırlılıklar	3
1.5. Tanımlar	3

BÖLÜM II

KURAMSAL ÇERÇEVE VE LİTERATÜR

2.1 Uluslararası Sınavlar ve Önemi.....	5
2.1.1 PISA	6
2.1.1.1 PISA'nın Yapısı	7
2.1.1.2 PISA Örnekleme	9
2.1.1.3 PISA'nın Hazırlanma Süreci	9
2.1.1.4 Türkiye'nin PISA Performansı	10
2.2 Yapı Geçerliliği ve Ölçme Değişmezliği	11
2.2.1 Yapısal Değişmezlik (Configural Invariance)	15
2.2.2 Zayıf Faktöriyel Değişmezlik (Metric Invariance).....	15
2.2.3 Güçlü Faktöriyel Değişmezlik (Scalar Invariance)	16
2.2.4 Katı Faktöriyel Değişmezlik (Strict Factorial Invariance)	17
2.3 Yapısal Eşitlik Modellemesi	17
2.3.1 Doğrulayıcı Faktör Analizi.....	18
2.3.2 Uyum İndeksleri	19
2.3.3 Çok Gruplu Doğrulayıcı Faktör Analizi	22
2.3.3.1 Çok Gruplu Doğrulayıcı Faktör Analizi Varsayımları	22
2.3.3.2 Çok Gruplu Doğrulayıcı Faktör Analizi Aşamaları	25
2.3.3.3 Model Karşılaştırmalarında Alternatif Uyum İndeksleri	26
2.4 Ölçme Değişmezliği ve Kültür İlişkisi	28
2.4.3 Psikolojik Ölçme Araçlarını Uyarlama Süreci	30
2.4.3.1 Ölçek Uyarlama Adımları.....	31
2.4.3.2 Uyarlama Çeşitleri	34
2.5 Ölçme Değişmezliğini Etkileyen Faktörler	35
2.6 İlgili Araştırmalar	36

BÖLÜM III

YÖNTEM

3.1	Araştırmanın Modeli	42
3.2	Araştırma Verileri	42
3.3	Verilerin Düzenlenmesi	43
3.4	Verilen Analizi.....	43
3.4.1	Varsayımların Test Edilmesi	44

BÖLÜM IV

BULGULAR

4.1	1. Alt Probleme İlişkin Veri Setlerinden Elde Edilen DFA, Kovaryans Matrisleri Eşitliği Testi ve ÇDFA Sonuçları.....	47
4.2	2. Alt Probleme İlişkin Veri Setlerinden Elde Edilen DFA, Kovaryans Matrisleri Eşitliği Testi ve ÇDFA Sonuçları.....	50
4.3	3. Alt Probleme İlişkin Veri Setlerinden Elde Edilen DFA, Kovaryans Matrisleri Eşitliği Testi ve ÇDFA Sonuçları.....	52

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

5.1	Sonuç ve Tartışma.....	56
5.2	Öneriler	59
KAYNAKÇA		62
EKLER.....		72
Ek 1: Doğrulayıcı Faktör Analizi Örnek Lisrel Sözdizimi.....		72
Ek 2: Kovaryans Matrisleri Eşitliği Testi Örnek Lisrel Sözdizimi		72
Ek 3: Ölçme Değişmezliği Modelleri Testi Model A Örnek Lisrel Sözdizimi		73
Ek 4: Ölçme Değişmezliği Modelleri Testi Model B Örnek Lisrel Sözdizimi.....		74
Ek 5: Ölçme Değişmezliği Modelleri Testi Model C Örnek Lisrel Sözdizimi		75
Ek 6: Ölçme Değişmezliği Modelleri Testi Model D Örnek Lisrel Sözdizimi		76
ÖZGEÇMİŞ.....		78
BİLDİRİM		79

TABLULAR LİSTESİ

Tablo 3.1.1 1. Alt Probleme İlişkin Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları	45
Tablo 3.1.2. 2. Alt Probleme İlişkin Test İstatistikleri, Normallik Testleri ve güvenilirlik Katsayıları	45
Tablo 3.1.3. 3. Alt Probleme İlişkin Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları	45
Tablo 4.1.1. İngilizce Dil Formlarına İlişkin Doğrulayıcı Faktör Analizi	47
Tablo 4.1.2. İngilizce Dil Formlarına İlişkin Kovaryans Matrislerinin Eşitliği.....	48
Tablo 4.1.3. Amerika, Avustralya, İngiltere, İrlanda, Kanada, Singapur ve Yeni Zelanda'ya İlişkin Çok Gruplu Doğrulayıcı Faktör Analizi	48
Tablo 4.2.1 Fransızca Dil formlarına İlişkin Doğrulayıcı Faktör Analizi.....	50
Tablo 4.2.2 Fransızca Dil Formlarına İlişkin Kovaryans Matrislerinin Eşitliği	51
Tablo 4.2.3 Belçika, Fransa, İsviçre, Kanada ve Lüksemburg'a İlişkin Çok Gruplu Doğrulayıcı Faktör Analizi	51
Tablo 4.3.1 Farklı Dil Formlarına İlişkin Doğrulayıcı Faktör Analizi.....	52
Tablo 4.3.2 Farklı Dil Formlarına İlişkin Kovaryans Matrislerinin Eşitliği	53
Tablo 4.3.3. Amerika, Tayvan ve Türkiye'ye İlişkin Çok Gruplu Doğrulayıcı Faktör Analizi	53

BÖLÜM I

GİRİŞ

Bu bölümde araştırmanın problem durumu, amacı ve önemi üzerinde durulmuştur.

1.1. Problem Durumu

Günümüzde, uluslararası düzeyde düzenlenen geniş ölçekli sınavlar daha fazla önem kazanmaktadır. Her geçen gün bu tür sınavların sayısı ve bu sınavlara katılan ülkelerin sayısı artmaktadır. Gelişmiş ve yüksek gelir sağlayan ülkelerin neredeyse tamamının katıldığı bu sınavlar, sadece ülkelerin bulunduğu sırayı göstermekle kalmayıp aynı zamanda her bir ülkenin eğitim sistemine ışık tutmakta, eğitim sisteminde hangi alanlarda eksikliklerin olduğunun ve bu eksikliklerin kaynağının belirlenmesinde büyük rol oynamaktadır. Bu denli büyük bir öneme sahip olan ölçme araçları oluşturulurken ve başka kültürlere uyarlanırken titizlikle davranılmaktadır. Fakat bu duruma ek olarak, geliştirilen ölçme araçları başka bir kültüre uyarlandığında, söz konusu ölçme aracının ölçmeyi hedeflediği özellik ya da özellikleri tam ve doğru ölçüp ölçmediğine ilişkin sorgulamaların yapılması ve geçerlilik kanıtlarının ortaya konulması gerekmektedir.

Bir testin sahip olması gereken en önemli özelliği geçerliliğidir. Geçerlilik ilk olarak 1937 yılında Garret tarafından “ölçülmek istenen özelliğin amaca uygun olarak ölçülme derecesi” olarak tanımlanmıştır (Lissitz, 2009). Ölçme aracının farklı gruplarda amacına hizmet edebilmesini sağlamak birçok işlem gerektiren bir süreçtir. Bu sürecin ardından da ölçülmesi hedeflenen özelliklerin sözü geçen ölçme aracı tarafından amaca uygun olarak ölçme yapıp yapılamadığı deneysel işlemlerle test edilmelidir. Bir ölçme aracı, ‘farklı gruplarda da aynı özelliği ölçer’ düşüncesini kabul ederek hazırlanır ve yapılan karşılaştırma ve puanlara ait analizlerin doğruluğu ancak bu kabul kanıtlanabildiğinde anlamlı hale gelir. Yapılan bir ölçmedeki ölçme sonuçlarındaki farklılık bireylerin farklı özellikler taşımasından kaynaklanabilir, fakat bu farklılığın sadece birey özelliklerinden kaynaklandığını söylemek doğru değildir (Uzun ve Öğretmen, 2010). Bir ölçme aracının geçerliliği, o ölçme aracı ile elde edilebilecek varyansın (değişkenliğin) ne kadarının, ölçme işlemi uygulanan

bireylerin ölçülmesi hedeflenen özelliklere sahip oluş derecelerindeki farklılardan kaynaklandığını gösterir (Şekercioğlu, 2009). Ölçme sonuçları, bireyin özellikleri dışında olabildiğince daha az farklı değişkenden etkilendiği zaman daha geçerli hale gelir.

Özellikle uluslararası alanda düzenlenen sınavların artmasıyla ve katılan ülkelerin performans sıralarının duyurulmasıyla, ölçme araçlarının farklı toplumlarda ne kadar güvenilir ve geçerli bir ölçme yaptığı sorgulanmaya başlanmış ve bu doğrultuda, farklı gruplara uygulanan faktör yapılarının eş değer olup olmadığını belirlemek amacıyla yapılan ölçme değişmezliği analizleri ve çalışmaları da hız kazanmıştır. Bu araştırmanın problemi, aynı dilin kullanıldığı farklı ülkeler arasında ve farklı dillerin kullanıldığı ülkeler arasında PISA (Programme for International Student Assessment) 2012 okuduğunu anlama testinin ölçme değişmezliğinin sağlanıp sağlanmadığının belirlenmesidir.

1.2. Araştırmanın Amacı

Bu çalışmanın genel amacı, PISA 2012 okuduğunu anlama becerisi testinin faktör yapısının dil değişkenine göre eşitliğinin test edilmesidir. Bu sayede genel amaç doğrultusunda aşağıdaki alt problemlere cevap aranmıştır.

(1) PISA 2012 verilerine göre okuduğunu anlama becerisi ölçeği ölçek dili İngilizce olan ve İngilizce konuşan gruplar arasında (Amerika, Avustralya, İngiltere, İrlanda, Kanada, Singapur, Yeni Zelanda) ölçme değişmezliğine sahip midir?

(2) PISA 2012 verilerine göre okuduğunu anlama becerisi ölçeği ölçek dili Fransızca olan ve Fransızca konuşan gruplar arasında (Belçika, Fransa, İsviçre, Kanada, Lüksemburg) ölçme değişmezliğine sahip midir?

(3) PISA 2012 verilerine göre okuduğunu anlama becerisi ölçeği, ölçek dilinin ve konuşulan dilin birbirinden farklı olduğu gruplar arasında (Amerika, İspanya, Tayvan, Türkiye) ölçme değişmezliğine sahip midir?

1.3. Araştırmanın Önemi

Bir ölçme aracından elde edilen puanlar doğrultusunda, kültürlerarası karşılaştırma yapılırken, ölçme aracının bütün gruplar için aynı faktör desenine sahip olduğu varsayılmakta ve kıyaslamalar ölçme değişmezliğinin sağlanıp sağlanmadığı

dikkate alınmadan gerçekleştirilebilmektedir. Özellikle ülkelerin eğitim sistemlerine yönelik olarak yapılan uygulamalarda, ölçeğin çok sayıda dile çevirisinin yapılması nedeniyle ölçme aracında herhangi bir yanlılık olmadığına ve tüm ülkelerde eşit bir ölçme yapıldığına dair kanıt sunulması büyük önem taşımaktadır. Bu sayede, ülkelere göre ölçme değişmezliğinin var olduğuna yönelik bir kanıt elde edilirse uygulamaya katılan öğrencilerin performanslarını etkileyen diğer unsurlar üzerinde durularak, eğitimdeki eşitsizlikler ve eğitim sistemindeki eksiklikler giderilmeye çalışılabilir. Ölçme değişmezliğinin sağlanamadığına yönelik bir kanıt elde edilir ise grupların performanslarına göre yapılan sıralama, çeşitli yanlılıklar ya da ölçek uyarlama problemlerinin varlığı nedeniyle geçerliliğe şüpheyle yaklaşılır.

Yapılan alanyazın taramalarına göre; bu tür uygulamalara yönelik ölçme değişmezliği çalışmalarını genellikle ölçek verilerinin oluşturduğu, performansı belirleyen madde puanları üzerinde yapılan çalışmaların daha az sayıda olduğu görülmektedir. Bu nedenle, bu çalışmanın ölçme değişmezliği kanıtı sunmasının yanı sıra, alanyazına da katkı sağlayacağı beklenmektedir.

Sonuç olarak bu çalışma, PISA 2012 okuduğunu anlama becerisi için sözü geçen ülkelere yönelik başarı ortalamalarının ve sıralamalarının doğruluğu açısından yeterli kanıt toplamak için bir gerekliliktir.

1.4. Sayıltı ve Sınırlılıklar

Araştırma kapsamında kullanılan PISA 2012 uygulamasının, katılımcılar tarafından gerçeği yansıtacak biçimde ve uygun ortam ve koşullarda cevaplandırıldığı varsayılmaktadır.

Bu araştırma PISA 2012 uygulamasının okuduğunu anlama testinin 2 numaralı kitapçığında bulunan veriler ile sınırlıdır.

1.5. Tanımlar

Bu çalışmada geçen temel kavramlar aşağıda belirtilmiştir.

Geniş ölçekli sınavlar: Ulusal ya da uluslararası çapta büyük örneklem gruplarına uygulanan ve belli bir amaç doğrultusunda bireylerin ya da belli bir grubun düzeyini saptamaya çalışan testlerdir.

Ölçme deęişmezlięi: Bir ölçme aracının psikometrik niteliklerinin, farklı gruplar için benzer olup olmadığının değerlendirilmesidir.

BÖLÜM II

KURAMSAL ÇERÇEVE VE LİTERATÜR

2.1 Uluslararası Sınavlar ve Önemi

Uluslararası alanda ülkelerin eğitim sistemini değerlendirmesi bağlamında ele alınan bu sınavlar çok eskiye dayanmamaktadır. İlk pilot çalışma 1960larda IEA (The International Association for the Evaluation of Educational Achievement) tarafından 12 ülkede, 13 yaş grubu öğrencilerinin akademik başarısını ölçmek için düzenlenmiştir (Keeves, 2011). Husen ve Tuijnman (1994)'e göre, eğitim sistemlerinin karşılaştırılmasındaki bu ampirik yaklaşım, ölçek uygulama teknikleri, ölçek geliştirme ve veri analizi kapsamındaki gelişmelerle mümkün hale gelmiştir.

IEA ve OECD (Organization for Economic Co-operation and Development) tarafından düzenlenen sınavlar, farklı ülkelerin çeşitli konu alanlarında nasıl performans sergilediğini, öğrencilerin gerçek hayata ne kadar hazır olduklarını ve eğitim sistemlerinin ne kadar başarılı ve etkili olduğunu ortaya koymaktadır. Bu sınavlar sadece ülkeler arası farklılıklara ışık tutmakla kalmaz aynı zamanda ulusal açıdan var olan öğrenciler arasındaki bireysel farklılıklara, bu farklılıkların kaynağına da ışık tutmaktadır. Eğer bu tür sınavlar, uygulanmakta oldukları ülkelere iyi bir biçimde uyarlanmışsa ve sonucunda iyi bir değerlendirmeyle doğru çıkarımlar yapılmışsa kullanılan müfredatta düzenlemelere ve iyileştirmeye gidilebilir ve eğitim sisteminin daha güçlü hale gelmesi sağlanabilir.

Türkiye'nin Uluslararası çerçevede katıldığı üç sınav bulunmaktadır. Türkiye ilk olarak 1999 yılında TIMSS (Trends in International Mathematics and Science Study), 2001 yılında PIRLS (Progress in International Reading Literacy Study) ve 2003 yılında da PISA'ya katılmıştır. TIMSS ilk olarak 1995 yılında, PIRLS 2001 ve PISA 2000 yılında uygulanmıştır (Milli Eğitim Bakanlığı (MEB), 2003; MEB, 2005; MEB, 2015; IEA, 2003). TIMSS ve PIRLS sınavları IEA tarafından düzenlenirken PISA OECD tarafından düzenlenmektedir. Her bir sınav farklı amaca hizmet etmekte, farklı yaş gruplarındaki öğrencileri değerlendirmekte ve farklı konular bazında düzenlenmektedir.

PIRLS okuma becerisini değerlendirmeye yönelik bir sınav olup, hedef grup 4. sınıftaki öğrencilerdir (IEA, 2003). TIMSS ise 4. ve 8. sınıf öğrencilerinin

matematik ve fen okuryazarlığının değerlendirildiği bir sınavdır (MEB, 2003). Araştırmanın amacının, PISA 2012 okuduğunu anlama bölümünden elde edilen sonuçların ve yapılan çıkarımların doğruluğunu incelemek olduğundan, sonraki bölümlerde PISA'nın tarihi gelişimi ve 2012 uygulaması hakkında bilgiler sunulmuştur.

2.1.1 PISA

PISA uygulaması 2000 yılından bu yana üçer yıl arayla yapılan, OECD'ye üye ve üye olmayan ülkelerin öğrencilerinin katıldığı bir sınavdır. 15 yaş grubundaki öğrencilere uygulanmakta ve öğrencilerin okuma, matematik ve fen bilimlerindeki başarısını ölçmektedir. İlk değerlendirme 2000 yılında yapılmış ve sonrasında her üç yılda bir tekrarlanmıştır (OECD, 2012). PISA'nın amacı, farklı eğitim sistemlerini değerlendirmek ve öğrencilerin gerçek hayata ne kadar hazır olduklarını saptamaktır. Bu nedenle PISA öğrencilerin okul ve okul dışında öğrendiklerini, ev ve iş hayatlarında karşılaşılabilecekleri durum ve olaylar karşısında nasıl uygulayacaklarını değerlendirmeye çalışır. Değerlendirilen bilgi sadece okulda öğrenilenlerle sınırlı değildir (Shivraj, 2014). Öğrencilerin müfredata göre ne öğrendiklerini belirlemeye çalışan bir sınav olmaktan daha çok öğrendikleriyle yaşamda neler yapabildiğini belirleyen bir sınavdır (MEB, 2011). Temel olarak ölçülen beceriler ekonomik hayata katılım için gerekli beceriler olsa da, PISA temel becerilerin yanında eleştirel düşünce, analiz, sentez ve yaratıcılık gibi farklı becerileri de ölçmektedir. Ayrıca uygulanan anketlerle öğrencilerin demografik bilgileri, okullarda uygulanan eğitim öğretim tekniklerine yönelik bilgiler vb. data da toplanarak, araştırmacılara ve kurumlara, elde edilen sonuçlarla birlikte bu verileri de kullanarak, derinlemesine analiz yapmaları ve doğru çıkarımlara ulaşmaları konusunda kaynak sağlamaktadır.

PISA, sadece uygulamaya katılan ülkelerin okuma, matematik ve fen alanındaki başarı sıralamalarını ve ortalama puanlarını vermekle kalmayıp; düşük ya da yüksek düzeyde başarı gösteren öğrencilerin oranını ve bunların cinsiyet açısından dağılımını, erkek-kız öğrenciler arasındaki puan farklılığı ile ilgili birtakım göstergeler de sunmaktadır (Özmuşul ve Kaya, 2014). Öğrencilerin üç temel konu alanındaki bilgi ve becerilerini değerlendirmenin yanında, PISA projesinde öğrencilerin öğrenme stratejileri, problem çözme becerileri ve değerlendirilen konu alanına yönelik ilgi ve tutumları da araştırılmaktadır (MEB, 2010).

PISA bilgi ve beceriyi dünyada en kapsamlı ve nitelikli bir biçimde ölçen ve değerlendiren sistemlerden birisi olarak kabul edilmektedir. Bu nedenle sonuçları kamuoyunda büyük bir yankı uyandırmakta ve eğitim sistemiyle ilgili tartışmalar başlamaktadır. PISA'nın önemli olmasının bir diğer nedeni de, bireyin ekonomik alandaki etkinliğini ve üretkenliğini belirleyen faktörleri ölçmesidir. PISA sonuçları, OECD'nin uygulamakta olduğu diğer bir program olan Uluslararası Yetişkin Becerilerini Değerlendirme Program (International Assessment of Adult Competencies - PIAAC) ile karşılaştırıldığında, PISA'da öğrencilerin gösterilen performansla, yetişkinlerinin PIAAC'da gösterdiği performans arasında güçlü bir ilişki olduğu görülmüştür (OECD, 2013b).

2.1.1.1 PISA'nın Yapısı

2000 yılından bu zamana kadar düzenli olarak her üç yılda bir matematik, fen ve okuma becerilerini ölçmeye yönelik olarak uygulanan sınavda her dönem bir beceri ağırlıklı olarak ölçülmektedir. 2000 ve 2009 yıllarında okuduğunu anlama becerisi ağırlıklı, 2006'da fen bilimleri ve 2003 ile 2012 yıllarında ise matematik becerisi ağırlıklı olarak ölçülmüştür.

PISA 2012 üç beceriyi ölçmüştür. Bunlar matematik okuryazarlığı, okuma okuryazarlığı ve fen okuryazarlığıdır. Bu becerilerin kapsamı ve tanımları OECD (2013a) tarafından aşağıdaki biçimde sunulmuştur:

Matematik Okuryazarlığı: Bireyin formüle etme, matematiği kullanma ve yorumlama kapasitesidir. Bu kapasite matematiksel olarak akıl yürütmeyi, gereken sonuca ulaşmak amacıyla matematiksel kavramları, işlemleri ve araçları kullanmayı gerektirir. Matematik okuryazarlığı sayesinde birey matematiğin evrensel önemini fark edebilir ve ihtiyaçları doğrultusunda kullanabilir. Matematik okuryazarlığı, teknolojinin her alanda kullanılması nedeniyle bilgisayar tabanlı olarak da değerlendirilebilir. Bu tür değerlendirme, üye ve katılımcı ülkelere isteğine bağlı olarak sunulmaktadır. PISA 2012'deki bilgisayar tabanlı değerlendirmede, hesap makinesi program içinde sunulmakta ve kâğıt kalem sınavlarında hesap makinesi bulundurmanın öğrenciler için faydalı olacağı hatırlatılmaktadır (OECD, 2013a).

Fen Okuryazarlığı: Bireyin fen alanında sahip olduğu bilgiyi; bilimsel gerçekleri açıklamak, yeni bilgiler edinmek ve ispata dayalı sonuçlar üretmek için kullanmasıdır. Ayrıca birey fen ve teknolojinin içinde bulunduğumuz dünyayı nasıl biçimlendirdiğinin farkında olmalıdır. İçinde bulunduğu çevreye ne kadar duyarlı

olduđu fen bilimlerine verdiđi deđerle dođru orantılıdır. Fen okuryazarlıđı bu deđeri de iinde barındırmaktadır ve bireyin dnyayı anlaması, bilimsel yaklaşımlar geliřtirmesi beklenmektedir (OECD, 2013a).

Okuma Becerileri: Bireyin bir metni basit bir biimde analiz etmesinden ok, szck bilgisine, dilbilgisine sahip olmasına, metinsel yapı ve zellikleri bilmesine ihtiya vardır. Bununla birlikte iinde bulunduđu dnya ile ilgili bilgi sahibi olarak, metni zmlerken uygun stratejileri kullanabilmelidir (OECD, 2013a).

Ayrıca 2012 yılında bazı lkelerde problem özme ve finansal okuryazarlık becerileri de llmřtr. Teknik nedenlerden dolayı problem özme becerisi btn lkelerde llememiř ve finansal okuryazarlık testi ise isteđe bađlı olarak uygulanmıřtır (OECD, 2013a).

Problem özme: özm yolu ok net belli olmayan problem durumlarında bireyin problemi anlaması ve özm yolu geliřtirmesine ynelik biliřsel srelerden geme kapasitesidir. Bireyin problem özmek iin hazırbulunuřluluk dzeyi ve isteđi nemlidir (OECD, 2013a).

Finansal Okuryazarlık: Bireyin finansal kavram ve riskleri anlaması ile bu konuda sahip olduđu bilgi ve beceridir. Toplumun ve bireylerin finansal aıdan iyi oluřları ve geliřmeleri ynnde, bireyin sahip olduđu bilgi, beceri ve motivasyonunu kullanarak, eřitli finansal konularda etkili kararlar almasıdır (OECD, 2013a).

Deđerlendirilen beceriler incelendiđinde, llen bilgi ve becerilerin gnlk hayatta yařamı devam ettirebilmek ve iinde bulunulan toplumun faydalı bir yesi olabilmek iin gerekli olan yapılardan oluřtuđu grlebilir. PISA kapsamında deđerlendirmeye alınan  temel beceri ierisinden bir tanesi ve diđer becerilerde bařarıyı yakalamak iin nemli rol oynayan beceri, okuduđunu anlama becerisidir. Okuduđunu anlama becerisi diđer becerilerle i ie ve diđer becerilerin kazanılması iin n kořuldur. rneđin matematik, sadece kurallar, semboller, řekiller ve iřlemlerin oluřturduđu bir yapı deđil, aynı zamanda iinde bir anlam btnlđ olan dzenler ve iliřkiler ađıdır (MEB, 2009). đrencilerin iki dođal sayıyı toplayabilmelerinin yanı sıra, hangi durumlarda toplama yapmanın uygun olacađını kavraması veya toplamada eldenin ne anlama geldiđini anlaması da nemsenmektedir (MEB, 2009).

Okuduđunu anlama, okuyucunun n bilgileri ile metinlerden đrendiklerini karřılařtırıp, sentez ederek yeni bir dřnceye ulařmasıdır (Akyol, 2005). Okuyucular materyali okuduktan sonra, sahip oldukları bilgi ile metinde sunulan

bilgi arasında köprü kurmalıdırlar. Eđer okuduđunu anlamada başarısız oluyorsa bir kiři bunun çeřitli nedenleri olabilir. İlk akla gelen okuyucunun ön bilgi eksikliđidir. Fakat yapılan arařtırmalara göre okumayı anlamlandırmak için, hatırlama ve dikkat mekanizmalarını, dilin kullanımını ve dođasını, konuřmanın anlaşılmasını, bireyler arası iliřkileri, sosyo-kültürel farklılıkları göz önünde bulundurmak gereklidir (Yılmaz, 2008). Bař (2013) tarafından, Türk ve Amerikan kültürünü içeren metinler kullanılarak yapılan arařtırmaya göre, ölçme aracında bulunan okuma parçaları, katılımcıların kültürel açıdan ařına oldukları, günlük hayatlarında karřılařtıkları konu ya da olaylarla ilgili olduđunda, o zaman okuduđunu anlama becerisi daha yüksek düzeyde gözlenirken, okuduđunu anlama metinleri, katılımcıların daha önce kendi kültürlerinde hiç tecrübe etmedikleri ya da gözlelemedikleri konulardan olduđunda ise okuduđunu anlama becerisi daha düşük düzeyde olmaktadır. Görüldüğü gibi, okuduđunu anlama becerisi farklı faktörlerden etkilenebilmektedir ve bu nedenle deđerlendirme konusunda oldukça titiz davranılmalıdır.

2.1.1.2 PISA Örnekleme

PISA uygulamaları için örneklem, benzer yař grubundaki öğrencilerden oluřmaktadır. Uygulamanın yapılacađı tarih itibariyle 15 yıl 3 ay ve 16 yıl 2 ay yař grubu arasındaki öğrenciler PISA öğrenci evrenini oluřtırmaktadırlar. Uygulamaya katılacak olan okullar tabakalı örnekleme yöntemiyle belirlenir (MEB, 2011). Tabakalı örneklemede, örneklem seçimine bařlamadan önce, evren, belirlenmiř özellikler açısından tabakalara ayrılır ve her tabakadan rastgele seçim yapılır. Tabakalandırma yapmadan rastgele seçim yapıldığında, bazı gruplardan hiç seçim yapılmama olasılıđı vardır (Özdemir, 2009). Tabakalar okul türleri, bölgeleri vb. ölçütler dikkate alınarak oluřturulur (MEB, 2011).

2.1.1.3 PISA'nın Hazırlanma Süreci

OECD (2012) tarafından sunulan bilgilere göre, PISA uygulaması için testler hazırlanırken, uygulamanın bütün ülkeler için anlaşılır olmasına ve güvenilir olmasına dikkat edilmekte olup, çeviri ve uyarlama sürecinin titizlikle yürütüldüğü belirtilmektedir. İlk ařamada, kültürel farklılıklar da göz önünde bulundurularak iki yönlü tercüme (double translation) iřlemi yapılmakta ve ölçek İngilizce ve Fransızca olarak geliřtirilmektedir. Ölçeklerin son halleri uzmanlar (özellikle iki dilli

uzmanlar) tarafından doğrulanıp onaylandıktan sonra, pilot çalışması yapılmaktadır. 2009 PISA uygulaması yapılmadan önce geliştirilen ölçeklerin Fransızca formu, Fransızca konuşan 5 ülkede uygulanmış ve sonuçlar doğrultusunda anket soruları, ölçek maddeleri ve okuma becerileri ölçeğinde düzenlemelere gidilmiştir. Okuma becerisi ölçeğinin bir bölümünde cümle yapısında yanlıklar olduğu gerekçesiyle çeşitli düzeltmeler yapılmış ve düzeltilmiş formdan çevirisinin yapılarak kullanılması için yeniden düzenlenen bölüm ve maddeler PISA 2012 uygulamasının madde havuzuna alınmıştır (OECD, 2012).

OECD (2014) PISA 2012 Technical Report isimli rapora göre, uygulamanın diğer dillere çevrilebilmesi konusunda ise çeviri ve uyarlama konusunda ulusal merkezlere ve görevlilere rehberlik edecek eğitimler düzenlenmekte ve oluşturulan formlar tekrar gözden geçirilmekte ve uygulanmadan önce geniş çaplı (pilot uygulamaya dâhil edilen ülkeler belirtilmemiştir) pilot uygulaması yapılmaktadır. Pilot uygulamaların sonucuna göre, ülkeler kendi ulusal formlarında yapmak istedikleri değişiklikleri bildirebilmektedirler.

Yukarıda verilen bilgiler doğrultusunda, PISA uygulamasının geçerli ve güvenilir bir ölçme işlemi yapması konusuna özen gösterildiği ve uygulama öncesinde geçerliliği ve güvenilirliği artıracak çalışmalar ve işlemler yapıldığı raporlarda dile getirilmektedir.

2.1.1.4 Türkiye'nin PISA Performansı

PISA 2012 bulguları incelendiğinde, Türkiye'nin okuduğunu anlama bölümünde 65 ülke içerisinde 475 puanla 42. sırada olduğu görülmektedir. Şangay-Çin ise tüm alanlarda ilk sıralarda yer almaktadır. Şangay'ı Hong Kong, Singapur, Japonya, Kore, Finlandiya takip etmektedir. Üst performans düzeyine ulaşmış öğrenci sayıları incelendiğinde, sıralamalarda önemli farklılıklar ortaya çıkmaktadır. Türkiye okuduğunu anlamada 10. sıraya, matematik alanında 12 ve fen alanında 17. sıraya yükselmiştir. Okuduğunu anlama bölümünde üst performans düzeyine ulaşmış öğrenci sayılarına bakıldığında Türkiye Çin'i geride bırakmıştır (MEB, 2013). Üst performans düzeyine ulaşan öğrenci sayıları, yeterlilik düzeyine ulaşan öğrenci sayısının katılan öğrenci sayısına oranlanmasıyla elde edilen yüzdenin ülkedeki bütün 15 yaş grubunda bulunan öğrenci nüfusuyla çarpılmasıyla elde edilmiştir. Fakat aynı işlem asgari performans düzeyindeki öğrenci sayılarını hesaplamak için yapıldığında, sıralamada okuduğunu anlama bölümü için 39, diğer bölümler için ise

daha alt sıralara yerleşmiştir. Bu sonuçlarda Türkiye nüfusunun yüksek olması rol oynamış olup, sıralamalar arasındaki büyük fark ülkemizde öğrenci başarıları arasında ciddi farklılıkların ve eşitsizliklerin olduğunun bir göstergesi olabilir.

PISA 2003'te ana alan olan matematik okuryazarlığında, Türkiye okullar arası eşitsizlik bakımından OECD ülkeleri arasında birinci sırada yer almaktadır (Dinçer ve Kolaşın, 2009). Türkiye'de öğrencilerin PISA 2012 matematik puanlarındaki varyansın (değişkenliğin) %62'sinin okullar arasındaki farklılıktan kaynaklandığı hesaplanmaktadır. Diğer iki beceriyi de temsilen sadece matematik okuryazarlığı için hesaplama yapılmıştır. Bu oran OECD ülkelerinde ortalama %37'dir (MEB, 2013). Bu doğrultuda, bu eşitsizliğin giderilmesine yönelik çalışmalar gerçekleştirilerek, eşitsizliğin belirleyicileri tespit edilmelidir. Öğrenci başarısındaki farklılıkları açıklamak için çeşitli değişkenler göz önünde bulundurulmalıdır. Bu değişkenlerden bazıları bölgesel, kültürel ve dilsel farklılıklar olabilir.

PISA gibi uluslararası düzeyde karşılaştırma ve sıralama yapan sınavlar farklı kültürlerde kullanılırken, uyarlanırken ya da tercüme edilirken çok hassas olunmalıdır. Uluslararası karşılaştırmanın anlamlı olabilmesi için kullanılan testlerin ve anketlerin ölçtüğü yapıların eşdeğer olması gerekmektedir (Asil ve Gelbal, 2012). Diğer bir ifadeyle ölçme değişmezliği sağlanmalıdır. Eğer sınavların ölçme değişmezliği sağlanıyorsa dünya ölçeğinde yapılan sıralamalara ve varyansın kaynaklandığı nedenlere (okullar arasındaki farklılık, bölgeler arasındaki farklılık vb.) güvenilebilir. Ancak faktör yapıları eşit değil ve ölçme değişmezliği sağlanamıyorsa farklılığın kaynaklandığı başka nedenler araştırılmalıdır.

2.2 Yapı Geçerliliği ve Ölçme Değişmezliği

Yapı birbiriyle alakalı öğelerden oluşan bir bütündür (Örs, 2010). Yapı geçerliliği kavramı ilk olarak Cronbach ve Meehl (1955) tarafından ortaya atılmıştır (akt: Kline, 1986). Yapı geçerliliği ölçme aracında yer alan maddelerin ölçülmek istenen özellikle yüksek düzeyde ilişki vermesi ve faktörler arasındaki ilişkilerin de kurama uygun olması durumunu ifade etmesi olarak kabul edilebilir (Yaşar, 2014). Ölçme aracının ölçmek istediği kuramsal temeli yansıtabilmesi olarak da tanımlanabilir. Bu yüzden yapı geçerliliği testin performansıyla testin altında yatan kuramın birbiriyle örtüşmesidir (Bachman, 1990).

Bir bireyin sahip olduđu psikolojik özellikleri, sahip olduđu fiziksel özellikleri gibi doğrudan ölçülemez. Doğrudan ölçülemeyen bu özellikler ‘yapı’ olarak tanımlanır (Şekerciođlu, 2009). Zekâ, yaratıcılık, dışadönüklük-içedönüklük gibi psikolojik özellikler yapıya örnek olarak verilebilir (Crocker ve Algina, 1986’dan akt: Şekerciođlu, 2009). Zekâ ölçülmek isteniyorsa zekânın göstergeleri olan ve birbiriyle ilişkili olan farklı davranışlar vardır. Ölçek geliştirilirken, birbiriyle ilişkili olan bu alt öğeleri test etmeye yönelik test maddeleri hazırlanır. Testteki maddelerin ilgili alanla ne kadar ilişkili olduğuna ve sözü geçen maddeler arasında ne derece ilişki olduğuna yönelik bir analizle yapı geçerliliđi test edilebilir. Yapı geçerliliđinde, ölçülmesi istenen yapıya iyi derecede sahip bir grupta düşük derecede sahip bir grubun ne derece doğru ve hatasız bir biçimde birbirlerinden farklı oldukları incelenir. Ölçme aracının birden çok grupta aynı yapıya sahip olması, o ölçeğin maddelerinin faktör yük değerlerinin ve hata varyanslarının birbirinden farklılaşmadığını anlamına gelir (Jöreskog ve Sörbom, 1993). Gruplarda maddelerin faktör yükleri birbirine eşit çıkarsa maddeler ile o maddelerin ölçmeye çalışıldığı gizil değişken arasındaki ilişki tüm gruplarda benzerdir. Böyle bir durumda, aynı psikolojik özellik, farklı grup da olsa aynı güçlükte ölçülmekte denilebilir (Akyıldız, 2009).

Aynı ölçme aracı kullanılarak ölçülen iki grubun faktör yapılarının birbirinden farklı olması, o ölçme aracının her grupta farklı bir psikolojik özelliđi ölçtüđü anlamına gelir. Eğer böyle bir durum söz konusu olursa söz konusu farklı gruplardan elde edilen sonuçları göz önünde bulundurarak, gruplar arasında bir karşılaştırma ya da sıralama yapmak doğru olmayacaktır. Yapılan bir ölçme uygulamasından elde edilen sonuçlar bireysel farklılık gösterir, fakat bu farklılığın her zaman bireyin performansından ya da becerilerinden kaynaklandığı söylenilemez. Gerçekte bu farklılığın kaynađı ölçme aracı olabilir (Başusta, 2010). Eğer ölçme aracı farklı gruplarda aynı özelliđi ölçemiyorsa gerçekleştirilen uygulama ve yapılan karşılaştırmalar anlamlı olmayacaktır (Başusta, 2010). Yapı eşdeğerliliđi sağlanmadan yapılan karşılaştırmalar elma ve portakalın kıyaslanması gibidir (He ve Van de Vijver, 2012). Yapı eşdeğerliliđinin kültürler arası karşılaştırmalarda ön koşul olduğu belirtilmektedir (Berry, 1969; Rensvold ve Cheung, 2001). Benzer olarak, Horn ve Ardle (1992)’e göre gruplar arası karşılaştırma yapabilmek için, öncelikle ölçme değişmezliđi incelenmelidir. Böylelikle ölçme aracında yer alan maddelerin farklı cinsiyete, kültüre vb. sahip katılımcılar tarafından aynı biçimde

algılanıp algılanmadığı gözlenebilir. Ölçme değişmezliği gözlenmediği takdirde, araştırmanın niceliksel olarak yorumlanmasından çok niteliksel açıdan yorumlanması daha uygun olacaktır (Milfont ve Fischer, 2010).

Byrne (2008)'e göre farklı kültüre uyarlanarak uygulanan ölçme araçlarında, öncelikli olan uygulanacak olan aracın hedef dile çevrilmesidir, bunun ardından orijinal ölçeğin faktör yapısı dikkate alınarak, yeni oluşturulan ölçeğin geçerliliğinin test edilmesi gerekir. Son olarak ise yeni oluşturulan ölçeğin orijinal formuyla ölçme değişmezliği ve yapı geçerliliği sağlayıp sağlamadığı test edilmelidir; fakat gerçekte bu üç süreç daha karmaşık olabilir ve daha çok sayıda işlem ve test gerektirebilir (Byrne, 2008). Örneğin ölçme aracı ilk adım olan hedef dile çevrilirken (uyarlanırken) bile izlenmesi gereken birçok adım ve yöntem bulunmaktadır.

Ölçme değişmezliği, farklı koşullarda gözlenen ya da üzerinde çalışılan durumun, ölçme işlemleri sonucunda aynı nitelikte değerlere sahip olabilmesi olarak tanımlanmaktadır (Horn ve Arde, 1992). Diğer bir ifadeyle ölçme değişmezliği gruplar arası gözlenen farklılığın yanıltıcı olmadığını kanıtlama işlemidir.

Byrne ve diğerleri (2002)'e göre ölçme değişmezliği incelenmeden, gözlenen puanlardaki farklılıkları ve bu farklılıkların kaynağını yorumlamak oldukça zordur. Diğer bir ifadeyle, gözlenen puanlardaki farklılık gruplar arasındaki performans farklılığından kaynaklanacağı gibi, gruplar arasında eşit olmayan gözlenen puanla örtük değişken arasındaki ilişkiden de kaynaklanabilir. Ölçme değişmezliğinin sağlandığı kanıtlandığında, gruplar arasında, örtük değişkenle gözlenen değişken arasındaki ilişki değişmez olarak tanımlanır ve bu durumda gözlenen puanlardaki farklılığın gruplar arasında var olan gerçek performans farklılığından kaynaklandığı söylenebilir.

Öncelikle bir gruba özgü olarak geliştirilen ölçme araçları, daha sonra farklı gruplar için de kullanılmakta ve işte bu noktada ölçme değişmezliği problemi ortaya çıkmaktadır (Borsboom, 2006; Byrne ve Campbell, 1999). Böyle bir durumda, eğer ölçek farklı gruplarda ölçme yaparken benzer nitelikte çalışıyorsa o zaman ölçme eşdeğerliliğinden bahsedilebilir (Gere ve Macdonald, 2012).

Wicherts ve Dolan (2010)'e göre psikolojik ölçme araçlarının kamu, kurum ve işletmelerde ve eğitim gibi alanlarda yansız ve geçerli bir biçimde kullanılabilmesi için, sözü geçen ölçme aracı amacına hizmet edebilmelidir, bu nedenle test puanlarındaki bireysel farklılıklar katılımcıların sahip olduğu farklı demografik özelliklerden (etnik köken, cinsiyet vb.) kaynaklanmıyor olmalıdır.

Ölçme değişmezliğinin var olabilmesi için, gizil değişkenlerin göstergeleri olan gözlenen cevaplar, katılımcıların farklı grubun bir parçası olmasından etkilenmemelidir (Meredith ve Millsap, 1992). Farklı cinsiyet, etnik grup ya da kültür gibi değişkenlerden etkilenen ölçümlerde ölçmenin değişmezliğinden değil, ancak yanlılığından söz edilebilir.

Vandenberg ve Lance (2000)'e göre ölçme değişmezliği test edilirken, gözlenen değişkenlere ait olan ölçüm özelliklerinin ne ölçüde bütün gruplara genellenebileceği sorusuna yanıt aranır. Karşılaştırmalı grup çalışmalarında, geçerli ve uygun bir karşılaştırma yapabilmek için en önemli nokta, karşılaştırılan gruplar arasında ölçme değişmezliği varsayımlarının sağlanmasıdır. Eğer araştırmacılar bu varsayımı göz ardı ederlerse o zaman puanlar arasındaki farklılığın gerçekte ölçülen yapıdan mı kaynaklandığını ya da testin gruplar için farklı biçimde mi çalıştığı bilinemez (Millsap ve Kwok, 2004). Böyle bir durumda yapılan karşılaştırmalar, uygunluk ve geçerlilik anlamında şüphe taşımaktadır. Tyson (2004)'a göre de ölçme değişmezliği, gruplar arası ölçülen yapıya ait değerlerin geçerlik ve güvenilirlik düzeyleri açısından birbirine yakın/benzer olma derecesidir ve ölçme değişmezliğini test etmenin asıl amacının, elde edilen değerlerin karşılaştırılan gruplar için geçerliliğinin sağlanması olduğunu ifade etmektedir. Horn (1991) "Ölçme değişmezliğine kanıt sağlanmadan, araştırma sonuçlarının bir dayanağı olmayacaktır" biçiminde ifade ederek ölçme değişmezliğinin test edilmesinin araştırmanın geçerliliği açısından gerekliliğini vurgulamıştır.

Ölçme değişmezliğinin test edilmesi sadece tek bir testle mümkün değildir, ölçme değişmezliği konusunda yorum yapabilmek için bir biri ardına, hiyerarşik olarak birkaç test uygulanmaktadır (Byrne, 2008). Ölçme değişmezliğinin test edilmesi için araştırmacılar tarafından çeşitli yöntemler önerilmektedir. Bu yöntemlerin sayısı ve isimleri araştırmacılara göre farklılık göstermektedir. Meredith (1993), çalışmasında ölçme değişmezliğinin 4 aşamasından bahsetmektedir. Bunlar, yapısal değişmezlik modeli (configural), zayıf değişmezlik modeli (weak factorial invariance), güçlü değişmezlik modeli (strong factorial invariance), katı değişmezlik modeli (strict factorial invariance) dir. Vandenberg ve Lance (2000) ise ölçme modeller test edilmeye başlanmadan önce, değişmezliği analizinin sayıltılarından birisi olan kovaryans matrisleri testinin yapılması gerektiğini belirtmişlerdir. Ölçme değişmezliği aşamalarını da şu biçimde isimlendirmişlerdir; yapısal değişmezlik (configural invariance), metrik değişmezlik (metric invariance), ölçek değişmezliği

(scalar invariance), özgün faktör değişmezliği (invariant uniqueness). İlgili alanyazında zayıf değişmezlik için metrik değişmezliği, güçlü değişmezlik için ölçek değişmezliği, katı değişmezlik için de özgün faktör değişmezliği isimlerinin kullanıldığı görülmektedir. Vandenberg ve Lance (2000) ölçme değişmezliğinin incelenmesi aşamalarında fikir birliğinin olmadığını, daha farklı ve daha fazla aşamayla inceleme ya da test yapılabileceği gerçeğiyle birlikte, modelleri test etmeden yapılacak olan kovaryans matrisleri eşitliğinin test edilmesi ve temel model olarak nitelendirilen yapısal değişmezlik (configural invariance) modelinin test edilmesinde genel bir görüş birliği olduğunu belirtmişlerdir. Meredith (1993) de araştırmacılar için katı değişmezlik modelinin gerekli ve önemli olduğunu belirtmiştir. Bu bilgiler doğrultusunda Meredith (1993)'ün önerdiği ölçme değişmezliği test aşamaları dikkate alınarak, çalışma kapsamında test edilecek olan ölçme modelleriyle ilgili açıklamalar aşağıda sunulmuştur.

2.2.1 Yapısal Değişmezlik (Configural Invariance)

Alanyazında 'temel model' olarak ifade edilmektedir (Marsh, 1994; Reise, Widaman ve Pugh, 1993). Bu aşamanın amacı, ölçme aracında bulunan maddelerin gizil değişkenlerle ilişkileri bakımından aynı örüntüye sahip olup olmadığının incelenmesidir. Bu aşamada, ölçme aracına ilişkin faktör yapısının gruplar arası benzer olması gerekmektedir (Önen, 2009).

Yapısal değişmezlik test edilirken faktör yükleri, faktör korelasyonları ve hata varyansları sabit tutulmakta ve bu modelde bir sınırlandırma yapılmamaktadır. Bu modelde hiçbir parametre üzerinde eşitlik sınırlaması yapılmadığı için, grup farklılıklarına yönelik bir belirleme yapılamaz (Byrne, 2012).

Yapısal değişmezliğin sağlandığına yönelik bir kanıt elde edilmiş ise diğer bir ölçme değişmezliği test aşaması olan zayıf faktöriyel değişmezlik test edilmelidir.

2.2.2 Zayıf Faktöriyel Değişmezlik (Metric Invariance)

Zayıf değişmezlik ile grupların gizil değişkene ilişkin ölçme biriminin benzer olup olmadığı test edilir. Bu aşamada faktör yüklerinin eşitliği incelenerek, ölçme aracını oluşturan maddelere ilişkin faktör yüklerinin gruplar arası eşit olduğu biçimindeki bir hipotez test edilmektedir (Başusta ve Gelbal, 2015). Faktör yükleri, gizil değişkendeki bir birimlik değişim için gözlenen değişkendeki beklenen değişim

miktarını temsil etmektedir (Önen, 2009). Bu test yapısal değişmezlik testinden daha güçlüdür, çünkü değişmez faktör örüntüsünün tanımlanması haricinde bu aşamada maddelerin faktör yüklerinin gruplar arası eşit olduğu sınırlaması vardır (Bollen, 1989'dan akt: Önen, 2009).

Eğer zayıf değişmezlik gözlenmez ise bu durum grupların farklı ölçme birimlerine sahip olduğunu gösterir (Sommer ve diğerleri, 2009). Zayıf değişmezlik ile aynı gruptaki kişilerin puanları (aynı cinsiyete sahip olanların puanları vb.) karşılaştırılabilir. Farklı grupların arasındaki korelasyon incelenebilir, fakat farklı grupların puanları direkt olarak karşılaştırılmaz. Bu durum şöyle bir örnekle açıklanabilir: Uzunluk birimi kilometre cinsinden olan farklı mesafeler doğrudan karşılaştırılabilirler, fakat farklı uzunluk birimlerini aynı birim cinsine dönüştürmeden geçerli ve uygun bir kıyaslama yapmak mümkün değildir (He ve Van de Vijver, 2012). Bu nedenle bu tür çalışmalar gerçekleştirilirken iki kategori arasındaki uzaklıkların gruplar arasında eşit olup olmadığına bakmak gerekir. Bu durumda zayıf değişmezliğin sağlandığına dair bir kanıt ulaşıldığında, karşılaştırılan gruplar arasında birbirine denk ölçme birimleri var demektir (Salzberger ve diğerleri 1999'dan akt: Önen, 2009).

2.2.3 Güçlü Faktöriyel Değişmezlik (Scalar Invariance)

Zayıf değişmezliğin sağlanması, karşılaştırılan gruplar arasındaki gözlenen puanların ilgili ölçme aracının ölçmeyi hedeflediği psikolojik yapıdan kaynaklandığını ifade etmek için yeterli değildir (Little ve Slegers, 2005). Skalar değişmezlik olarak da adlandırılan güçlü değişmezlik de sağlanmalıdır. Bu aşamada, önceki modelde yapılan sınırlamaya (faktör yükleri) ek olarak, gözlenen değişkenlerin regresyon sabit değerleri de sınırlandırılır ve bu model ile regresyon sabitinin gruplar arasında eşdeğerliliği incelenir (Wu ve diğerleri, 2007). Madde sabiti, gizil değişkenin göstergesi olan ve gözlenen değişken olarak da ele alınan madde puanını yordamak amacıyla oluşturulan regresyon eşitliğinde bulunan sabit değerdir (Önen, 2009).

Güçlü değişmezlik, farklı gruplarda aynı yetenek düzeyine sahip bireylerin, ölçme aracından da aynı puana sahip olup olmadıklarını incelemek açısından gereklidir, bu nedenle bu modelin avantajı ölçme aracının yapı eşdeğerliliğini büyük ölçüde ortaya koymasıdır (Little ve Slegers, 2005).

2.2.4 Katı Faktöriyel Değişmezlik (Strict Factorial Invariance)

Güçlü değişmezlikte sınırlananlara (faktör yükleri, regresyon sabit değerleri) ek olarak, bu aşamada hata varyansları da sınırlandırılır (Little ve Slegers, 2005). Bu nedenle bu modele hata varyansları değişmezliği adı da verilir. Gruplar arası gözlenen değişkenlerin hata varyanslarının farklılaşıp farklılaşmadığı incelenerek gruplar arası eşdeğerliliği test edilir (Vandenberg ve Lance, 2000). Bu modelin çok sınırlandırılmış bir model olduğu ve pratikte karşılanmasının zor olduğu düşünülmektedir (Uyar ve Doğan, 2014).

Bu değişmezlik modelleri, hiyerarşik bir sıra izlenilerek ve birçok model uyum katsayıları kullanılarak test edilir. Farklı gruplarda yapı geçerliliğini test etmek için önerilen yöntem ise yapısal eşitlik modellemesi çatısı altında bulunan çok gruplu doğrulayıcı faktör analizidir.

2.3 Yapısal Eşitlik Modellemesi

Yapısal Eşitlik Modeli (YEM), faktör analizi ve regresyon analizi olmak üzere birden fazla tekniğin bir araya gelmesiyle oluşmuş olan istatistiksel bir modeldir (Hox ve Bechger, 2007; Akyıldız, 2011). YEM, gözlenen ve gizil değişkenler arasındaki ilişkiye yönelik hipotezlerin test edilmesinde kullanılan kapsamlı bir yaklaşımdır (Hoyle, 1995). Jöreskog ve Sörbom (1993) ise bu modeli, gözlenen değişkenler aracılığıyla gizil değişkenlerin incelenmesini sağlayan tekniklere verilen genel isim olarak nitelendirmektedir. Diğer bir ifadeyle, YEM bir analiz değil, birden fazla analiz tekniğini altında toplayan bir kavramdır. Byrne (1998)'e göre, YEM; bazı olgulara dayalı yapısal kuramların çok değişkenli çözümlenmesine yönelik doğrulayıcı bir yaklaşım sunan istatistiksel bir yöntemdir. YEM ile bir grup değişken arasındaki korelasyon /kovaryans örüntülerini anlamak ve belirlenen model ile bu değişkenlerin uyumunu açıklamak amaçlanmaktadır (Kline, 2011).

YEM, bir çok yönden geleneksel olarak kullanılan korelasyon, regresyon ve varyans analizine benzemektedir. Fakat diğer analizlerde sadece gözlenen değişken incelenebiliyorken, onlardan farklı olarak çok değişkenli bir teknik olan YEM ile hem gözlenen hem de gizil değişkenler analiz edilebilmektedir. Byrne (1998), gizil ve gözlenen değişkenler arasındaki ilişkiyi gösteren en eski ve en bilinen yöntemin

faktör analizi olduğunu belirtmektedir. YEM ise faktör analizinden yararlanarak, değişkenler arasındaki regresyon yapısının belirlenmesine olanak sağlamaktadır.

YEM, deneysel yaklaşımlar kullanılarak yapılan araştırmaların sınırlılıklarını gidererek daha kolay bir araştırma sağlamaktadır. YEM’de, diğer istatistiksel yöntemlere kıyasla, açıklayıcı ve keşfedici bir yapı yerine, doğrulayıcı bir yapı söz konusudur (Ayyıldız ve Cengiz, 2006).

Günümüzde birçok bilim dalı tarafından kullanılmakta olan YEM, belli teoriye dayalı olarak model içerisinde değişkenlerin nedensel ilişkisinin incelenmesine olanak sağlar. YEM analizinde, uyum indekslerinden faydalanılarak modelin uyumu test edilir. Modelin test edilmesinde, ulaşılan uyum indeksleri model ile veri arasında uyum olduğunu gösteriyorsa oluşturulan hipotezler kabul edilir ve model doğrulanır, fakat uyum indeksleri gerekli düzeylerde değilse ilk başta düzenlenen modelin yeniden belirlenmesi gerekmektedir. YEM’de ölçme modeli doğrulayıcı faktör analizi (DFA) ile değerlendirilir.

2.3.1 Doğrulayıcı Faktör Analizi

Doğrulayıcı faktör analizi gözlenen değişkenler ya da göstergeler (test maddeleri, test puanları vb.) ile gizil değişkenler arasındaki ilişkiyi/uyumu inceleyen yapısal eşitlik modeli altında ele alınan bir analizdir (Brown, 2006). YEM’in en önemli kavramlarından biri olan gizil değişkenler, doğrudan gözlenemeyen ve araştırmaların asıl amaçları doğrultusunda açığa çıkarılmaya çalışılan zekâ, duygu, başarı, tutum gibi soyut kavramları ya da psikolojik yapıları ifade eder. DFA’da araştırmacı, faktör sayısı, faktör yük değerleri gibi parametreleri önceden belirler ve böylece DFA kullanılarak önceden belirlenen ölçme modeline göre ölçme yapıp yapılamadığı test edilir. AFA (Açımlayıcı Faktör Analizi), değişkenler arasındaki ilişkilerden yola çıkarak faktör bulmak, teori üretmek için uygulanır; DFA ise değişkenler arasındaki ilişkiye yönelik daha önce saptanan bir hipotezin test edilmesi için uygulanır (Tabachnick ve Fidell, 2013). DFA’yı AFA’dan ayıran en belirgin özelliklerden bir tanesi de budur; diğer bir ifadeyle AFA’da bir model tanımlamadan, değişkenler arasındaki ilişkiler incelenerek değişkenlerin gruplanmasına ilişkin bir araştırma yapılır.

Doğrulayıcı faktör analizi modelleri şematik olarak şekil ve semboller ile gösterilebilir; oval ya da daireler gizil değişkeni, kareler ise göstergeleri ya da gözlenen değişkenleri simgeler, değişkenler arasındaki ilişkiyi göstermek için

çizgiler kullanılır. Tek yönlü oklar, doğrudan ilişkileri temsil etmek için kullanılırlar ve ok yönündeki değişken bağımlı değişkendir, iki yönlü oklar ise gizil değişkenler arasında nedensel olmayan (korelasyonel) bir ilişkiyi ifade eder (Kline, 2011).

2.3.2 Uyum İndeksleri

DFA sonuçları değerlendirilirken, bir takım uyum indeksleri kullanılmakta ve uyumun iyi olup olmadığına yönelik bazı kesme noktaları dikkate alınmaktadır. Model uyumuna yönelik indekslerden belli başlı olanları aşağıda açıklanmıştır.

χ^2 (Chi Square Goodness of Fit) : Bu değer yapısal eşitlik modeli çalışmalarında öncelikli olarak incelenmesi gerektiği düşünülen uyum değeri olmakla birlikte, gözlenen ve beklenen veri matrisi arasındaki farkın manidar olup olmadığı konusunda bilgi verir (Brown, 2006). İyi bir model uyumu için, χ^2 'nin manidar olmaması beklenmekte ve bu nedenle $p > .05$ olması beklenen bir durumdur. Bu durumda, gözlenen ve beklenen kovaryans matrisleri arasında fark yoktur hipotezi kabul edilir (Çokluk, Şekercioğlu ve Büyüköztürk, 2010). χ^2 testi sıklıkla kullanılmasına rağmen, χ^2 örneklem büyüklüğüyle bağlantılı ve ona duyarlı olması nedeniyle örneklem büyüklüğü arttıkça χ^2 değerinin de manidar olma olasılığı artar ve bu durum modelin çok da geçerli olmayan sebeplerle reddedilmesine yol açabilir (Brannick, 1995; Bollen, 1989; Tucker ve Lewis, 1973). Benzer bir biçimde örneklem büyüklüğü 200 ya da daha küçük ise χ^2 değerinin de küçülme olasılığı artmakta ve model daha iyi uyum gösterebilmektedir ve bu nedenle yanlış yorumlamalar yapılabilir (Tabachnick ve Fidell, 2013). χ^2 ;

$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

formülü ile hesaplanır. Formülde bulunan o değeri; gözlenen, e değeri de beklenen değerleri ifade etmektedir (Cramèr, 1946'dan akt: Crack, 2018).

χ^2/sd (Chi Square/ Degrees of Freedom) : χ^2 testinin sınırlılıklarından dolayı, örneklem büyüklüğünün etkisini en aza indirmek için, Wheaton ve diğerleri (1977) χ^2/sd oranının kullanılabileceğini önermişlerdir. Bu oranın uygun kesme noktaları konusunda fikir birliği olmamakla birlikte Wheaton ve diğerleri (1977) tarafından $\chi^2/sd \leq 5.0$ olabileceği, Tabachnick ve Fidell (2013) tarafından, bu oranın ≤ 2.0 olması gerektiği belirtilmiştir. Kline (2011) ise bu oranın 3.0 veya altında olmasının iyi bir uyumun göstergesi olduğunu belirtmektedir. Bu doğrultuda, daha

hassas ve kesin sonuç için 2.0'ın altında bir değere sahip olmak istenen bir durumdur ve böylece uyumun mükemmele yakın olduğu konusunda karar alınabilir.

GFI (Goodness of Fit Index) : GFI Jöreskog ve Sörbom tarafından χ^2 'ye alternatif olarak geliştirilmiş ve kestirilen örneklem kovaryansı tarafından açıklanan varyans oranını hesaplar (Tabachnick ve Fidell, 2013). Null modele kıyasla test modelinin ne kadar iyi uyuma sahip olduğunu belirler (Schermelleh-Engel, Moosbrugger, Müller, 2003). 0 ile 1 arasında değere sahip olup, 1'e yakın olması iyi uyum göstergesidir. Bu değer $\geq .90$ olması beklenmektedir (Hu ve Bentler, 1999; Hooper, Coughlan ve Mullen, 2008). Örneklem büyüklüğüne daha az duyarlı olduğundan, büyük örneklerde kullanılması daha uygun görülmektedir. (Tabachnick ve Fidell, 2013, Rhodes-Purdy, 2017). Hu ve Bentler (1999), GFI değerinin diğer uyum indekslerine kıyasla çok iyi performans göstermediğini belirtmiştir. Bazı araştırmacılar da bu değer yorumlanmasını tarihsel önemine bağlamaktadırlar (Hooper, Coughlan ve Mullen, 2008). GFI;

$$GFI = 1 - \frac{X_t^2}{X_b^2}$$

eşitliği ile hesaplanmaktadır (Schermelleh-Engel, Moosbrugger, Müller, 2003). X_t^2 test modelinin ki karesini, X_b^2 oluşturulan null (baseline) modelin ki karesini ifade etmektedir.

CFI (Comparative Fit Index) : Örneklem büyüklüğü de dikkate alınarak hesaplanan bu değer küçük örneklerle de uygun sonuçlar vermektedir (Tabachnick and Fidell, 2013) . Bu indeks bütün gizil değişkenler arasında ilişkinin olmadığını varsayar ve bir null (yokluk) modeli oluşturarak, test edilen modelin uyumunu null modelle karşılaştırır. 0 ile 1 arasında değere sahip olup, 1'e yakın olması iyi uyum göstergesidir (Hooper, Coughlan ve Mullen, 2008). Hu ve Bentler (1999), CFI ve benzer uyum indekslerinin .90'nın altında olmamasını önermekle birlikte, model uyumu için kesme noktasının .95'e yakın olması gerektiğini belirtmişlerdir. Bu doğrultuda, iyi bir model uyumunun kanıtı için CFI $\geq .95$ olması istenen bir durumdur. CFI değeri,

$$CFI = 1 - \frac{\max[(X_t^2 - sd_t), 0]}{\max[(X_t^2 - sd_t), (X_b^2 - sd_b), 0]}$$

eşitliği ile hesaplanır ve model uyumunu, test modelini sınırlandırılmış temel (baseline) modelle karşılaştırarak ortaya koyar (Yu,2002).

NNFI (Non-Normed Fit Index), TLI (Tucker-Lewis Index) : TLI olarak da bilinen NNFI, NFI (Normed Fit Index)'nin örneklem büyüklüğünden etkilenmesi nedeniyle, bu dezavantajı ortadan kaldırmak için geliştirilmiştir (Cangur ve Ercan, 2015). 1980'de Bentler ve Bonett tarafından geliştirilen ve temel model (baseline) olarak bir null model kullanan bu indeks genel olarak .90 ve üzeri olması durumunda kabul görmektedir (Wegener ve Fabrigar, 2000, Little, 2013). Bentler ve Hu (1999) ise NNFI/TLI değerinin .95 ve üzerinde olmasını önermektedirler. Null model ve test modelinin χ^2/sd oranları kullanılarak hesaplanan (Little, 2013) bu indeks,

$$NNFI = \frac{(X_b^2/sd_b) - (X_t^2/sd_t)}{(X_b^2/sd_b) - 1}$$

formülü ile hesaplanır.

RMSEA (Root Mean Square Error of Approximation) : Steiger tarafından 1980 yılında geliştirilen RMSEA değeri,

$$RMSEA = \frac{\sqrt{X_m^2 - sd_m}}{\sqrt{sd_m(N - 1)}}$$

formülü ile hesaplanır (Hooper, Coughlan ve Mullen, 2008; Yu, 2002). N, örneklem büyüklüğü; sd, modelin serbestlik derecesini ifade etmektedir. Eğer χ^2 değeri sd değerinden küçükse RMSEA 0'a eşit olur (Mondal, Das, Musunuru, Dash, 2017). Hu ve Bentler (1999) çalışmalarında, RMSEA değerinin .06'ya yakın da bu değer altında olmasını gerektiğini ifade etmişlerdir. Bir çok araştırmacı genel olarak, RMSEA'nın .08 ve altında bir değer almasını iyi uyumun göstergesi olarak kabul etmektedirler (MacCallum, Browne, Sugawara,1996'dan akt: Hooper, Coughlan ve Mullen, 2008). .05'e yakın veya daha altında bir değer alması ise uyumun çok yüksek olduğunun göstergesi olarak ele alınmıştır (Browne ve Cudeck, 1993).

SRMR (Standardized Root Mean Square Residual) : Bentler tarafından 1995'te tanıtılan bu değer, Jöreskog ve Sörbom tarafından 1981'de geliştirilen RMR'nin standardize edilmiş bir versiyonur (Yu, 2002). Bu değer, örnekleme ait kovaryans matrisiyle kestirilen kovaryans matrisleri arasındaki farktır. SRMR, RMR'ye kıyasla daha çok tercih edilmektedir, çünkü değişken metrikleri bu değeri etkileyebilmekte ve değer yorumlanmasını zorlaştırmaktadır (Brown, 2006). Hu ve Bentler (1999) SRMR değerinin .08 altında bir değer almasının kabul edilebilir

olduğunu belirtmektedir. .05'in altında bir değer alırsa iyi bir uyumun varlığına işaret etmektedir (Hu ve Bentler, 1999; Kline, 2011).

$$SRMR = \frac{\sqrt{2 \sum \sum [(s_{ij} - \sigma_{ij}) / (s_{ii} s_{jj})]^2}}{\sqrt{p(p+1)}}$$

SRMR değeri yukarıda bulunan formülle hesaplanmakta olup, p gözlenen değişken sayısını, s_{ij} gözlenen kovaryansları, σ_{ij} yeniden üretilmiş kovaryansları, s_{ii} ve s_{jj} gözlenen standart sapmaları ifade etmektedir (Hu ve Bentler, 1999; Chen, 2007) .

2.3.3 Çok Gruplu Doğrulayıcı Faktör Analizi

Çoklu grup doğrulayıcı faktör analizi (ÇG-DFA) ise testin faktör yapısı değişmezliğini veya eşitliğini test etmek için kullanılan bir yöntem olup, ölçme aracının kültürlerarası geçerliğini incelemekte son yıllarda yaygın bir biçimde kullanılmaya başlamıştır. Buradaki temel amaç, cinsiyet, yaş, etnik köken vb. özellikler bakımından birbirlerinden farklılaşan gruplar için, uygulanan belli bir ölçme aracının sonucunda ortaya çıkan faktör yapısının gruplar arası farklılaşp farklılaşmadığını tespit etmektir. Başka bir ifadeyle, farklı grupların aynı yapı ile ölçülüp ölçülemediği test edilir. Çoğunlukla, psikolojik testler geliştirilirken, testin sahip olması beklenen bazı psikometrik özelliklerin, bütün hedef gruplarda eşit olup olmadığının belirlenmesi için ÇG-DFA kullanılır (Thompson, 2004'ten akt. Akyıldız, 2011).

ÇG-DFA ile teorik yapı iki veya daha fazla grupta gözlenen yapıyla karşılaştırılır. Faktör yapılarını test etmeye yönelik analizlerde, üç tür hipotez test edilir. Eğer bir ölçme aracı birden fazla grupta aynı yapıya sahipse o ölçme aracının maddeleri faktör yükleri, faktörler arası korelasyon katsayıları ve hata varyansları bakımından farklılaşmadığını anlamına gelmektedir (Byrne, 1998; Jöreskog ve Sörbom, 1993).

2.3.3.1 Çok Gruplu Doğrulayıcı Faktör Analizi Varsayımları

Çok gruplu doğrulayıcı faktör analizini gerçekleştirmeden önce test edilmesi gereken varsayımlar bulunmaktadır. Bunlar, kayıp değer ve uç değerlerin bulunup bulunmadığı, çoklu bağlantı (multicollinearity) probleminin var olup olmadığı,

grupların normal dağılıma sahip olup olmadığı ve veri setinin güvenilir olup olmadığı ile ilgili analizleri kapsar. Çokluk, Şekercioğlu ve Büyüköztürk (2010) tarafından bu gerekliliklerin test edilmesi önerilmektedir. Aynı zamanda değişmezlik modelleri test edilmeden önce, DFA ile her bir gruba ilişkin verilerin test edilmesi ve kovaryans matrislerinin eşitlenip eşitlenmediğinin belirlenmesi gerekmektedir.

Alanyazında, örneklem büyüklüğüne göre ne kadar kayıp değer toleran edilebileceğine ilişkin bir ölçüt olmamasıyla birlikte uygulanabilecek bazı alternatiflerin olduğu belirtilmektedir (Tabachnick ve Fidell, 2013). Bunlar; kayıp verileri silmek ya da kayıp değerler yerine değer atamaktır. Eğer kayıp değerler örneklem büyüklüğünü ciddi boyutta etkileyerek araştırmayı zorlaştırabileceyse kayıp değerlere veri atama yapılabileceği, aksi takdirde az sayıda katılımcı kayıp değere sahipse o katılımcıların veri setinden çıkarılabileceği önerilmektedir.

Uç değerler, Barnet ve Lewis (1994) tarafından şöyle tanımlanmıştır; uç değerler örnekleme bulunan diğer üyelerden önemli derece farklılaşan öğelerdir. Uç değerler tek yönlü ve çok yönlü olabilir, tek yönlü uç değer; deneklerin bir tek değişkene yönelik çok farklı ya da aykırı değerlere sahip olmalarını, çok yönlü uç değer ise birden fazla değişkene yönelik aykırı değerlerin kombinasyonunu ifade eder (Çokluk ve diğerleri, 2010). Tek yönlü uç değerler kutu diyagramı, histogram ya da standart Z puanının hesaplanması gibi yöntemlerle belirlenebilirken, çok yönlü uç değer Mahalanobis uzaklığı olarak adlandırılan istatistiksel bir işlemle belirlenebilir. Uç değerlerin test sonuçlarını etkileyebileyecek olmaları nedeniyle analizlere başlamadan önce tespit edilerek, kaynakları araştırılarak veri setinden çıkarılmaları ya da bu değişkenlerin istatistiksel bir işlemle dönüştürülmeleri gerekmektedir (Mertler ve Vannatta, 2005).

Çoklu bağlantı problemi (multicollinearity) ise değişkenler arasındaki ilişkinin güçlü ve yüksek olması durumudur. Değişkenler arasında böyle bir problem varsa bu duruma neden olan değişken ya da değişkenlerin analizden çıkarılması önerilmektedir (Tabachnick ve Fidell, 2013). Çoklu bağlantı problemini test edilmesi amacıyla kullanılan bazı yöntemler olup (Çokluk ve diğerleri, 2010), en çok tercih edilenler; değişkenler arası korelasyonların incelenmesi ve bağımsız değişkenler için tolerans değerleridir. Korelasyon incelemesinde $r < .90$ olması beklenmektedir. $.10$ 'dan küçük olan tolerans değerleri de çoklu bağlantı problemi olduğunun göstergesidir (Field, 2009).

Normal dağılım birçok istatistiksel analizin dayandığı bir varsayım olarak örneklemin normal olarak dağıldığını varsaymaktadır. Normal ya da normale yakın bir dağılımın olması geçerli ve güvenilir sonuçlar sağlamak için gereklidir (Field, 2009). Normallik varsayımı betimsel ve grafiksel yöntemlerle incelenebilmekte ve verilerin, aritmetik ortalama, mod, medyan, çarpıklık ve basıklık katsayıları gibi hesaplamalar kullanılarak incelenmesi betimsel yöntemler olarak ifade edilmektedir (Abbott, 2011). Ayrıca örneklem büyüklüğüne göre kullanılması tercih edilebilecek olan ve en yaygın kullanılan hipotez testleri şunlardır; Shapiro-Wilks W Testi ve Kolmogorov-Smirnov (Hair, Anderson, Tatham ve Black, 1998).

Güvenilir sonuçlara ulaşmak için, öncelikle incelenen ölçek yapısının güvenilir olup olmadığının incelenmesi gerekmektedir. Güvenilirlik farklı yollar izlenerek belirlenebilir. Bu yöntemler; bir ölçeği bir kez uygulamak, bir ölçeği iki kez uygulamak veya iki eşdeğer ölçeği bir kez uygulamaktır. Bir ölçek bir kez uygulanmışsa iç tutarlılık güvenilirliği incelenir ve iç tutarlılığı test etmek için kullanılan güvenilirlik katsayılarıyla maddelerin birbiri arasında ve ölçeğin tamamıyla ne derece tutarlı olduğu belirlenebilir (Ercan ve Kan, 2004).

ÇGDFA ile ölçme değişmezliği test etmeden önce gruplar için DFA'nın test edilmesi önerilmektedir. Van de Schoot, Lugtig ve Hox (2012) tarafından yapının teorik anlamda işlevselliğini görmek açısından, ölçme değişmezliği çalışmalarına her bir grup için DFA yapılarak başlanması gerektiği belirtilmektedir. DFA modelinin her bir grup için iyi uyum göstermesi gerekmektedir.

Ölçme değişmezliğinin diğer bir varsayımı olan kovaryans matrisleri eşitliği testinin ölçme değişmezliği modellerinden önce test edilmesi gerekliliği ilk olarak 1971'de Jöreskog tarafından önerilmiştir. Kovaryans matrisleri eşitliğinin test edilmesinin nedeni; DFA model parametrelerinde grup farklılıkları varsa kovaryans matrisindeki değerlerin de gruplara göre değişecek olmasıdır. Bu nedenle kovaryans matrisleri eşitliği testi bağımlı değişkenler için gözlenen kovaryans matrislerinin gruplar arasında eşit olduğu yönündeki hipotezi reddetmezse analizlere devam edilmesine gerek kalmayabileceği belirtilmektedir (Steenkamp ve Baumgartner, 1998; Vandenberg ve Lance, 2000; Brown, 2006). Diğer bir ifadeyle, bazı araştırmacılara göre eğer kovaryans matrisleri gruplar arasında farklılık göstermiyorsa ölçme değişmezliğinin sağlandığı kabul edilebilir, fakat kovaryans matrisleri gruplar arasında farklılık gösteriyorsa bu farklılığın kaynağını belirlemek için ölçme modellerinin test edilmesi gerekmektedir (Vandenberg ve Lance 2000).

Fakat kovaryans matrisleri arasında farkın olmadığı kabul edilse bile, sonraki aşamalarda uygulanacak olan değişmezlik modellerinin kabul edilmeyebileceği ya da tam tersi durumun söz konusu olabileceği de belirtilmektedir. Bu uyumsuzluğun kaynağının da kovaryans matrislerinin karşılaştırılabileceği bir temel model olmaması olarak ifade edilmiştir (Byrne, 1998). Kovaryans matrisleri eşitliğinin öncelikli olarak test edilip edilmemesi konusunda araştırmacılar arasında fikir ayrılıkları olsa da, bu test grup farklılıklarına ya da benzerliklerine ilişkin bilgi sunmaktadır. Fakat tartışmalardan edilen bilgiler doğrultusunda bu test tek başına ölçme değişmezliğin sağlanıp sağlanmadığının kanıtı da olamaz, bu nedenle bu test sonucu gruplar arasında uyum olduğu yönünde olsa bile daha güçlü ve daha net bilgiler sağlayacak olan çok gruplu doğrulayıcı faktör analizi ile ölçme değişmezliğine yönelik karar alınmalıdır.

2.3.3.2 Çok Gruplu Doğrulayıcı Faktör Analizi Aşamaları

Ölçme değişmezliğine yönelik gerekli varsayımlar test edilerek verilerin analize uygun olduğunun anlaşılmasıyla analizin ilk adımı olan yapısal değişmezlik modeli test edilir (Model 1). Değerlendirilen uyum indeksleri doğrultusunda, bu modelin iyi uyum göstermesi beklenir ve diğer aşama olan zayıf faktöriyel değişmezlik modeli (Model 2) test edilir. Diğer aşamada, güçlü faktöriyel değişmezlik modeli (model 3) bulunmaktadır. Son aşamada da katı faktöriyel değişmezlik modeli (model 4) test edilir (Jöreskog ve Sörbom, 1993). Alanyazında, en az sınırlandırılmış modelle daha çok sınırlandırılmış model arasında karşılaştırma yapılarak ölçme değişmezliğinin var olup olmadığıyla ilgili bir sonuca ulaşılmaktadır. Modeller arası karşılaştırma yapmak için kullanılan farklı yaklaşımlar vardır. Bunlardan bir tanesi olan ve daha iyi sonuçlar sunduğu düşünülen yaklaşımda, yapı en az sınırlandırılmış modelden daha çok sınırlandırılmış modele doğru oluşturulur (Harrington, 2009). Bu yaklaşımda, daha çok sınırlandırılmış modeller, daha az sınırlandırılmış modellerde yuvalanmış (nested) olduklarından, yapılan ek sınırlandırmaların model uyumluluğunu önemli bir ölçüde etkileyip etkilemediğini test etmek amacıyla χ^2 üzerindeki değişim dikkate alınabilir (Brown, 2006). Bu doğrultuda, yuvalanmış modeller karşılaştırılırken; model 1 ve 2, model 2 ve 3, model 3 ve 4 kendi aralarında karşılaştırılarak değişim incelenir, ancak sadece χ^2 üzerindeki farkı değerlendirmek yeterli olmamaktadır. Ölçme değişmezliğinin sağlanması için, diğer bir ifadeyle grupların birbirinden manidar bir biçimde

farklılaşmadığını söyleyebilmek için, χ^2 'nin manidarlık düzeyinin anlamlı olmaması beklenmektedir. Büyük örneklerde bu durum sorun teşkil edebileceği için, bazı araştırmacılar model uyumlarının değerlendirilebilmesine yönelik alternatif uyum indekslerinin de incelenmesi gerektiği üzerinde durmuşlardır.

2.3.3.3 Model Karşılaştırmalarında Alternatif Uyum İndeksleri

Ölçme değişmezliği modelleri arasındaki uyum incelenirken ilk olarak Δ (χ^2/sd) dikkate alınabilir. Bu oranda bir iyileşme varsa χ^2 farkının manidar olup olmadığına yönelik S-B χ^2 fark derecesi (scaled difference) için \check{T}_s değeri hesaplanmalı ve bu değer χ^2 dağılım tablosunda bulunan $p<.01$ ya da $p<.05$ düzeylerindeki kritik değerle karşılaştırılmalıdır (Satorra ve Bentler, 2001). Eğer \check{T}_s değeri kritik tablodaki değerden küçükse modeller arasında fark olmadığına yönelik olarak kurulan hipotez kabul edilir. S-B χ^2 değeri kullanarak \check{T}_s değeri hesaplanmalı ve ölçme değişmezliğinin sağlanıp sağlanmadığı konusunda alternatif bir karar sunulmalıdır. \check{T}_s değeri Satorra ve Bentler (2001) tarafından sunulan formülle hesaplanabilir:

$$\check{T}_s = T_d / c_d$$

T_d , T_0 (yuvalanmış modelin normal maksimum olasılık χ^2 değeri) ile T_1 (daha az sınırlandırılmış modelin normal maksimum olasılık χ^2 değeri) arasındaki farktır.

$$T_d = T_0 - T_1$$

c_d ise düzeltme derecesidir ve aşağıda bulunan eşitlikle hesaplanmaktadır.

$$c_d = (r_0 c_0 - r_1 c_1) / (r_0 - r_1)$$

r_0 yuvalanmış modelin serbestlik derecesini, r_1 ise daha az sınırlandırılmış modelin serbestlik derecesini ifade etmektedir. c_0 yuvalanmış modelin düzeltme derecesi, c_1 ise daha az sınırlandırılmış olan modelin düzeltme derecesi olup, c_0 ve c_1 ;

$$c_0 = T_0 / \check{T}_0$$

$$c_1 = T_1 / \check{T}_1$$

eşitlikleri ile hesaplanmaktadır. \check{T}_0 yuvalanmış modelin S-B χ^2 değeri, \check{T}_1 daha az sınırlandırılmış olan modelin S-B χ^2 değeridir.

Cheung ve Rensvold (2002) tarafından yapılan bir simülasyon çalışmasına göre, gruplar arası değişmezliği test etmek için kullanılabilir istatistiklerden en iyi ve en güçlü olanları ΔCFI , $\Delta \text{Gamma hat}$ ve $\Delta \text{McDonald's NCI}$ olarak belirtilmiştir. Aynı çalışmada kesme noktalarının da üzerinde durulmuştur; eğer CFI değerindeki fark (Δ) $-.01$ değerinden daha küçük ya da $-.01$ ise modeller arasında farkın olmadığına yönelik kurulan H_0 hipotezi kabul edilmeli, ΔCFI $-.01$ 'in üzerindeyse bu hipotez reddedilmelidir. $\Delta \text{Gamma hat}$ ve $\Delta \text{McDonald's NCI}$ 'ya yönelik olarak sunulan kesme noktaları ise sırasıyla $-.001$ ve $-.02$ olarak belirtilmiştir.

Çalışmasında CFI, RMSEA, Gamma hat, Mc ve SRMR değerleri üzerinde durmuş olan Chen (2007) CFI, Gamma hat, Mc ve SRMR değerlerinin nispeten örneklem büyüklüğünden bağımsız olduklarını, fakat RMSEA değerinin diğerlerine kıyasla örneklem büyüklüğünden daha fazla etkilendiğini belirtmiştir. Aynı zamanda Gamma hat değeri CFI değeriyle hemen hemen aynı sonuçları vermiş olup, Mc değeri SRMR ve RMSEA'nın gerisinde kalmıştır. Bu bilgiler doğrultusunda en sık kullanılan ve en çok önerilenler CFI ve SRMR'dir. Aynı çalışmada üzerinde durulan kesme noktaları şu biçimdedir; grupların örneklem sayılarının birbirinden farklı olduğu ve 300 ve altında bir örneklem sayısına sahip olan gruplar için, zayıf faktöriyel değişmezlik aşamasında ΔCFI için $-.005$ ve üzerinde bir değişim, ΔSRMR için $.025$ ve üzerinde bir değişim olması değişmezliğin sağlanamadığı yönünde yorumlanabilir. Güçlü ve katı değişmezlik aşamalarında ise ΔCFI için $-.005$ ve üzeri, ΔSRMR için $.005$ ve üzeri bir değişim değişmezliğin sağlanamadığı belirtilebilir. Ayrıca, grupların örneklem sayılarının eşit olduğu 300 ve daha fazla örneklemin bulunduğu durumlarda, sırasıyla zayıf faktöriyel değişmezlik aşamasında ΔCFI için $-.010$ ve üzeri bir değişim, ΔSRMR için $.030$ ve üzeri bir değişim, güçlü ve katı faktöriyel değişmezlik aşamasında ΔCFI için $-.010$ ve üzeri değişim, ΔSRMR için $.010$ ve üzeri değişim gözlenmesi değişmezliğin sağlanamadığına yönelik şüpheyi artırmaktadır.

Model karşılaştırmalarında aynı zamanda sıklıkla $\Delta \text{NNFI/TLI}$ raporlaştırılmaktadır. Schnabel, Kelava, Van de Vijver, Seifert (2015) tarafından yapılan çalışmada $\Delta \text{TLI} \leq -.010$ olarak kabul edilmiştir. Benzer biçimde, Sözer, Saritaş, Çavdar, Çorbacı ve Bağcı (2017) tarafından, ölçme değişmezliği incelenirken ΔCFI ile birlikte $\Delta \text{NNFI/TLI}$ değeri de yorumlanarak, NNFI/TLI fark derecesi için kesme noktası, ΔCFI 'ya benzer biçimde $.010$ olarak dikkate alınmıştır.

Modeller karşılaştırılırken, model uyumunun karşılanması için sınırlandırılmış modelin sınırlandırılmamış modelden daha kötü sonuç vermesi beklenmektedir (Şekercioğlu, 2018). Bu doğrultuda, küçüldükçe daha iyi uyum gösteren χ^2/sd oranının artması, artmalı uyum indekslerinden olan CFI ve NNFI'nın aynı kalması ya da düşmesi beklenirken, azaltmalı uyum indekslerinden olan SRMR'nin ise düşmemesi beklenmektedir. Eğer sınırlandırılmış modelde, aksi durum söz konusuysa iyileşme olduğu belirtilir. O durumda Δ 'ya yönelik kesme noktalarından yararlanılarak, farklılığın manidar olup olmadığı değerlendirilir.

2.4 Ölçme Değişmezliği ve Kültür İlişkisi

Gelişen teknoloji ve bilimle birlikte, dünyada kültürlerarası iletişim artmaktadır. Kültürlerarası iletişim, farklı kültüre sahip insanların günlük ve toplumsal hayatın çeşitli alanlarında birbirleriyle kurdukları iletişimi konu alır (Sarı, 2004). Çağa ayak uydurabilmek için farklı kültürlerin yaptığı çalışmaları da takip etmek ve onlardan yararlanmak gerekir. Bunu yaparken kültürel farklılıkların farkında olunmalıdır. O nedenle öncelikle o kültürün özellikleri iyi tanımak ve anlamak önemlidir. Kültür, edinilen bir davranış kabıdır ve insanın yaptığı her şeyi kapsayan bir yaşam biçimidir. Kültür, bu özelliğiyle gelenekleri, töreleri, konuşma tarzlarını, müzik anlayışlarını, yiyecek ve içecek alışkanlıklarını, giyim tarzları ve yaşam alanları gibi birçok geniş kapsamlı davranış kalıplarını içeren bir bütündür (Engin, 1990). Bütün bunlar dikkate alındığında kültürün bireyin davranışında, tepkilerinde, kişilik ve zekâ gelişiminde etkisi ve rolü yadsınamaz. Bu nedenle uluslararası yapılan ölçme işlemleri de ölçme değişmezliğinin sağlanması açısından, katılımcıların mensup olduğu kültürler göz önünde bulundurularak uygulanmalı ve değerlendirilmelidir. Cronbach (1990) farklı kültürlerin benzer uyarıcılara farklı biçimde tepkiler verdiğini belirtmekte olup, yetenek testi uygulanan bazı kültürlerin, gerçekte sahip oldukları yetenekleri o testle gösteremeyeceklerini belirtmektedir.

Serpell (1979)'in bulgusuna göre, farklı kültürlerdeki çocuklardan bazı şekilleri tel, kâğıt ve kalem, çamur kullanarak yeniden oluşturmaları istendiğinde, Zambialı çocuklar tel konusunda İngiliz çocuklar ise kâğıt-kalem konusunda daha iyi performans sergilemişlerdir. Bu bulgudan yola çıkarak kâğıt-kalem kullanılarak yapılan zekâ testlerinin batı kültürü lehine çalıştığı söylenebilir.

Demetriou ve diğeri (2005) tarafından yapılan bir çalışmada çoğunlukla görsel ve mekânsal öğelerden oluşan bir testte, Çinli çocukların Yunan çocuklardan daha üstün performans gösterdikleri görülmüştür (akt: He ve Van de Vijver, 2012). Buradan da anlaşılacağı üzere, kültürel açıdan dikkat edilmesi gereken sadece ölçme aracını oluşturan soruların içeriği değil aynı zamanda soru tipleridir. Bu durum öğrencinin geçmişten o güne kadar olan alışkanlığıyla da ilgilidir. Bulunduğu bölgede ya da okulunda uygulanan tekniklere daha çok aşinalık kazanılmış ve farklı bir soru tarzıyla karşılaşıldığında gerçek performans sergilenemeyebilir. Malda, Van de Vijver ve Temane (2011) tarafından Güney Afrika'da yaşayan beyaz ve siyah ırka yönelik ayrı ayrı ölçme araçları geliştirilerek uygulanan çalışmada, çocukların kendi toplumları için geliştirilen formlarda daha başarılı olduğu görülmüştür.

Test ve envanterlerin başka kültürler çevrilmesi, farklı kültürlerdeki insanların başarı, tutum, kişilik ve diğer psikolojik özelliklerinin karşılaştırılması için yaygın bir yöntemdir. Kültürlerarası testler genellikle farklı ana dil konuşan gruplara uygulanmaktadır ve test sadece İngilizce olarak geliştirilebilmektedir. Bu durumda ana dili İngilizce olmayanlar birçok yapıya aşina olmayabilirler ve bu durum onlar için dezavantaj olabilir. O testi hedef dile çevirmek oldukça mantıklı olabilir, ancak bu çeviri işi diğer dilde kelimeleri yeniden yazmak olarak anlaşılmamalıdır. Kültüre bağlı olarak değişen anlamsal farklılıklar ya da karşılığı olmayan ifadeler varsa bunlar dikkatli bir biçimde çevrilmeli ve uyarlanmalıdır. Örneğin bazı kültürler için elektrik süpürgesinin anlamsal bir karşılığı olmayabilir (Malda ve diğeri, 2008).

Araştırma konusunun kültürel ya da evrensel olduğuna yönelik iki yaklaşım bulunmaktadır. Araştırma konusunu, kendi bağlamında değerlendirilmesi gerektiğini ifade eden emik yaklaşım, araştırma konusunun yerel-kültürel olduğunu varsaymaktadır (Sargut, 2001). Emik yaklaşım belirli bir kültürde var olan davranış ve tutumların, yalnızca o kültürün kavramlarıyla açıklanabileceği temeline dayanır ve diğer kültürlerle karşılaştırma imkânı tanımaz (Çukur, 2007). Örneğin Türk kültürel öğeler taşıyan masallar, emik yaklaşımın parçalarıdır ve bu nedenle her kültür tarafından bilinmeyeceği gibi farklı biçimlerde de yorumlanabilirler. Bu nedenle karşılaştırmalı araştırmalarda, ölçme aracının emik öğeler içermemesine dikkat edilmelidir.

Etik yaklaşım ise araştırma konusunun evrensel olduğunu varsaymaktadır (Sargut, 2001). Etik yaklaşım, araştırma konularının evrensel nitelik taşıdığını savunmakta ve tüm kültürlerde ortak ve toplumdan bağımsız öğelerin var olduğunu

belirtmektedir. Buradan yola çıkarak, karşılaştırmalı kültür çalışmaları yapmak temel amacdır ve karşılaştırılan ögeler, kıyaslama yapılması hedeflenen gruplar arasında var olmalı ve aynı anlamsal özellikleri taşımalıdır (Çukur, 2007). Örneğin ‘savaş’ bütün kültürlerde bilinen, tanınan bir içeriğe sahip olduğu için etik yaklaşımın bir ögesidir.

Karşılaştırmalı çalışmalarda yapı eşdeğerliliğinin sağlanmadığına yönelik bulgulara ulaşılması, özgün bir kültürün ya da o kültürün emik ögeler taşıdığına bir göstergesi olabilir (Hulin ve Mayer, 1986).

Kültür davranışlarımıza şekil veren büyük bir etmen olduğu için yapılacak ölçek uyarlama çalışmalarında dilsel ve kültürel farklılıklar göz önünde bulundurulmalı ve farklı kültürlerde farklı biçimde algılanan olaylar ya da tam karşılığı olmayan terimler dikkatle kullanılmalıdır. Hambleton ve Patsula (1999), uyarlama çalışmalarında amaç kültürlerarası karşılaştırma yapmak değilse yeni bir ölçek geliştirmenin, uyarlama sırasında karşılaşılabilecek birçok sorunu ortadan kaldırması bakımından daha uygun olacağını düşünmektedirler.

Öğrencilerin okuma, matematik ve fen becerilerinin ölçüldüğü ve ülkelerin eğitim sistemlerine yönelik bir değerlendirme ve karşılaştırma imkânı sunan The Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS) ve Progress in International Reading Literacy Study (PIRLS) gibi sınavlar birçok dilde uygulanmaktadır. Deniz (2007)’e göre başka bir kültürde kullanılmak istenen bir ölçek varsa uyarlanacak olan ölçeğin içeriği oluşturulurken, ölçeğin hem asıl kültürüne hem de uyarlanacağı kültüre hâkim ve söz konusu olan kültürleri birçok yönden kıyaslayabilecek bilgiye sahip uzmanlara ulaşılmalıdır. Eğer sözü geçen ölçek farklı bir dile uyarlanacaksa her iki dile de hâkim olan alan uzmanlarına ulaşılmalıdır.

2.4.3 Psikolojik Ölçme Araçlarını Uyarlama Süreci

PISA uygulamalarında farklı dillere uyarlanarak ölçme yapılması dolayısıyla bu tür uygulamaların uyarlama süreci büyük önem taşımaktadır. Bir ölçek üzerinde farklı bir kültürde kullanılmak üzere gerekli değişikliklerin yapılarak, o kültürü iyi yansıtabilmesini sağlamak amacıyla çeşitli yolların izlenmesi süreci ölçek uyarlama süreci olarak adlandırılabilir. Ölçeğin başka bir dile tercüme edilerek kullanılması dışında o ölçeğe yönelik olarak geçerlilik ve güvenilirlik gibi temel psikometrik

işlem ya da analizlerin de yapılması, ölçeğin farklı bir dil ve kültüre uyarlanması olarak tanımlanmaktadır (Deniz, 2007).

Hambleton ve Patsula (1999) ölçek uyarlama çalışmalarına neden başvurulduğu konusunda beş konunun önemli olduğunu vurgulamaktadır:

1. Ölçek uyarlamak, yeni bir ölçek geliştirmekten çok daha kolay ve hızlı olarak düşünülmektedir.
2. Eğer testin amacı kültürlerarası ya da uluslararası bir değerlendirme yaparsa testi uyarlamak eşdeğer bir test geliştirmekten daha etkili bir yöntemdir.
3. Yeni bir ölçek geliştirmek için yeterli uzmanlığa sahip olunmayabilir.
4. Eğer ölçeğin orijinali iyi biliniyorsa ve tanınıyorsa o ölçeği uyarlamak yenisini geliştirmekten daha memnun edici ve güven vericidir.
5. Katılımcılar için adil olmak, ancak o testin birden çok dildeki formlarının olmasıyla sağlanabilir.

Yukarıda yer alan maddelerde uyarlama çalışması yapmanın yeni bir test geliştirmekten daha kolay olduğunun düşünüldüğü belirtilmesine rağmen, uyarlama çalışması yapmak düşünüldüğü kadar kolay olmamakla birlikte, dikkat edilmesi gereken kuralları ve takip edilmesi gereken bazı adımları vardır. Test veya ölçeklerin çevrilmesi/uyarlanması Alfred Binet tarafından geliştirilen zekâ testlerine kadar uzanmasına rağmen, son zamanlarda testlerin bir dil ya da kültürden diğerine çevrilmesinde ya da uyarlanmasında gelişmiş, çeşitli teknik ve yöntemlerin kullanılması yönünde oldukça fazla öneri bulunmakla birlikte, araştırmacıların ve ölçek geliştiren uzmanların bu konuya normalde verdikleri önem ve dikkatten daha fazlasını vermeye özen göstermeleri gerektiği belirtilmektedir (Hambleton ve Patsula, 1999).

2.4.3.1 Ölçek Uyarlama Adımları

Uyarlanan ölçekler bu süreçte çeşitli adımlardan geçmekte ve orijinaline yakın testler oluşturmak için çeşitli yollar izlenmektedir. Ölçek uyarlama sürecinde izlenecek adımlar Hambleton ve Patsula (1999) tarafından 13 başlıkta ele alınmış ve ilgili başlıklar aşağıda sunulmuştur.

1. İlgili grupların dil ve kültürel açıdan yapı eşdeğerliliğini sağladıklarından emin olunması: Bu adımda bazı sorulara cevap aranmaktadır:

- a) Araştırılacak olan yapı her iki kültürde de mevcut mudur?
 - b) Bu farklı iki kültürü bu yapıda karşılaştırmak mümkün ve mantıklı mıdır?
 - c) Bu yapı üzerinde kültürler arası karşılaştırma yapmak anlamlı bir sonuç ortaya koyar mı?
 - d) Ölçülen yapı bütün kültürler için aynı anlamı ifade etmekte midir?
2. Ölçek uyarlamasının mı yoksa yeni bir ölçek geliştirmenin mi daha uygun olacağına karar verilmesi: Ölçek uyarlamasının ve ölçek geliştirmenin avantaj ve dezavantajları dikkate alınarak hangi durumlarda uygun olacağı düşünülmelidir.
 3. Yüksek nitelikli çevirmenler seçilmesi: İki dilde de akıcı ve üst seviyede dil becerisine sahip, iki kültüre de hâkim ve aynı zamanda ölçülmesi hedeflenen yapı konusunda bilgi sahibi çevirmenlere ya da uzmanlara ulaşılmalıdır. Üstelik farklı bakış açılarından yararlanmak ve çevirinin uygun biçimde gerçekleştiğine yönelik karar vermek için birden fazla çevirmene ihtiyaç vardır.
 4. Ölçeğin hedef dile çevrilmesi ve uyarlanması: Çeviri konusunda yapılacak uygulama bir dilden diğer dile çevirme (forward translation) ya da bir dilden diğerine çeviri yapıldıktan sonra tekrar asıl diline çevirme (back translation) biçiminde olmalıdır.
 5. Ölçeğin uyarlanmış halinin gözden geçirilmesi ve gerekli düzeltmelerin yapılması: Çeviri, tek yönlü tercüme (forward translation) yöntemi kullanılarak yapıldığında farklı bir grup çevirmen, uyarlanan ölçeği inceleyerek, hedef kültürde ölçeğin yaratabileceği olası anlam farklılıkları olup olmadığını kontrol etmelidirler. Çift yönlü tercüme (Back translation) yöntemi kullanılarak yapılan çeviri işlemlerinde ise çevirmenler ölçeğin orijinal formunu ve çift yönlü tercüme yapılmış formunu karşılaştırarak, iki formun eşdeğer olup olmadığını incelemelidirler. Bu iki tercüme yöntemi uyarlama için yeterli olmayabilir, gerektiğinde ölçeğin orijinal halinde ya da uyarlama aşamasında değişiklikler yapılabilir.
 6. Ölçeğin uyarlanmış halinin pilot uygulamasının yapılması: Uyarlanan ölçeğin geçerliliğinin sağlanması için sadece çevirmenlerin ya da uzmanların görüşlerinin alınması yeterli değildir. Pilot uygulama yapılarak, uyarlanan ölçeğin hedef kitleyi temsil edecek büyüklükteki bir grup katılımcı üzerinde

uygulanmasıyla ölçekle ilgili gerekli düzeltmeler olup olmadığı incelenmelidir.

7. Olası hedef kitleyi temsil edecek daha büyük bir grup üzerinde yapılan uygulamadan elde edilen veriler ile ölçeğin geçerlik, güvenirlik hesaplamaları ve madde analizlerinin yapılması: Belirlenen daha büyük bir grup üzerinde uygulama yapılarak elde edilmiş ölçeğin geçerliğine yönelik elde edilen istatistiksel veriler, ölçeğin özgün formunun sahip olduğu değerlerle karşılaştırılarak yorumlanmalıdır.
8. Ölçeğin özgün formu ve hedef dil formundan elde edilen puanlar arasındaki ilişkiyi bulmak için istatistiksel bir desen belirlenmesi: Bu adım, özellikle kültürlerarası karşılaştırma yapılacak olan araştırmalarda ya da kaynak kültürde norm ya da performans standartları varsa uygulanmalıdır. Bu aşamada hangi desenin kullanılacağına karar verilmelidir. Desenler, farklı formlardan elde edilen puanların ortak bir ölçeğe yerleştirilmesine yardımcı olur. En yaygın kullanılan desenler aşağıda gösterilmektedir:
 - a) İki dilli grup deseni,
 - b) Tek dilli karşılaştırmalı grup deseni,
 - c) Tek dilli grup deseni.
9. Kültürlerarası karşılaştırma yapılacaksa testin farklı dildeki formlarının eşit olması: Ölçek evrenden alınan örnekleme uygulanarak, testin farklı formlarının eşitliğini test etmek için genellikle madde yanlılığına yönelik çalışmalar yapılır. Maddeler gruplar arasında farklı fonksiyon gösteriyorsa bu maddeler yeniden yazılmalı, yeniden tercüme edilmeli, yeniden uygulanmalı ve yeniden analiz edilmelidir.
10. Geçerlik çalışmalarının yapılması: Geçerliğin test edilmesi için uygun bir yöntem kullanılmalıdır; yapı geçerliğini sağlamaya yönelik olan faktör analizi ya da deneysel veya ilişkisel çalışmalar yapılmalıdır.
11. Testi uyarlama sürecini içeren bir rapor ve uyarlanan testin kullanıcıları için bir el kitabı hazırlanması: 1. adımdan 10. adıma kadar yapılanlar ve elde edilen sonuçlar bir rapor halinde hazırlanmalı ve raporun içerisine uyarlanan testi kullanmak isteyenlere yönelik bir de kılavuz oluşturularak eklenmelidir.
12. Kullanıcılara eğitim verilmesi: Hazırlanan rapor ve kılavuz, kullanıcılara yol gösterecek olmasına rağmen, gerektiğinde kullanıcıların olası sorularını

cevaplamak amacıyla testi uygulayacak ya da kullanacak olan kişilere eğitim verilmelidir.

13. Uyarlanan testin zaman içerisinde takibinin yapılması: Uyarlama yapan araştırmacılar ölçeğin sürekliliği sağlamak için ölçeği takip etmelidirler. Tekrardan geçerlilik ve güvenilirlik çalışmaları yapılmalıdır, çünkü zaman içerisinde ölçeğin geçerlilik ve güvenilirlik değerleri değişen koşullar, müfredat vb. nedenlerden dolayı değişebilir. Karşılaştırmalı kültür araştırmalarında kullanılan testler genelde bir defaya mahsus oldukları için bu tür testlerde bu duruma ihtiyaç olamayabilir, fakat zekâ, yetenek ve kişilik testleri geliştirilirken ya da uyarlanırken sürekli kullanım hedeflendiği için bu testlerin zaman içerisinde değişen durum ve koşullar altında hala geçerli olup olmadıkları takip edilmelidir.

2.4.3.2 Uyarlama Çeşitleri

Malda, Van de Vijver, Srinivasan, Transler, Sukumar, Rao, (2008) uyarlama sürecine kolaylık sağlaması ve katkıda bulunması açısından uyarlama yöntemlerini sınıflandırmışlardır.

- a) Yapıya yönelik uyarlama: Bazı psikolojik kavramlar farklı biçimde tanımlanabilir. Örneğin zekâ ölçülmek isteniyorsa hedef kültürdeki zekâ tanımına uygun bir biçimde uyarlama yapılmalıdır.
- b) Dile yönelik uyarlama: Hedef dilde tercüme edilecek kelimenin anlamsal karşılığı olmayabilir. Hollanda’da distress kelimesinin tam karşılığı yoktur. Friend kelimesi İngilizce’de hem kız hem de erkeğe yönelikken, Almanca’da freund kelimesi erkeğe, freundin kıza yöneliktir.
- c) Kültüre yönelik uyarlama: Kültürel normlar, değerler, gelenekler farklıdır. Doğum günü kutlama, Cadılar bayramı kutlama vb.
- d) Kurama yönelik uyarlama: Kuramsal nedenlerden dolayı bazı kelimelerin benzer uzunlukta kalması gerekebilir. Örneğin zekâ testlerinde çok sık kullanılan ‘digit span’ olarak adlandırılan teknik kısa süreli bellekte depolanan kavramları hatırlama tekniğidir. Bu kavramların uyarlama esnasında uzunluklarının benzer biçimde kalması gerekir.
- e) Aşına olmaya/tanımaya yönelik uyarlama: Testte yer alan maddeler yabancı olabilir, daha önceden bilinmiyor olabilir. Eğer kişinin çevresinde ya da bir

resimde daha önceden görmediği bir nesneden bahsediliyorsa testin amacına zarar vermeyecek biçimde o ifadeler yerel ifadelerle değiştirilebilir.

2.5 Ölçme Değişmezliğini Etkileyen Faktörler

Farklı gruplar arası yapılan çalışmalarda, eşitliğin sağlanamamasının en temel nedeni çeşitli yanlılıklardır (Van de Vijver ve Leung, 1997). Yanlılık, ölçülen yapıyla ilişkili gözlenen puanların sistematik olarak gerçek puanları yansıtamadığı durumlarda ortaya çıkar. Yanlılık, farklı kültürlerde ya da gruplarda uygulanan ölçme aracının geçerliliğine zarar veren sorunlu, ölçülmesi hedeflenen yapıdaki grup farklılıkları ile açıklanamayan faktörlerdir (He ve Van de Vijver, 2012). Bu nedenle yanlılık, ölçme aracının ölçme değişmezliğini ve dolayısıyla geçerliliğini olumsuz yönde etkileyerek, gruplar arası yapılan karşılaştırmaların da geçerliliğini zayıflatır (Kankaraş ve Moors, 2010). Geçerli ve güvenilir bir karşılaştırma sunabilmek için, yanlılık ve değişmezlik karşılaştırmalı kültür çalışmalarının anahtar parçalarıdır (He ve Van de Vijver, 2012). Yanlılık konusunu Van de Vijver ve Tanzer (2004) bir örnekle şöyle açıklamışlardır:

Bir coğrafya testinde ‘Polonya’nın başkenti neresidir?’ biçiminde bir soru olduğunu ve bu testin geniş çapta birçok ülkede uygulandığını düşünelim. Doğru cevabın verilmesi birçok faktöre bağlıdır: Öğrencilerin zihinsel yeteneklerine, verilen coğrafya eğitimin kalitesine ve testin uygulandığı ülkelerin Polonya’ya olan uzaklıklarına. Eğer örneklem dikkatli bir biçimde oluşturulursa o zaman bu soru, bu bilgi düzeyindeki farklılıkların kıyaslanmasına olanak sağlar, fakat daha büyük bir çapta örneklem seçilir ve bu soru kullanılırsa işte o zaman ülkelerin Polonya’ya uzaklığı problem yaratan (yanlılığa neden olan) bir faktör olacaktır.

Yanlılık karşılaştırılan gruplara göre ortaya çıkmaktadır; karşılaştırılan bir grup arasında yanlı olduğunu düşünülen bir maddede, karşılaştırılan başka bir grupta yanlılık tespit edilmeyebilir. Bu durum da, karşılaştırılan grupların kültürel, coğrafi vb. özelliklerinin benzer ya da benzer olmamasıyla açıklanabilir.

Teorik yapı, ölçme aracı ya da belirli maddelerden kaynaklanmak üzere üç tür yanlılığın bulunduğu belirtilmektedir. Bunlar, yapı yanlılığı (construct bias), yöntem yanlılığı (method bias) ve madde yanlılığı (item bias)’dır (Van de Vijver ve Leung, 1997; Van de Vijver ve Tanzer, 2004). Yapı yanlılığı, yöntem yanlılığı ve madde

yanlılığını en aza indirmek ya da bu yanlılık türlerinden her hangi biriyle karşılaşmamak için uygulanacak ölçeğin hedef dillere tercümesi yapılmadan önce, ya da farklı kültürlere uyarlanmadan önce, ölçeğin kapsam ve yapısının uygulanacak olan kültüre ne derece uygun olduğu belirlenmelidir. Tercüme işlemine gelindiğinde ise birçok araştırmacı (Hambleton ve Patsula, 1999; Van de Vijver ve Hambleton, 1996; Van de Vijver ve Leung, 1997; Hambleton ve de Jong, 2003) Uluslararası Test Komisyonu tarafından önerilen tercüme basamaklarını da dikkate alarak izlenmesi gereken adımlardan ve bunların gerekliliğinden bahsetmişlerdir.

Yukarıda uluslararası düzeyde uygulanan PISA sınavı, ölçme değişmezliği ve uyarlama çalışmaları ve yanlılıklar konusunda verilen bilgiler doğrultusunda, bu araştırmanın problemini PISA 2012 okuduğunu anlama becerisi uygulamasının teknik niteliklerinden bir tanesi olan yapı geçerliliğinin ortaya koyulması gerekliliği oluşturmaktadır. Ölçme aracının faktör yapılarının seçilen ülkelerde farklı çıkması durumunda, söz konusu ülkeler arasında yapılan sıralamanın ya da alınan puanların karşılaştırılmasına şüpheyle yaklaşılabilecektir. Çünkü faktör yapıları açısından farklı olan ölçme araçları aynı kabul edilen ölçme araçları olsa da, farklı ölçekler gibi rol oynamaktadırlar ve bu nedenle birbirlerine denk olarak kabul edilmezler. Böyle bir sonuç elde edilirse PISA 2012 uygulaması bakımından oluşturulan başarı sıralarının sözü geçen ülkeler için geçerliliği şüpheli hale gelecek ve çeşitli yanlılık şüpheleri artacaktır. Gruplar arasında faktör deseninin eşit olduğuna yönelik bir sonuç elde edilir ise bu uygulamanın uyarlama çalışmalarının kültürel farklılıklar göz önünde bulundurularak yapıldığı ve uygulama sonucundaki başarı puanları arasındaki farklılığın gerçek anlamda katılımcıların performansından kaynaklandığı söylenilebilir.

2.6 İlgili Araştırmalar

Literatürde, farklı gruplara ya da kültürlere uygulanan geniş ölçekli testler için ölçme değişmezliğini sağlayıp sağlamadığı veya yapısal açıdan eşdeğer olup olmadıklarını araştıran ilgili sınırlı sayıda çalışma bulunmaktadır.

Öğretmen (2006), PIRLS 2001 verilerini kullanarak gerçekleştirdiği araştırmasında, Amerika Birleşik Devletleri ve Türkiye örneklemelerinden elde edilen verilerin faktör yükleri ve hata terimlerinin gruplar arası farklılık sergilediği belirlemiş ve bu durumda yapı eşdeğerliliğine sahip olmadığını belirtmiştir.

Akyıldız (2009) tarafından yapılan başka bir arařtırmada ise PIRLS 2001 sınavının faktör yapısının, uygulanan bütün ülkelerde (35 ülke) orta düzeyde eşdeğer olduğu belirtilmiştir.

Uzun ve Öğretmen (2010) tarafından yapılan başka bir arařtırmada ise TIMSS 1999 uygulamasında fen başarısını etkileyen faktörlerin (özyeterlik, önem, tutum ve sınıf içi öğrenci etkinlikleri) cinsiyete göre ölçme değişmezliği araştırılmıştır. Modelde metrik değişmezliğin sağlandığı, fakat katı değişmezlik koşulunun sağlanmadığı belirtilmiştir.

Elosua ve Mujika (2013) PISA 2009 uygulaması okuma becerilerine yönelik olarak İspanya'nın çeşitli bölgeleri için ölçme değişmezliğinin sağlanıp sağlanmadığını incelemiştir. Bask, Katalan ve Galiçya dil formlarının eşdeğer olup olmadığının araştırıldığını bu çalışmada, testlerin yapısının eşdeğer olduğu sonucuna ulaşılmıştır.

Uyar ve Doğan (2014) tarafından yapılan bir arařtırmada, PISA 2009 Türkiye örnekleminde öğrenci anketinde, öğrenme stratejilerine dair bir model test edilmiş ve modelin cinsiyet, okul türü ve istatistiksel bölge gruplarında ölçme değişmezliği incelenmiştir. Modelin cinsiyet ve okul türü gruplarında sadece yapısal ve metrik değişmezliği sağladığı, bölgelerde ise tüm değişmezlik koşullarını sağladığı sonucuna ulaşılmıştır.

Wu ve diğerleri (2007), TIMSS 1999 uygulamasında incelenen matematik testlerinin, farklı kültürlerde yapısal (Yeni Zelanda- Japonya, Amerika- Japonya) ve metrik değişmezliğin (Avustralya-Japonya, Avustralya-Amerika, vb.) sağlandığını, benzer kültürlerde (Kanada-Amerika, Avustralya-Yeni Zelanda vb.) ise katı değişmezliğin sağlandığını göstermişlerdir.

Güzeller (2011) PISA 2009 uygulamasına ait öğrenci anketinde yer alan bilgisayar tutum boyutunun kültürlerarası eşitliğini incelemiştir. Arařtırmaya, Çek Cumhuriyeti, Avustralya, Belçika, Danimarka, Macaristan, İrlanda, İsrail, Japonya, Norveç, Çin olmak üzere 10 ülke dâhil edilmiştir. Bu amaçla yapılan çoklu grup doğrulayıcı faktör analizi sonucunda, bilgisayar tutumuna yönelik aracın kültürlerarası eşdeğer olduğu ortaya koyulmuştur.

Ardıç ve Güzeller (2015) tarafından yapılan diğer bir arařtırmanın kapsamını ise PISA 2012 uygulamasına ait bilgi ve teknoloji kullanımına yönelik olan ölçek maddeleri oluşturmaktadır. Türkiye, Yeni Zelanda ve Singapur örneklemi

arasındaki ölçme değişmezliği aşamalı olarak analiz edilmiş ve analiz sonucunda ülkeler arası sadece yapısal değişmezlik koşulunun sağlandığı belirtilmiştir.

Evran (2012) çok gruplu faktör analizi yöntemiyle, PISA 2009 uygulamasının öğrenci anketinde yer alan sorularına yönelik, Türkiye ve Amerika katılımcıları arasında ölçme değişmezliğinin sağlanıp sağlanmadığına kanıt olabilecek madde yanlılığını incelemiştir. Madde yanlılığı bulunduğu biçiminde bazı bulgulara rastlansa da, bu bulguların anlamlı olup olmadığı konusunda bir netliğe ulaşamamıştır.

Yıldırım (2006) TIMSS 1999 ve PISA 2003 uygulamalarının matematik testlerinin kültürler arası eşdeğerliliğini ve matematik başarı kavramının altında kültürlere has öğeler ya da yapılar olup olmadığını araştırmıştır. Farklı kültür lehine ya da aleyhine çalışan maddeler değerlendirilmiştir. Araştırma sonucunda, TIMSS ve PISA maddelerinde kültürel açıdan denk olma konusunda problemler olduğunu görülmüştür; özellikle TIMSS uygulamasının farklı kültürlerdeki faktör yapısının birbirine denk olmadığı saptanmıştır.

Arim ve Ercikan (2014) TIMSS 1999 uygulamasının matematik başarı testi sonuçlarına göre Türk ve Amerikalı öğrencilerin karşılaştırılıp karşılaştırılmayacağını araştırmıştır. Matematik başarı testinde yer alan maddelerin %23 ünün bu iki grup arasında farklı biçimde çalıştığını belirtmiş ve testlerin faktör yapılarının da denk olmadığını ortaya koymuştur.

Ercikan ve Koh (2009) başka bir araştırmada, TIMSS 1995 uygulamasının İngilizce ve Fransızca dil formlarındaki matematik ve fen bilimleri testlerinin eşdeğerliliğini incelemişlerdir. Araştırmada Kanada, Amerika, İngiltere ile Fransa örneklemini ayrı ayrı karşılaştırılmış ve faktör yapısının en çok farklılaştığı grubun matematik testi verilerinde Amerika-Fransa olduğu belirtilmiştir. Çok gruplu doğrulayıcı faktör analizi sonucu, matematik testi için yapı eşdeğerliliğinin Amerika ve Fransa karşılaştırmasında hiç bir kitapçık için sağlanamadığı, Kanada ve İngiltere ile yapılan Fransa kıyaslamasında ise arasından sadece üç kitapçığın yapı eşdeğerliliğine sahip olduğu sonucuna ulaşılmıştır. Fen bilimleri testi için ise Amerika-Fransa arasında hiçbir kitapçığın ölçme değişmezliği koşulunu sağlamadığı, Kanada ve İngiltere ile karşılaştırılan Fransa örneklemine ait verilere göre ise sadece bir kitapçığın ölçme değişmezliği koşulunu sağladığı belirtilmiştir.

Arıkan (2010) tarafından gerçekleştirilen bir çalışmanın bir bölümünde Öğrenci Seçme Sınavı (ÖSS) matematik bölümlerinin cinsiyet ve okul türleri

bakımından eşitliği incelenmiş ve faktör yapıları bakımından ÖSS'nin bu gruplar açısından benzerlik gösterdiği görülmüştür. Karşılaştırılan gruplar (kız-erkek, devlet okulu-özel okul) arasında değişmezlik modellerin iyi uyum gösterdiği belirtilmiş ve faktör yükleri, faktör korelasyonları ve hata varyanslarının eşdeğer olduğu ifade edilmiştir.

Başusta ve Gelbal (2015) tarafından gerçekleştirilen çalışmada PISA 2009 öğrenci ölçek verileri kullanılmış ve dört aşamalı (biçimsel değişmezlik, metrik değişmezlik, ölçek değişmezliği ve katı değişmezlik) analiz sonrası elde edilen değerlerin cinsiyete göre ölçme değişmezliğini sağladığı bulgusuna ulaşılmıştır. Böylece öğrencilerin fen bilimlerine yönelik tutumlarının, ilgilerinin vb. ortaya çıkarılmaya çalışıldığı ölçme aracının sahip olduğu psikometrik özelliklerinin farklı cinsiyet gruplarına genellenebileceği ortaya çıkmıştır.

Bahadır (2012), PISA 2009 uygulamasına ait okuma becerileri verilerini kullanarak, okuma becerilerini etkileyen değişkenlerin modellemesini yapmış ve modele alınan değişkenlerin bölgeler arasında ölçme değişmezliğine sahip olup olmadığını incelemiştir. Ölçek değişmezlik koşulu sağlanmadığı için kısmi ölçme değişmezliğini test edilmiş ve bölgeler arasındaki en büyük farklılığın anne- baba eğitim düzeyi ve öğrenciye sağlayan olanaklardan kaynaklandığı belirtilmiştir.

Stricker ve Rock (2008) TOEFL IBT testinin, ana dili İngilizce olanlar ve yabancı dil ya da ikinci dil olarak İngilizce öğrenenler katılımcılardan oluşan alt gruplarda faktör yapısının değişmez olup olmadığını araştırmışlardır. Alt gruplar; Hint-Avrupa dillerini konuşanlar ya da konuşmayanlar, dilin eğitim amaçlı ya da mesleki amaçlı kullanımı, sınıf ortamında İngilizce eğitim aldığı ya da bu dile maruz kaldığı yıl sayısı biçimindedir. Yapılan analiz sonucunda katılımcıların oluşturduğu bütün alt gruplarda, testin aynı biçimde işlev gösterdiği, diğer bir ifadeyle ölçme eşdeğerliliğine sahip olduğu belirtilmiştir.

Shulz (2005), PISA 2006 ön test uygulamasında, öğrenci anketinde yer alan maddelerin 16 ülke arasında ölçme değişmezliği koşulunu sağlayıp sağlamadığını incelemiştir. Analiz sonucunda benzer dil ve kültüre sahip ülkeler arasında sadece zayıf değişmezliğin sağlandığı bulgusuna ulaşılmıştır.

Çetin (2010), PISA 2006 uygulamasında yer alan anket sorularının kültürler arası genellenebilirliğini incelemek amacıyla çok gruplu doğrulayıcı faktör analizi yöntemini kullanarak, 10 ülke arasında faktör yapılarının değişmezliğini

araştırmıştır. Analiz sonucunda, faktör yüklerinin ülkeden ülkeye değişmediği, fakat hata varyanslarının değişebildiği görülmüştür.

Asil ve Gelbal (2012) tarafından da benzer bir araştırma yürütülerek, PISA 2006 kapsamında kullanılan öğrenci anketinin kültürler arası eşdeğerliliği araştırılmıştır. Araştırmaya Avustralya, Yeni Zelanda, Türkiye ve Amerika Birleşik Devletleri örneklemi dâhil edilmiştir. Doğrulayıcı faktör analizi sonucunda, ölçme modelinin bütün örneklerde aynı faktör yapısına sahip olduğu bulgusuna ulaşılsa da DMF gösteren maddelerin var olması nedeniyle uygulanan anketlerin eşdeğer olmadığı söylenmiştir. En çok DMF gösteren madde sayısı (9) Avustralya ve Türkiye karşılaştırılmasındadır. Avustralya ve Yeni Zelanda arasında ise her hangi bir yanlılığa rastlanmamıştır. Avustralya ve Amerika örneklemi için ise iki maddenin yanlı çalıştığı belirtilmiştir. Bu durum uzmanlar tarafından incelenerek, bu duruma kaynak teşkil edebilecek iki neden öne sürülmüştür, bunlar; uyarlama ve çeviri ile ilgili problemler ya da kültürel farklılıklardır.

Çıkrıkçı ve Ulutaş (2015), PISA 2006 Fen okuryazarlığı testine yönelik, Türkiye ve Amerika örneklemi arasında faktör yapısının farklılaşp farklılaşmadığını görmek amacıyla çok gruplu doğrulayıcı faktör analizi yaparak, her iki kitapçığın eş değer olduğu kararına varmışlardır. Madde yanlılığının da araştırıldığı bu çalışmada, maddelerde gözlenen yanlılık düzeyinin katılımcıların performansını değiştirecek düzeyde olmadığı belirtilmiştir.

Alivernini (2011) PIRLS 2006 uygulamasına yönelik olarak okuma becerileri alt testinde cinsiyete göre, bölgeye göre ve katılımcıların göç etmiş olup olmama durumlarına göre, ölçme değişmezliğini incelemiştir. Bölgelere göre yapılan analizde şekilsel değişmezliğin sağlanamadığını ortaya çıkmıştır. Göç durumu ve cinsiyet değişkenleri açısından metrik değişmezliğin sağlandığı, fakat bu değişkenler için ölçek değişmezliği kanıtına ulaşamadığı biçiminde sonuçlanmıştır. Araştırma, cinsiyet ve göç durumuna göre oluşturulan grupların birbiri ile karşılaştırılabileceği, ancak farklı coğrafik bölgede yaşayan katılımcılar için karşılaştırmanın uygun olmayacağı, bu katılımcıların soruları farklı biçimde algılamış olabileceği biçiminde yorumlanmıştır.

Asil ve Brown (2016), PISA 2009 uygulamasında katılan ülkelerin, okuduğunu anlama testinden elde edilen sonuçların ne derece birbiriyle karşılaştırılabilir olduğunu incelemek amacıyla bütün ülkelerde ortak olarak kullanılan altı kitapçıktan bir tanesini seçerek, Avustralya örneğini baza alarak ülkeler arası

değişmezliği incelemişlerdir. Araştırmaya dâhil edilen 55 ülkeden elde edilen verilen şekilsel değişmezlik koşulunu sağladığı, metrik değişmezliğin kabul edilebilir düzeyde olduğu, fakat neredeyse bütün ülkeler için ölçek değişmezliği koşulunun sağlamadığı sonucuna ulaşılmıştır. Ölçek değişmezliği Avustralya ile karşılaştırılan sadece üç ülke (Yeni Zelanda, Kanada ve Amerika) ile sağlanabilmiştir.

Arffman (2010), PISA 2000 uygulamasının alt testi olan okuma becerileri testinin tercümesinin eşdeğer olup olmadığını araştırmak için, bu testte kullanılan üç okuma parçası seçerek eşdeğerliliği etkileyebilecek altı problemi (dil bilgisi farklılıkları, yazım farklılıkları, anlam farklılıkları, kültür farklılıkları, çevirmenlerin etkisi, basımla ilgili problemler) metinlerin İngilizce ve Fince formlarını karşılaştırarak tespit etmeye çalışmıştır. Yapısal, anlamsal, sözdizimi, kelime kullanımı vb. açılardan ele alınan metinlerin farklı dil formlarının tam olarak birbirleriyle eşdeğer olmadığı belirtilmektedir. Tercüme edilen metinler, seçilen kelime ve cümle yapısına göre bazen orijinaline kıyasla kolay anlaşılır, bazen de daha zor anlaşılır hale gelmektedir. Bu doğrultuda, birden fazla dilde formu bulunan metinlerin tam anlamıyla ölçme değişmezliğini sağlamanın pek mümkün olmadığı belirtilmektedir. Grisay ve Monseur (2007) de benzer olarak, bir ölçme aracının başka bir dilde kullanılmak üzere tercümesi yapıldığında, o aracın eş değerliliğinin kaybolacağını belirtmektedirler.

Alanyazındaki çalışmalar incelendiğinde, çalışmaların genelinin örneklemine öğrenci anketlerinin ya da matematik ve fen okuryazarlığı becerisinin verilerinin oluşturduğu, fakat PISA kapsamında okuma becerisine yönelik öğrenci performansının ölçme değişmezliğinin (madde düzeyinde ve ölçek düzeyinde) araştırmacılar tarafından yeterince ele alınmadığı görülmektedir. Bu araştırma kapsamında ele alınacak konunun, PISA 2012 uygulaması okuma becerilerinin ölçme değişmezliğinin test edilmesi olması, bu konuda alanyazında bulunan eksikliklerin giderilmesi açısından oldukça önem taşımaktadır. Okuduğunu anlama performansı, direkt olarak katılımcıların sahip oldukları dil özellikleri ve kültürden etkilenebileceği için, bu alt testin gruplar arası karşılaştırma yapmaya uygun olup olmadığının araştırılması ve bu yönde ölçeğe yönelik gerekli psikometrik niteliklerin belirlenmesi gerektiği düşünülmektedir.

BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın modeli, evreni, verilerin toplanması, verilerin analizi ve varsayımların test edilmesi sunulmuştur.

3.1 Araştırmanın Modeli

Bu araştırma ile PISA 2012 uygulaması alt testinden bir tanesi olan okuma becerilerine ait verilerin, aynı dil formları ve farklı dil formları arasında, ölçme değişmezliğinin incelenmesi hedeflenmektedir. Bu çerçevede yapılan araştırmanın ilişkisel araştırma modeli bağlamında yürütüldüğü ifade edilebilir.

3.2 Araştırma Verileri

PISA 2012 uygulamasına katılan 65 ülke olup, on beş yaş grubundan yaklaşık olarak 510000 öğrenci katılmıştır (MEB, 2015). Bu çalışma için okuma becerileri testine yönelik, kitapçıklardaki madde sayıları ve kitapçığı kullanan örneklem büyüklüğü dikkate alınarak 13 farklı kitapçık arasından, analize dâhil edilecek bütün gruplarda kullanılmış olması ve en çok madde sayısına sahip olması nedeniyle 2 numaralı kitapçık seçilmiş ve çalışmaya birinci alt problem için; Amerika, Avustralya, İngiltere, İrlanda, Kanada, Singapur ve Yeni Zelanda ülkeleri dâhil edilmiştir. Bu ülkelerde bütün kitapçıklardaki örneklem sayısı yaklaşık 50000'dir. 2 numaralı kitapçıkta kayıp verilerin ve İngilizce dışında uygulanan ölçek formlarının çıkartılmasıyla birlikte Amerika örnekleme 389, Avustralya örnekleme 1032, İngiltere örnekleme 836, İrlanda örnekleme 358, Kanada örnekleme 1101, Singapur örnekleme 378, Yeni Zelanda örnekleme 296 olmak üzere toplam sayı 4390 olmuştur. Bu doğrultuda araştırmanın örneklemini yedi ülkeden toplamda 4390 öğrenci oluşturmaktadır.

İkinci alt problemin örneklemini; Belçika, Fransa, İsviçre, Kanada ve Lüksemburg ülkelerindeki katılımcılar oluşturmaktadır. Bu ülkelerde bütün kitapçıklardaki örneklem sayısı yaklaşık 20000'dir. 2 numaralı kitapçıkta kayıp verilerin ve Fransızca dışındaki ölçek formlarının çıkartılmasıyla birlikte Belçika

örnekleme 167, Fransa örnekleme 306, İsviçre örnekleme 307, Kanada örnekleme 356, Lüksemburg örnekleme 74 olmak üzere, araştırmanın örneklemini toplamda 1210 öğrenci oluşturmaktadır.

Üçüncü alt problemin örneklemini ise Amerika, İspanya, Tayvan ve Türkiye katılımcıları oluşturmakta olup, bu ülkeler için bütün kitapçıklardaki toplam katılımcı sayısı yaklaşık 35000 olup, kayıp verilerin atılmasıyla 2. kitapçıkta bulunan Amerika örnekleme 389, İspanya örnekleme 1489, Tayvan örnekleme 441 ve Türkiye örnekleme 349 katılımcıdan oluşmakta ve toplam sayı 2668'dir.

Gruplar belirlenirken, ilk olarak aynı dillerin konuşulduğu ve aynı ölçek formlarının uygulandığı farklı ülkelerdeki değişmezliği, ikinci olarak da farklı dillerin konuşulduğu ülkelerdeki değişmezliği incelemek hedef alınmıştır. Aynı dili kullanan ülkeler İngilizce ve Fransızcanın hâkim olduğu ülkeler olarak belirlenmiştir. Farklı dili konuşan ülkeler belirlenirken Türkiye'nin dünyada en yaygın konuşulan dillerinin hâkim olduğu ülkelerle kıyaslanması amaçlanmış olup, bu doğrultuda, söz konusu dillerin en fazla konuşulduğu ülkeler dikkate alınarak; çalışma grubu İngilizceyi temsilen Amerika, İspanyolca'yı temsilen İspanya ve Mandarini (Çince) temsilen Tayvan olarak belirlenmiştir.

3.3 Verilerin Düzenlenmesi

Araştırmada PISA 2012 uygulamasında okuma becerileri ölçeğindeki maddeler kullanılmış ve elde edilen veriler araştırmacılar için OECD veri tabanında sunulmakta olup, bütün veriler veri tabanından edinilmiştir. Verilere ulaşıldıktan sonra ilk olarak kullanılan kitapçıklar incelenmiş ve analize dâhil edilecek olan gruplarda ortak olarak kullanılan ve ölçme aracında en çok maddeyi içeren 2 numaralı kitapçık seçilmiştir. Seçilen kitapçıkla birlikte karşılaştırılacak olan grup verileri veri setinden ayrıştırılarak varsayım testleri ve diğer analizler için hazır hale getirilmiştir.

3.4 Verilen Analizi

Bu araştırmada alt problemler doğrultusunda ölçme değişmezliğinin sağlanma durumu Çok Gruplu DFA ile test edilmiştir. Bu doğrultuda, ilk olarak analizler için gerekli olan varsayımlar test edilmiştir.

3.4.1 Varsayımların Test Edilmesi

Varsayım analizleri öncesinde, veri setinde kayıp değer olup olmadığı kontrol edilmiştir. PISA datasında 7 ve 8 ile kodlanmış olan kayıp veriler veri setinden çıkartılmıştır. Kayıp verilerin çıkartılmasıyla 2 numaralı kitapçığın uygulandığı gruplarda birinci alt problem için toplamda 4390, ikinci alt problem için 1210, ve üçüncü alt problem için 2668 katılımcı analize dâhil edilmiştir. Örneklemin geniş olması nedeniyle, kayıp verilere sahip katılımcıların veri setinden çıkarılması kararı alınmıştır. Ayrıca 2 numara ile kodlanan açık uçlu bir soru da veri setinden çıkartılmış ve analizler 28 madde üzerinden yapılmıştır. Tek yönlü ve çok yönlü uç değer taraması yapılmış ve herhangi bir uç değer olmadığı belirlenmiştir.

Çoklu bağlantı probleminin test edilmesi amacıyla kullanılan bazı yöntemler bulunmaktadır (Çokluk ve diğerleri, 2010). Bu çalışmada değişkenler arası korelasyonlar incelenmiş ve bağımsız değişkenler için tolerans değerleri dikkate alınmıştır. İncelenen korelasyon matrislerinde, birinci alt problem için, $r < .90$ olup, korelasyonlar .00 ve .87 arasında değişmektedir. Tolerans değerlerinin de .10'dan büyük olduğu tespit edilmiştir .10'dan küçük olan değerler çoklu bağlantı problemi olduğunun göstergesidir (Field, 2009). İkinci alt problem için, $r < .90$ olup, korelasyonlar yine .00 ve .89 arasında değişmektedir. Diğer analiz çıktısı incelendiğinde tolerans değerlerinin oldukça yüksek olduğu görülmüştür. Üçüncü alt problem için, $r < .90$ olup, korelasyonlar .00 ve .81 arasında değişmekte olup tolerans değerleri de oldukça yüksektir. Bu sonuçlara göre; veri setleri çoklu bağlantı problemine sahip değildir.

Varsayım testlerinin diğer bir aşamasında verilerin normal dağılıp dağılmadığı incelenmiştir. Tablo 3.4.1.1, 3.4.1.2, 3.4.1.3'de normallik ve güvenilirlik testlerine ilişkin bulgular yer almaktadır.

Tablo 3.4.1.1. 1. Alt Probleme İlişkin Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları

Ülkeler	N	\bar{X}	X_{ort}	S	Ranj	Çarpıklık Katsayısı	Basıklık Katsayısı	KR-20
Amerika	389	17.21	18.00	5.33	27	-.420	-.321	.83
Avustralya	1032	17.25	18.00	5.24	27	-.517	-.132	.83
İngiltere	836	17.25	18.00	4.73	28	-.452	-.047	.79
İrlanda	358	18.13	18.00	4.52	21	-.151	-.660	.78
Kanada	1101	18.01	18.00	4.94	26	-.614	.211	.81
Singapur	378	19.07	20.00	5.13	25	-.711	.051	.83
Yeni Zelanda	296	18.29	19.00	5.28	27	-.690	.195	.84

Tablo 3.4.1.2. 2. Alt Probleme İlişkin Test İstatistikleri, Normallik Testleri ve güvenilirlik Katsayıları

Ülkeler	N	\bar{X}	X_{ort}	S	Ranj	Çarpıklık Katsayısı	Basıklık Katsayısı	KR-20
Belçika	167	18.53	20.00	5.43	25	-.787	.196	.84
Fransa	306	18.95	20.00	5.42	26	-.738	.004	.84
İsviçre	307	17.09	17.00	5.21	24	-.227	-.685	.81
Kanada	356	17.21	18.00	5.53	26	-.518	-.289	.84
Lüksemburg	74	17.23	17.50	6.18	28	-.499	-.063	.87

Tablo 3.4.1.3. 3. Alt Probleme İlişkin Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları

Ülkeler	N	\bar{X}	X_{ort}	S	Ranj	Çarpıklık Katsayısı	Basıklık Katsayısı	KR-20
Amerika	389	17.21	18.00	5.33	27	-.420	-.321	.83
İspanya	1489	17.09	18.00	5.28	27	-.590	-.086	.82
Tayvan	441	17.49	19.00	5.11	26	-.794	.204	.82
Türkiye	349	15.30	15.00	5.22	27	-.076	-.496	.80

Tablolarda görüldüğü üzere, merkezi eğilim ölçülerinin göreceli olarak birbirlerine yakın oldukları söylenilebilir. Çarpıklık ve basıklık katsayılarının +1 ile -1 aralığında olması, dağılımın normale yakın olduğunu göstermektedir (Metler ve Vannatta, 2005; Çokluk ve diğerleri 2010). Normal veya normale yakın bir dağılım olduğunu söyleyebilmek için, bazı araştırmacılara göre Skewness ve Kurtosis

katsayılarının +2 ile -2 arasında olması kabul edilebilirdir (Trochim ve Donnelly, 2006; Field, 2009; Gravetter ve Wallnau, 2014). Tabloda verilen katsayılar +1 ile -1 aralığında olmasına rağmen, puanların bir derece sola çarpık olduğu söylenebilir. Bu nedenle her bir grup için doğrulayıcı faktör analizleri yapılmış ve ölçme değişmezliğini test etmek adına uygulanan çoklu grup DFA asimptotik kovaryans matrisi üzerinden yapılarak, model uyumu için Satorra Bentler Chi-Square istatistiği dikkate alınmıştır. Hu ve Bentler (1999)'in önerisi doğrultusunda, χ^2 istatistiği olarak, büyük örneklerde, ya da karşılaştırılan gruplar arasında dağılımın normalden bir ölçüde sapması durumunda, dağılımın normal olduğu durumlarda elde edilen χ^2 değerine yakın değerler veren Satorra Bentler χ^2 (S-B χ^2) kullanılması tercih edilebilir. Ayrıca puanlamanın kategorik olması durumunda ve χ^2 değerleri arasında anlamlı bir farkın olup olmadığını belirlemede hesaplanacak T_s değerinde S-B χ^2 'ye ihtiyaç duyulmaktadır. Bu nedenle, ilerleyen aşamada ÇG-DFA S-B χ^2 istatistiği üzerinden yapılacaktır.

İç tutarlılığı test etmek amacıyla Kuder Richardson 20 (KR-20) formülü kullanılmıştır. Bu değer .50 ve üzerinde olması beklenmekte olup, .30'dan daha düşük değerler güvenilirliğin zayıf olduğu biçiminde yorumlanmaktadır (Chen, 2018). Ayrıca Kehoe (1995) 10-15 maddelik kısa testler için KR-20 \geq .50 olması, 50 madde ve üzerinde olan testler için KR-20 \geq .80 olması gerektiğini belirtmiştir. Bu açıklamaları dikkate alarak, alt problemlere yönelik olarak hesaplanan KR-20 güvenilirlik katsayısının bütün gruplarda .50 ve üzerinde ve 1'e yakın (KR-20 \geq .78) değerlere sahip olması, grupların güvenilirlik ilkesinin sağlandığı yönünde ifade edilebilir.

BÖLÜM IV

BULGULAR

Bu bölümde sırasıyla DFA, Kovaryans matrisi eşitliği testi ve son olarak çok gruplu DFA ile elde edilen bulgular sunulacaktır.

4.1 1. Alt Probleme İlişkin Veri Setlerinden Elde Edilen DFA, Kovaryans Matrisleri Eşitliği Testi ve ÇDFA Sonuçları

a) 1. alt probleme ilişkin grup bazında DFA sonuçları

28 maddeli İngilizce dil formlarına ilişkin doğrulayıcı faktör analizine yönelik değerler Tablo 4.1.1’de verilmiştir.

Tablo 4.1.1. İngilizce Dil Formlarına İlişkin Doğrulayıcı Faktör Analizi

Grup	S-B χ^2 (sd)	p	χ^2 /sd	RMSEA	CFI	NNFI	SRMR
Amerika	580.67(350)	.000	1.66	.041	.96	.96	.050
Avustralya	1318.44(350)	.000	3.77	.052	.94	.93	.047
İngiltere	824.51(350)	.000	2.36	.040	.94	.93	.042
İrlanda	436.97(350)	.001	1.25	.026	.97	.97	.048
Kanada	1196.66(350)	.000	3.42	.047	.93	.93	.045
Singapur	458.54(350)	.000	1.31	.029	.98	.98	.047
Yeni Zel.	561.42(350)	.000	1.60	.045	.96	.95	.058

Tablo incelendiğinde, S-B χ^2 ile serbestlik derecesi oranlarının 5.0’ın altında, RMSEA değerlerinin .05 ve altında, CFI değerlerinin .90 üzerinde, NNFI değerlerinin .90 üzerinde ve SRMR değerlerinin de .08’in altında olduğu görülmektedir. Bu durumda, uyum indekslerinin 1. alt probleme ilişkin tüm gruplar için iyi uyum gösterdiği belirtilebilir. Özellikle Amerika, İrlanda, Singapur ve Yeni Zelanda gruplarında, S-B χ^2 ile serbestlik derecesi oranlarının 2.0’ın altında, RMSEA değerlerinin .05’in altında, CFI değerlerinin .95 ve üzerinde, NNFI değerlerinin .95 ve üzerinde olması mükemmel bir uyumun olduğunu ifade etmektedir.

Ölçeğin 28 maddeli yapısı DFA ile bütün gruplar için doğrulanmış ve aşağıda bir sonraki aşama olan kovaryans matrisleri eşitliği testi ve çok-gruplu DFA sonuçları ile ilgili bilgiler sunulmuştur.

b) 1. alt probleme yönelik kovaryans matrisleri eşitliği testi

İngilizce dil formlarına ilişkin kovaryans matrisleri eşitliği testine yönelik değerler Tablo 4.1.2’de verilmiştir.

Tablo 4.1.2. İngilizce Dil Formlarına İlişkin Kovaryans Matrislerinin Eşitliği

Grup	S-B χ^2 (sd)	p	χ^2 /sd	RMSEA	CFI	NNFI	SRMR
İngilizce	3429.85(2436)	.000	1.41	.026	.98	.98	.097

Tabloda görülmekte olduğu gibi, S-B χ^2 ile serbestlik derecesi oranının 2.0’in altında, RMSEA değerinin .05’in altında, CFI değerinin .95 üzerinde, NNFI değerinin .95 üzerinde olması kovaryans matrisleri arasında bir uyumun olduğu yönünde ifade edilebilir.

c) 1. alt probleme yönelik ÇDFA sonuçları

İngilizce dil formlarına ilişkin çok gruplu doğrulayıcı faktör analizine yönelik değerler Tablo 4.1.3’de verilmiştir.

Tablo 4.1.3. Amerika, Avustralya, İngiltere, İrlanda, Kanada, Singapur ve Yeni Zelanda’ya İlişkin Çok Gruplu Doğrulayıcı Faktör Analizi

	S-B χ^2 (sd)	χ^2 /sd	$\Delta\chi^2/\Delta$ sd	CFI	Δ CFI	SRMR	Δ SRMR	KARAR
Model1	6671.88(2786)	2.39	--	.93	--	.11	--	--
Model2	6183.25(2618)	2.36	.03	.93	.00	.069	.042	RET
Model3	5369.15(2450)	2.19	.17	.95	-.02	.050	.019	RET
Model4	5910.99(2618)	2.26	-.07	.94	.01	.078	-.028	KABUL

*p<.05

Model 1 (Yapısal Değişmezlik)

Model 2 (Zayıf Değişmezlik)

Model 3 (Güçlü Değişmezlik)

Model 4 (Katı Değişmezlik)

Ölçme modelleri yapısal eşitlik modeliyle test edilmeye başlanmış ve elde edilen uyum indeksleri ile grupların faktör yükleri, faktör korelasyonları ve hata varyansları arasında manidar bir fark olmadığı görülmektedir. S-B χ^2 ve serbestlik derecesi oranının 3’ün altında, CFI .90 üzerindedir. SRMR .08’in üzerinde olmasına rağmen, uyum indeksleri genel olarak değerlendirildiğinde modelin doğrulandığı kabul edilebilir.

Yapısal model ve zayıf değişmezlik modeli arasında anlamlı bir fark olup olmadığını görmek için Model 1 ve Model 2 karşılaştırılmış, S-B χ^2 ve serbestlik derecesi oranına göre modeller arasındaki uyumun iyileştiği ifade edilebilir. Bu

iyileşmenin manidar olup olmadığını değerlendirmek için \bar{T}_s değeri hesaplanmıştır. Bu değer 451.75 olup, χ^2 dağılımı tablosundaki kritik değerden büyüktür, $\chi^2_{fark}(168)=199.24$, $p<.05$. Bu durum yapısal değişmezlik modeli ile zayıf faktöriyel değişmezlik modeli arasında manidar bir farkın olduğunu göstermektedir. CFI değerlerinde bir değişim gözlenmemiş olup, bir iyileşme söz konusu olmamıştır. SRMR değerinde gözlenen .042 oranındaki iyileşme ise Chen (2007) tarafından Δ SRMR için önerilen .030 değerinin üzerinde olup, modeller arasında manidar bir farkın olduğu yönünde ifade edilebilir. Uyum indeksleri genel olarak değerlendirildiğinde, yapısal değişmezlik modeli ile zayıf faktöriyel değişmezlik modeli arasında manidar bir farkın olduğu, dolayısıyla zayıf değişmezlik modelinin kabul düzeyini karşılamadığı belirtilebilir.

Bir sonraki aşamada, zayıf faktöriyel değişmezlik ve güçlü faktöriyel değişmezlik modelleri karşılaştırılmış olup ve S-B χ^2 ve serbestlik derecesi oranında bir iyileşme olduğu görülmektedir. Bu değişimin manidar olup olmadığını tespit etmek için \bar{T}_s değeri hesaplanarak χ^2 dağılımı tablosundaki kritik değerden küçük olup olmadığı kontrol edilmiştir. \bar{T}_s değeri 694.73 olarak hesaplanmış ve χ^2 dağılımı tablosundaki kritik değerden büyük olduğu belirlenmiştir, $\chi^2_{fark}(168)=199.24$, $p<.05$. Bu durum zayıf değişmezlik ve güçlü değişmezlik modelleri arasında manidar bir farkın olduğunu göstermektedir. Modeller arasında farkın olmadığı yönünde kurulan hipotezi reddetmeden önce, diğer uyum indeksleri de incelenmiştir. CFI değerinde ise iyileşme olduğu gözlenmiş ve bu iyileşme kesme noktasının üzerinde, diğer bir ifadeyle $\geq .01$ olması nedeniyle modeller arasında fark olduğunu göstermektedir. SRMR değerindeki iyileşme de kabul düzeyini aşmakta olup $\geq .010$, uyum indekslerinden üçü de manidar bir farkın olduğunu göstermekte ve bu nedenle güçlü faktöriyel değişmezlik modelinin zayıf faktöriyel değişmezlik modelinden farklılaştığı yönünde karar alınarak H_0 hipotezi reddedilmiştir.

Son aşamada güçlü faktöriyel değişmezlik modelini simgeleyen Model 3 ve katı faktöriyel değişmezlik modelini simgeleyen Model 4 karşılaştırılmıştır. Katı değişmezlik modeline ait S-B χ^2 ve serbestlik derecesi oranında, CFI ve SRMR indekslerinde iyileşme olmadığı, aksine indekslerin kötüleştiği görülmüş ve bu doğrultuda iki model arasında manidar bir fark olmadığı tespit edilmiştir. Katı faktöriyel değişmezlik modelinin, güçlü faktöriyel değişmezlik modelinden manidar bir biçimde farklılaşmadığı tespit edilse de, güçlü faktöriyel değişmezlik modeli zayıf faktöriyel değişmezlik modelinden manidar bir biçimde farklılaştığı için, güçlü

faktöriyel değişmezlik modelinin reddedilmesi dolayısıyla, ölçme değişmezliğinin ikinci ve üçüncü model bağlamında sağlanamadığı ifade edilebilir.

Bütün bu aşamalar ve elde edilen uyum indeksleri doğrultusunda, ölçme aracının (PISA 2012 okuduğunu anlama becerisine yönelik hazırlanan 2 nolu kitapçık) İngilizce dil formlarının uygulandığı ülkeler (Amerika, Avustralya, İngiltere, İrlanda, Kanada, Singapur, Yeni Zelanda) arasında, faktör yapısı eşit olmamakta ve genel anlamda ölçme değişmezliği bulunmamaktadır.

4.2 2. Alt Probleme İlişkin Veri Setlerinden Elde Edilen DFA, Kovaryans Matrisleri Eşitliği Testi ve ÇDFA Sonuçları

a) 2. alt probleme ilişkin grup bazında DFA sonuçları

Fransızca dil formlarına ilişkin doğrulayıcı faktör analizine yönelik değerler Tablo 4.2.1’de verilmiştir.

Tablo 4.2.1. Fransızca Dil formlarına İlişkin Doğrulayıcı Faktör Analizi

Grup	S-B χ^2 (sd)	p	χ^2 /sd	RMSEA	CFI	NNFI	SRMR
Belçika	427.35(350)	.003	1.22	.036	.97	.96	.065
Fransa	442.43(350)	.001	1.26	.029	.98	.98	.050
İsviçre	499.36(350)	.000	1.43	.040	.96	.95	.053
Kanada	616.50(350)	.000	1.76	.046	.95	.95	.053
Lüksemburg	449.90(350)	.000	1.29	.063	.93	.93	.093

Tablo incelendiğinde, S-B χ^2 ile serbestlik derecesi oranlarının bütün gruplarda 2.0’ın altında, ilk dört grup için incelenen RMSEA değerlerinin .05’in altında, CFI değerlerinin .95 ve üzerinde, NNFI değerlerinin .95 ve üzerinde ve SRMR değerlerinin de .08’in altında olduğu görülmektedir. İndeksler genel olarak değerlendirildiğinde, söz konusu ülkelerde mükemmel uyumun varlığından söz edilebilir. Lüksemburg için ise RMSEA’nın .08’in altında, CFI ve NNFI’nın .90’nın üzerinde olması nedeniyle iyi bir uyumun olduğu belirtilebilir.

Ölçeğin 28 maddeli yapısı DFA ile bütün gruplar için doğrulanmış ve aşağıda bir sonraki aşama olan kovaryans matrisleri eşitliği testi ve çok-gruplu DFA sonuçları ile ilgili bilgiler sunulmuştur.

b) 2. alt probleme yönelik kovaryans matrisleri eşitliği testi

Fransızca dil formlarına ilişkin kovaryans matrisleri eşitliği testine yönelik değerler Tablo 4.2.2’de verilmiştir.

Tablo 4.2.2. Fransızca Dil Formlarına İlişkin Kovaryans Matrislerinin Eşitliği

Grup	S-B χ^2 (sd)	p	χ^2 /sd	RMSEA	CFI	NNFI	SRMR
Fransızca	1766.38(1624)	.007	1.09	.019	.99	.99	.15

Tabloda görülmekte olduğu gibi, S-B χ^2 ile serbestlik derecesi oranının 2.0’ın altında, RMSEA değerinin .05’in altında, CFI değerinin .95 üzerinde, NNFI değerinin .95 üzerinde olması Fransızca dil formlarında kovaryans matrisleri arasında bir uyumun olduğunu ifade edebilir.

c) 2. alt probleme yönelik ÇDFA sonuçları

Fransızca dil formlarına ilişkin çok gruplu doğrulayıcı faktör analizine yönelik değerler Tablo 4.2.3’de verilmiştir.

Tablo 4.2.3. Belçika, Fransa, İsviçre, Kanada ve Lüksemburg’a İlişkin Çok Gruplu Doğrulayıcı Faktör Analizi

	S-B χ^2 (sd)	χ^2 /sd	$\Delta \chi^2 / \Delta sd$	CFI	ΔCFI	SRMR	$\Delta SRMR$	KARAR
Model 1	2801.67(1974)	1.42	--	.95	--	.16	--	--
Model 2	2613.97(1862)	1.40	.02	.96	-.01	.10	.060	RET
Model 3	2435.44(1750)	1.39	.01	.96	.00	.093	.007	KABUL
Model 4	2601.27(1862)	1.40	-.01	.96	.00	.14	-.047	KABUL

*p<.05

Yapısal eşitlik modeliyle test edilmeye başlanan ölçme modellerinde, ilk olarak elde edilen uyum indeksleri ile grupların faktör yükleri, faktör korelasyonları ve hata varyansları arasında manidar bir fark olmadığı söylenebilir. S-B χ^2 ve serbestlik derecesi oranı 2’nin altında, CFI .90’nın üzerindedir. SRMR .08’in üzerinde olmasına rağmen, uyum indeksleri genel olarak değerlendirildiğinde modelin doğrulandığı kabul edilebilir.

Zayıf değişmezliğin test edildiği diğer aşamada, SB χ^2 /sd oranında iyileşme olması nedeniyle, iyileşmenin manidar olup olmadığı incelenmiştir. \bar{T}_s değeri 191.50 olarak hesaplanmış ve χ^2 dağılımı tablosundaki kritik değerden büyük olduğu belirlenmiştir, $\chi^2_{\text{fark}}(112)=137.70$, p<.05. Bu nedenle Model 1 ve 2 arasındaki SB χ^2 /sd farkının manidar olduğu söylenebilir. $\Delta SRMR \geq .030$ olması da modeller arasındaki farkın manidar olduğunu gösterirken, diğer taraftan incelen diğer uyum

indeksi ΔCFI önerilen kesme noktasına uygundur. Uyum indekslerinin çoğunluğu iki model arasında manidar bir fark olmadığı yönündeki hipotezi reddetmektedir.

Güçlü faktöriyel değişmezlik modeli aşamasında, $S_B\chi^2/sd$ farkı için hesaplanan \tilde{T}_s değeri 169.90 olarak hesaplanmış olup, χ^2 dağılımı tablosundaki kritik değerden büyük olduğu belirlenmiştir, $\chi^2_{fark}(112)=137.70$, $p<.05$. Bu durum modeller arasında manidar fark olduğunu gösterirken, diğer uyum indekslerinin tümü modeller arasında manidar bir farkın olmadığını göstermektedir.

Katı faktöriyel değişmezlik aşamasında, Model 4'ün uyum indekslerinde Model 3'e göre iyileşme gözlenmemiş olup bu doğrultuda, modeller arasında manidar bir fark bulunmamakla birlikte, katı faktöriyel değişmezliğin sağlandığı ifade edilebilir.

Bütün bu aşamalar sonucunda ve elde edilen uyum indeksleri doğrultusunda, ölçme aracının (PISA 2012 okuduğunu anlama becerisine yönelik hazırlanan 2 nolu kitapçık) Fransızca dil formlarının uygulandığı ülkeler (Belçika, Fransa, İsviçre, Kanada, Lüksemburg) arasında, zayıf faktöriyel model modeller arasında daha iyi çalışmakta ve modeller arasında fark olmadığı yönündeki hipotezi reddettiği için ölçme değişmezliğinin sağlanamadığı bulgusuna ulaşılmıştır.

4.3 3. Alt Probleme İlişkin Veri Setlerinden Elde Edilen DFA, Kovaryans Matrisleri Eşitliği Testi ve ÇDFA Sonuçları

a) 3. alt probleme ilişkin grup bazında DFA sonuçları

Farklı dil formlarına ilişkin doğrulayıcı faktör analizine yönelik değerler Tablo 4.3.1'de verilmiştir.

Tablo 4.3.1. Farklı Dil Formlarına İlişkin Doğrulayıcı Faktör Analizi

Grup	S-B χ^2 (sd)	p	χ^2 /sd	RMSEA	CFI	NNFI	SRMR
Türkiye	522.72(350)	.000	1.49	.038	.95	.95	.050
Amerika	580.67(350)	.000	1.66	.041	.96	.96	.050
Çin-Tayvan	615.77(350)	.000	1.76	.042	.95	.95	.050
İspanya	1041.00(350)	.000	2.97	.036	.96	.96	.034

Tabloda verilen bilgiler doğrultusunda S-B χ^2 ile serbestlik derecesi oranlarının bütün gruplarda 3.0'ın altında, RMSEA değerlerinin .05'in altında, CFI değerlerinin .95 ve üzerinde, NNFI değerlerinin .95 ve üzerinde ve SRMR

değerlerinin ise .05 ve altında olduğu görülmektedir. İndeksler genel olarak incelendiğinde, uyumun mükemmele yakın olduğu ifade edilebilir.

Ölçeğin 28 maddeli yapısı DFA ile bütün gruplar için doğrulanmış ve aşağıda bir sonraki aşama olan kovaryans matrisleri eşitliği testi ve çok-gruplu DFA sonuçları ile ilgili bilgiler sunulmuştur.

b) 3. alt probleme yönelik kovaryans matrisleri eşitliği testi

Farklı dil formlarına ilişkin kovaryans matrisleri eşitliği testine yönelik değerler Tablo 4.3.2’de verilmiştir.

Tablo 4.3.2. Farklı Dil Formlarına İlişkin Kovaryans Matrislerinin Eşitliği

Grup	S-B χ^2 (sd)	p	χ^2 /sd	RMSEA	CFI	NNFI	SRMR
T,A,Ç,I	2078.79(1218)	.000	1.71	.033	.97	.97	.065

T:Türkiye, A:Amerika, Ç:Çin-Tayvan, I:İspanya

Tablo incelendiğinde, S-B χ^2 ile serbestlik derecesi oranının 2.0’ın altında, RMSEA değerinin .05’in altında, CFI değerinin .95’in üzerinde, NNFI değerinin .95’in ve SRMR değerinin de .08’in altında olması Türkçe, İngilizce, Çince ve İspanyolca dil formlarının kovaryans matrisleri arasında bir uyumun olduğunu ifade edebilir.

c) 3. alt probleme yönelik ÇDFA sonuçları

Farklı dil formlarına ilişkin çok gruplu doğrulayıcı faktör analizine yönelik değerler Tablo 4.3.3’de verilmiştir.

Tablo 4.3.3. Amerika, Tayvan ve Türkiye’ye İlişkin Çok Gruplu Doğrulayıcı Faktör Analizi

	S-B χ^2 (sd)	χ^2 /sd	$\Delta \chi^2 / \Delta sd$	CFI	ΔCFI	SRMR	$\Delta SRMR$	KARAR
Model 1	3628.78(1568)	2.31	--	.94	--	.072	--	--
Model 2	3367.07(1484)	2.27	.04	.94	.00	.058	.014	KABUL
Model 3	2761.86(1400)	1.97	.30	.96	-.02	.050	.008	RET
Model 4	3022.39(1484)	2.04	-.07	.95	.01	.066	-.016	KABUL

*p<.05

Yapısal eşitlik modeliyle test edilmeye başlanan ölçme modelinde, elde edilen uyum indeksleri ile grupların faktör yükleri, faktör korelasyonları ve hata varyansları arasında manidar bir fark olmadığı görülmektedir. S-B χ^2 ve serbestlik derecesi oranının 3’ün altında, CFI ve NNFI .90 üzerinde, RMSEA .05’in altında, SRMR .08’in altındadır.

Yapısal model ve zayıf değişmezlik modeli arasındaki farkı görmek için Model 1 ve Model 2 karşılaştırıldığında, S-B χ^2 ve serbestlik derecesi oranına göre modeller arasındaki uyumun iyileştiği ifade edilebilir. Bu iyileşmenin manidar olup olmadığını değerlendirmek için \bar{T}_s değeri hesaplanmıştır. Bu değer 261.59 olup, χ^2 dağılımı tablosundaki kritik değerden büyüktür, $\chi^2_{\text{fark}}(84)=106.40$, $p<.05$. Bu durum yapısal değişmezlik modeli ile zayıf faktöriyel değişmezlik modeli arasında manidar bir farkın olduğunu göstermektedir. Fakat CFI değerinde bir değişim gözlenmemiş olup, bir iyileşme söz konusu olmamıştır. Aynı zamanda $\Delta\text{SRMR}\leq.030$ olması nedeniyle, söz konusu değişmezlik modelleri arasında farkın olmadığı söylenebilir ve zayıf faktöriyel değişmezlik modelinin doğrulandığı kabul edilebilir.

Güçlü faktöriyel değişmezlik modelinin ise zayıf faktöriyel değişmezlik modelinden manidar bir biçimde farklılaştığı ortaya koyulmuştur. Öncelikle S-B χ^2 ve serbestlik derecesi oranında gözlenen iyileşme için manidarlık test edilmiş olup, \bar{T}_s değeri 546.94 olarak hesaplanmıştır. Bu değer, χ^2 dağılımı tablosundaki kritik değerden büyüktür, $\chi^2_{\text{fark}}(84)=106.40$, $p<.05$. Bu durum iyileşmenin manidar olduğu göstermekte ve aynı zamanda ΔCFI değişimi de önerilen kesme noktasının üzerinde olup, indekslerden çoğunun manidar bir biçimde iyileşmesi nedeniyle, söz konusu modellerin birbirinden manidar bir biçimde farklılaştığı belirtilerek H_0 hipotezi reddedilebilir. Bu doğrultuda, gruplar arasında güçlü faktöriyel değişmezlik sağlanamamıştır.

Son aşamada güçlü faktöriyel değişmezlik modeli (Model 3) ve katı faktöriyel değişmezlik modeli (Model 4) karşılaştırılmıştır. Katı değişmezlik modeline ait S-B χ^2 ve serbestlik derecesi oranında, CFI ve SRMR indekslerinde iyileşme olmadığı, aksine indekslerin kötüleştiği görülmüş ve bu doğrultuda iki model arasında manidar bir fark olmadığı tespit edilmiştir. Katı faktöriyel değişmezlik modelinin, güçlü faktöriyel değişmezlik modelinden manidar bir biçimde farklılaşmadığı tespit edilse de, güçlü faktöriyel değişmezlik modeli zayıf faktöriyel değişmezlik modelinden manidar bir biçimde farklılaştığı için, güçlü faktöriyel değişmezlik modelinin reddedilmesi dolayısıyla, ölçme değişmezliğinin sağlandığı ifade edilemez.

Aşamalı olarak test edilen modeller doğrultusunda, elde edilen uyum indekslerine göre; ölçme aracının (PISA 2012 okuduğunu anlama becerisine yönelik hazırlanan 2 nolu kitapçık) farklı dil formlarının uygulandığı ülkeler (Türkiye,

Amerika, İspanya, Çin-Tayvan) arasında, üçüncü model daha iyi çalıştığı için ölçme değişmezliğinin sağlanamadığı kabul edilmiştir.

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde araştırmanın sonuçlarına ve bu doğrultuda çalışmanın sonucuna yönelik öneri ve görüşlere yer verilmekle birlikte, gelecek araştırmalara katkı sağlayacağı düşünülen önerilerde bulunulmuştur.

5.1 Sonuç ve Tartışma

PISA 2012 okuduğunu anlama testinin, kullanılan dil bazında ülkelere göre ölçme değişmezliğinin test edildiği bu çalışmada, çalışma kapsamına dâhil edilen grupların hiç birinde ölçme değişmezliğinin sağlanamadığı belirlenmiştir.

1. alt probleme ilişkin analiz sonucunda İngilizce formların kullanıldığı Amerika, Avustralya, İngiltere, İrlanda, Kanada, Singapur ve Yeni Zelanda ülkeleri sadece faktör yapıları bakımından değişmezliğe sahiptirler, diğer bir ifadeyle ölçme değişmezliği sağlanamamıştır. Ölçme değişmezliğinin sağlanamamış olması, çeşitli ülkelerde aynı dil hâkim olsa bile, ölçme değişmezliğinin çeşitli yanlışlıklar ya da kültürel farklılıklardan dolayı sağlanamayabileceğini göstermektedir. Ülkeler arasında yapılan sıralamada, Avustralya, İrlanda, Kanada, Singapur ve Yeni Zelanda OECD ortalamasının üzerinde, Amerika ve İngiltere ise OECD ortalamasına yakın puanlara sahiplerdir. Çalışma sonucu doğrultusunda bu sıralamaya güvenilemeyeceği belirtilirken daha önce PISA 2009 uygulamasına yönelik gerçekleştirilen benzer bir çalışmada, farklı sonuç alındığı görülmektedir. Asil ve Brown (2016) tarafından, PISA 2009 uygulamasında katılan ülkelerin, okuduğunu anlama testinden elde edilen sonuçların ne derece birbiriyle karşılaştırılabilir olduğunu incelemek amacıyla, Avustralya örneklemini temel alarak ülkeler arası değişmezliği incelenmiştir. Avustralya ile karşılaştırılan 3 ülke (Yeni Zelanda, Kanada ve Amerika) arasında güçlü faktöriyel değişmezlik düzeyinde ölçme değişmezliğinin sağlandığı bulgusuna ulaşılmıştır. Bu bulgular doğrultusunda, PISA ile belirlenen performans puanlarının, her uygulama sonrasında incelenmesinde yarar vardır.

2. alt probleme yönelik Fransızca formların kullanıldığı Belçika, Fransa, İsviçre, Kanada ve Lüksemburg arasında bir karşılaştırma yapılırken, ikinci model daha iyi çalıştığı için ölçme değişmezliği sağlanamamış olup, değişken

ortalamalarındaki farklılıkların ölçüğün özelliklerinden kaynaklanabileceği sonucuna ulaşılmaktadır. Belçika, Fransa, İsviçre, Kanada uygulama sonucuna göre üst performans düzeyinde yer alırken, Lüksemburg alt performans düzeyinde yer almaktadır ve yapılan çalışma sonucunda bu farklılığın öğrenci performansından kaynaklanamayabileceği ifade edilebilir. Bu gruplarda uygulanan ölçme araçlarının faktör yapılarının eşit olmaması nedeniyle, çeşitli yanlılıkların söz konusu olabileceği ve bu gruplar arasında yapılan karşılaştırmaların ve alınan puanların geçerli olmadığı belirtilebilir.

3. alt probleme yönelik farklı dil ve farklı kültür kapsamında belirlenen Amerika, İspanya, Tayvan ve Türkiye ülkeleri için de ölçme değişmezliğinin sadece zayıf faktöriyel değişmezlik düzeyinde sağlanması, diğer bir ifadeyle grupların faktör yapıları ve faktör yükleri açısından benzer olması ölçme değişmezliğinin sağlanmasına yetmemektedir. Bu durum, üst performans düzeyinde yer alan Tayvan, OECD ortalamasında bulunan Amerika ve alt performans düzeyinde yer alan İspanya ve Türkiye'nin performanslarının doğrudan karşılaştırılmayacağı anlamına gelmektedir. Diğer bir ifadeyle; edinilen bulgular doğrultusunda, bir ülkenin diğerinden daha iyi ya da daha kötü olduğunu söylemek doğru olmayacaktır. Yapılan çalışma sonucunda bu sıralamanın öğrenci performanslarından kaynaklanamayabileceği ifade edilebilir. Bu sonuç, alanyazında bulunan farklı kültürlerle yönelik yapılan diğer ölçme değişmezliği çalışmalarıyla paralellik göstermektedir. Kıbrıslıoğlu (2015)'nin PISA 2012 matematik öğrenme modelini farklı dil ve kültüre sahip ülkeler arasında karşılaştırdığı çalışmada da, Çin-Şangay (üst performans düzeyinde), Türkiye (alt performans düzeyinde) ve Endonezya (alt performans düzeyinde) örneklemeleri arasında ölçme değişmezliğinin sağlanamadığı bulgusuna ulaşmıştır. Bu doğrultuda, uyarlama çalışmaları her ne kadar dikkatle yapılsa da, farklı kültüre uyarlanan ölçeklerin ölçme değişmezliğine sahip olması zor olabilmektedir. Benzer olarak, Wu ve diğerleri (2007)'nin inceledikleri TIMSS 1999 uygulamasının matematik testlerinin, farklı kültürlerde yapısal (Yeni Zelanda-Japonya, Amerika- Japonya) ve zayıf faktöriyel değişmezliğin (Avustralya-Japonya, Avustralya-Amerika, vb.) sağlanması, benzer kültürlerde (Kanada-Amerika, Avustralya-Yeni Zelanda vb.) ise katı değişmezliğin sağlanması, farklı kültürleri kapsayan uygulama sonuçlarının incelenmeden önce kabul edilmemesi gerektiğini desteklemektedir. Alanyazın incelendiğinde, uluslararası uygulamalara yönelik farklı kültürler arasında ölçme değişmezliği koşulunun sağlanamadığı çalışmalar, bu

koşulu sağlayan çalışmalardan sayıca üstündür. Öğretmen (2006)'nın, PIRLS 2001 verileriyle gerçekleştirdiği çalışmasında, Amerika ve Türkiye arasında ölçme değişmezliğinin sağlanmadığı belirtilmiştir. Ardıç ve Güzeller (2015) tarafından yapılan araştırma ise PISA 2012 uygulamasına ait bilgi ve teknoloji kullanımına yönelik olan ölçek maddelerini kapsamakta olup Türkiye, Yeni Zelanda ve Singapur örneklemi arasındaki ölçme değişmezliği aşamalı olarak analiz edilerek analiz sonucunda ülkeler arası sadece yapısal değişmezlik koşulunun sağlandığı belirtilmiştir. Örnek olarak verilen bu çalışmalar, kültürlerarası ölçme değişmezliğinin sağlanmasının çok fazla mümkün olmadığını göstermektedir. Bir ölçme aracı, farklı bir kültüre uyarlandığında, yapı eşdeğerliliğini kaybetme riski artmaktadır. Diğer taraftan, az sayıda da olsa alanyazında kültürlerarası karşılaştırmalarda ölçme değişmezliğinin sağlandığını ifade eden çalışmalar bulunmaktadır. Ulutaş ve Çıkrıkçı (2015) tarafından gerçekleştirilen çalışma, PISA 2006 Fen okuryazarlığı testinin Türkiye ve Amerika örneklemelerinin oluşturduğu gruplar arasında ölçme değişmezliğinin sağlandığını ifade etmektedir. Akyıldız (2009) PIRLS verilerini kullanarak gerçekleştirdiği çalışmasında, 35 ülkenin (alt problemler doğrultusunda gruplara ayrılmış) faktör yapısının orta düzeyde eş değer olduğu bulgusuna ulaşmıştır. Önemli olan, ilk önce testlerin psikometrik niteliklerini belirlemek ve sonrasında yapılan değerlendirmelere güvenebilmektir.

Bu çalışma kapsamında, ölçme değişmezliğinin sağlanamıyor olduğu bulgusu, gruplar arasında, örtük değişkenle gözlenen değişken arasındaki ilişkinin değişmez olarak tanımlanamayacağını ve bu durumda gözlenen puanlardaki farklılığın gruplar arasında var olan gerçek performans farklılığından kaynaklanmadığını ifade eder. Ayrıca bu bulgu, ölçek formlarının aynı niteliklere sahip olmadığını ve bu nedenle farklı nitelikleri taşıyan formlar aracılığıyla, gruplar arasında yansız bir karşılaştırma yapılamayacağını ve yapılan karşılaştırmaların ve sıralamaların da geçerliliğinin olmadığını belirtir. Bu çalışmanın alt problemlerine yönelik, değişmezliği etkileyen yanlılıkların, yapı yanlılığı ve yöntem yanlılığı ve madde yanlılığı olduğu söylenebilir. Uygulamada yapı yanlılığına neden olan ifadeler bulunuyor olabilir, farklı kültürlerde, benzer şeyler farklı biçimde ifade edilebileceğinden, o ifadeleri algılamak bazı öğrenciler için daha kolay bazıları için daha zor olabilir. Uyarlama sürecinde dilsel açıdan sistematik bir yol izlenmemesi yapı yanlılığına sebep olabilir. Ayrıca, Malda, Van de Vijver, Srinivasan, Transler, Sukumar ve Rao (2008)'nin üzerinde durduğu uyarlama gerektirecek kapsamlara

dikkat edilmemiş olabilir. Kültürel normların ve psikolojik kavramların anlamlı çevirisi, aşına olma durumunu dikkate alma, kelimelerin doğru karşılığının kullanılması ve tam karşılığı olmayan kelime ya da ifadeleri çevirirken benzer uzunlukta kalmalarına dikkat edilmesi gibi önemli noktalar alanyazında vurgulanmış olup, bu tür uyarılma hataları bu çalışmada ölçme değişmezliğin koşulunun sağlanmasını olumsuz yönde etkileyen faktörler arasında bulunabilir. Yöntem yanlılığına ilişkin neden ise kullanılan ölçme aracında yer alan yönergelerde kullanılan dilin ya da soru tiplerinin alışılmışın dışında olması olabilir. Bunların dışında, bazı kültürlerin lehine ya da aleyhine çalışan test maddeleri de olabilir. Bu tür yanlılığa daha çok kültürel ve coğrafik özellikler neden olduğu için, madde uyarılma ya da yazma sürecinde kültürel farklılıklar göz önünde bulundurulamamış olabilir.

5.2 Öneriler

Uygulama sonucuna yönelik ve ileriye ileriye yönelik olmak üzere oluşturulan öneriler aşağıda sıralanmıştır.

- Bu uygulama sonucuna göre, söz konusu gruplar arasındaki performansa yönelik olarak yapılan sıralamalara güvenilmemelidir.
- Aynı dili konuşan ülkeler bile kültürel farklılıklar taşıyabileceklerinden, çeşitli yanlılıklar söz konusu olabileceği için, mutlaka ölçme değişmezliği incelenmelidir. Öğrencilerin aynı dili konuşuyor olmaları, aynı dilde aynı sorulara tâbi tutulabilecekleri anlamına gelmemelidir. Aynı dilde formlar uygulanacaksa bile, sorular üzerinde kültüre göre gerekli değişiklikler yapılarak uygulanmalıdır. Anlam karışıklığına neden olacak ifadeler yer verilmemeli, çeviri işleminin en güvenilir biçimde yapılabilmesi için iki yönlü çeviri tercih edilmeli ve hiçbir yanlılığa neden olmayacak bir biçimde diğer ölçek uyarılma adımları özenle takip edilmelidir. Aynı dilin kullanıldığı ülke formlarında bile kültür farklılığının etkisiyle yapısal eşitliğin sağlanması zor olduğundan, karşılaştırmalı kültür uygulamaları sonuçlarına şüpheyle yaklaşılmalı ve sonuçları ancak gerekli değişmezlik ve yanlılık çalışmaları yapıldıktan sonra dikkate alınmalıdır.
- Söz konusu gruplarda ölçme değişmezliğinin sağlanamama nedenleri araştırılarak, asıl yanlılıkların neden kaynaklandığı tespit edilmeye

çalışılabilir. Yanlı olduğu düşünölen maddeler, ilerleyen yıllarda kullanılmak üzere PISA uygulaması madde havuzunda bulunabilir, bu doğrultuda gelecek çalışmalara o maddelerin dâhil edilmemesi ya da düzeltilmesi ile ilgili olarak kurumla iletişime geçilebilir. Böylece ilerleyen yıllarda, ölçme uygulamalarının daha güvenilir ve geçerli olabilmesi için, madde havuzunda birikecek olan soruların kalitesini artırmaya katkı sağlanabilir.

- Bu çalışma, 2012 PISA uygulamasının 2 numaralı kitapçığındaki alt ölçęi olan okuma becerileriyle sınırlı olması nedeniyle, farklı araştırmacılar tarafından farklı kitapçıkların, farklı alt ölçeklerin ya da farklı yıllardaki PISA uygulamalarının analizi yapılabilir.
- Farklı yıllardaki PISA okduğunu anlama becerisi testi incelenerek, PISA uygulamalarında ölçme deęişmezlięinin sağlanıp sağlanamadığına yönelik ek kanıt sunulabilir.
- Bu çalışmaya dâhil edilen grupların ölçme deęişmezlięi farklı bir teknik kullanılarak incelenebilir ve sonuçlar karşılaştırılabilir.
- Analiz, farklı ölkeler grupları ya da daha farklı deęişkenler belirlenerek yürütölebilir. Özellikle, aynı dili konuşan ve aynı kültür özelliklerini paylaşan gruplarda ölçme deęişmezlięi incelenerek, sonuçlar bu çalışmanın farklı kültöre sahip olan örneklem sonuçlarıyla karşılaştırılabilir.
- Analize dâhil edilen deęişkenler için ölçme deęişmezlięinin sağlanamadığı bulgusuna ulaşıldığı için, sonraki çalışmalarda kısmi ölçme deęişmezlięi test edilebilir.
- Ölçek uyarlama çalışmaları kültürlerin ve bireylerin özelliklerini dikkate alarak yapılmalıdır. Testler uygulanmadan önce uyarlama çalışmaları sonrasında, çeşitli pilot çalışmalarla yanlılıklardan arınık olup olmadığı da test edilmelidir. Aksi takdirde, performansa dayalı yapılan deęerlendirmeler hatalı ya da yanlı olabileceęi gibi, bireylerin ve toplumların yanlış yönlendirilmesine de neden olabilir.
- Alanyazında ölçme deęişmezlięi çalışmalarının kapsamının büyük çoğunluęunu ölçek verileri oluşturmakta ve bu nedenle çalışmalarda daha fazla başarı testinin ve performansı belirlemeye yönelik maddelerin

değerlendirilmesine yer verilmesi, geçerli ve güvenilir testler hazırlanmasını artırmak, yanlış algıları ortadan kaldırmak ve eğitim sistemlerin düzenlenmesine katkıda bulunmak açısından önerilmektedir.

- Yapılan bu çalışma, 2012 PISA okuduğunu anlama ölçeğine yönelik geçerlilik kanıtları sunuyor olsa da, aynı zamanda kullanılan yöntem ve izlenen yol açısından ölçme değişmezliği çalışmaları yapmak isteyen araştırmacılara gerekli bilgiyi sağlayabilir.

KAYNAKÇA

- Abbott, M. L. (2011). *Understanding educational statistics using Microsoft Excel and SPSS*. United States: John Wiley & Sons, Inc.
- Akyıldız, M. (2009). *PIRLS 2001 testinin ülkelerarası yapı geçerliliğinin incelenmesi*. (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara.
- Akyol, H. (2005). *Türkçe ilk okuma yazma öğretimi*. Ankara: Pegem A Yayıncılık.
- Alivernini, F. (2011). Measurement invariance of a reading literacy scale in the Italian Context: a psychometric analysis. *Procedia Social and Behavioral Sciences*, 15, 436-441.
- Ardıç, E. Ö. and Güzeller, C. O. (2015). *Crosscultural equivalence of PISA 2012 student questionnaire items on the use of information and communication Technologies*. Multidisciplinary Academic Conference on Education, Teaching and Learning, Prague.
- Arffman, I. (2010). Equivalence of translations in international reading literacy studies. *Scandinavian Journal of Educational Research*, 54(1), 37-59.
- Arıkan, S. (2010). *Construct validity and factor structure of student selection examination across subgroups* (Yayınlanmamış Doktora Tezi). Orta Doğu Teknik Üniversitesi, Ankara.
- Arim, R. G. and Ercikan, K. (2014). Comparability between the American and Turkish versions of the TIMSS mathematics test results. *Eğitim ve Bilim*, 39 (172), 33-48.
- Asil, M. and Brown, G. T. L. (2016) Comparing OECD PISA reading in English to other languages: Identifying potential sources of non-Invariance. *International Journal of Testing*, 16 (1), 71-93.
- Asil, M. ve Gelbal, S. (2012). PISA öğrenci anketinin kültürler arası eşdeğerliği. *Eğitim ve Bilim*, 37 (166), 236-249.
- Ayyıldız, H. ve Cengiz, E.(2006). Pazarlama modellerinin testinde kullanılabilir Yapısal Eşitlik Modeli (YEM) üzerine kavramsal bir inceleme. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi*, 11 (1), 63-84
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press
- Bahadır, E. (2012). *Uluslararası Öğrenci Değerlendirme Programı'na (PISA 2009) göre Türkiye'deki öğrencilerin okuma becerilerini etkileyen değişkenlerin bölgelere göre incelenmesi* (Yayınlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*. Chichester, United Kingdom: John Wiley & Sons. Inc. (ISBN: 978-0-471-93094-5)

- Baş, Ö. (2013). The effect of texts about Turkish and American cultures on reading comprehension level of primary school students. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education)* 28(3), 30-40.
- Başusta, N. B. (2010). Ölçme eşdeğerliği. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 58-64.
- Başusta, N. B. ve Gelbal, S. (2015), Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(4), 80-90.
- Berry, J. W. (1969). On cross-cultural comparability. *International Journal of Psychology*, 4, 119-128.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71, 425–440.
- Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, 16, 201–213.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Publications, Inc.
- Browne, M. W. and Cudeck, R. (1993). Alternative ways of assessing model fit. Kenneth A. Bollen and J. Scott Long (Ed.). *Testing Structural Equation Models* (p. 136-162). Newbury Park, CA: Sage
- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum.
- Byrne, B. M. and Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-cultural Psychology*, 30, 557–576.
- Byrne, B. M., Raju, N. S. and Laffitte, L. J. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 28 (3), 517-529.
- Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, 20 (4), 872-882.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Newyork: Taylor & Francis Group
- Cangur, S. and Ercan, I. (2015). Comparison of Model Fit Indices Used in Structural Equation Modeling Under Multivariate Normality. *Journal of Modern Applied Statistical Methods*, 14 (1), Article 14.

- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504.
- Chen, J. (2018). KR-20. Bruce B. Frey (Ed.). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (p. 932-936). Sage Publications, Inc.
- Cheung, G. W. and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255.
- Crack, T. F. (2018). A Note on Karl Pearson’s 1900 Chi-Squared Test: Two Derivations of the Asymptotic Distribution, and Uses in Goodness of Fit and Contingency Tests of Independence, and a Comparison with the Exact Sample Variance Chi-Square Result. SSRN Electronic Journal. (SSRN: <http://dx.doi.org/10.2139/ssrn.3284255>)
- Cronbach, L. J. (1990). *Essentials of psychological testing (5. Edition)*. New York: Harper&Row Publishers.
- Çetin, B. (2010). Cross-cultural structural parameter invariance on PISA 2006 student questionnaires. *Eğitim Araştırmaları-Eurasian Journal of Educational Research*, 38, 71-89.
- Çıkrıkçı Demirtaşlı, N. ve Uluştas, S. (2015). A Study on detecting of Differential Item Functioning of PISA 2006 Science Literacy items in Turkish and American samples. *Eurasian Journal of Educational Research*, 58, 41-60.
- Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları*. Ankara: Pegem
- Çukur, C. Ş. (2007). Kültürel süreçleri nitelenmek, ölçmek ve ilişkilendirmek: Karşılaştırmalı kültürel yaklaşımlar, Erdem R. ve Çukur C. Ş. (Ed.), *Kültürel bağlamda yönetsel-örgütsel davranış* (s. 33-46). Ankara: Türk Psikologlar Derneği Yayınları.
- Deniz, K. Z. (2007). Psikolojik ölçme aracı uyarlama. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 40 (1), 1-16.
- Dinçer, M. A. ve Kolaşın, G. U. (2009). Türkiye’de öğrenci başarısızlığında eşitsizliğin belirleyicileri. *Eğitim Reformu Girişimi*. İstanbul: Sabancı Üniversitesi.
- Elosua O. P. and Mujika L. J. (2013). Invariance levels across language versions of the PISA 2009 reading comprehension tests in Spain. *Psicothema*, 25 (3), 390–395.
- Engin, İ. (1990). Kültür- kişilik ilişkisi. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Dergisi*, 1 (2), 171-176.
- Ercan, İ. ve Kan, İ. (2004). Ölçeklerde güvenirlik ve geçerlik. *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 30 (3), 211-216
- Ercikan, K. and Koh, K. (2009). Examining the construct comparability of the English and French versions of TIMSS. *International Journal of Testing*, 5 (1), 23-35.

- Evran, D. (2012). *Assessing measurement invariance: Multiple group confirmatory factor analysis for differential item functioning detection in polytomous measures of Turkish and American students* (Yayınlanmamış Yüksek Lisans Tezi). University of Florida, Florida.
- Field, A. (2009) *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications Ltd.
- Gere, J. and Macdonald, G. (2012). Assessing relationship quality across cultures: an examination of measurement equivalence. *Personal Relationships*, 20, 422-442.
- Gravetter, F. and Wallnau, L. (2014). *Essentials of statistics for the behavioral sciences* (8th ed.). Belmont, CA: Wadsworth.
- Grisay, A. and Monseur, C. (2007). Measuring equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation* 33, 69-86.
- Güzeller, C. (2011). PISA 2009 öğrenci anketinde yer alan bilgisayar tutum boyutunun kültürlerarası eşitliğinin incelenmesi. *Eğitim ve Bilim*, 36 (162), 320-327.
- Hair, J. F., Anderson, R. E., Tatham, R., L. and Black, W. C. (1998). *Multivariate data analysis (Fifth edition)*. United States: Prentice-Hall, Inc.
- Hambleton, R. K. and Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology*, 1 (1), 1-13.
- Hambleton, R. K. and De Jong, J.H.A.L. (2003). Advances in translating and adapting educational and psychological tests. *Language Testing*, 20 (2), 127-134.
- Harrington, D. (2009). *Confirmatory factor analysis*. New York: Oxford University Press.
- He, j. and Van de Vijver, F. (2012). Bias and equivalence in crosscultural research. *Online Readings in Psychology and Culture*, 2 (2).
- Hooper, D., Coughlan, J. and Mullen, M.(2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6 (1), 53-60.
- Horn, J. L. (1991). Comments on 'Issues in Factorial Invariance. Linda M. Collins and John L. Horn (Eds.) *Best Methods for the Analysis of Change* (p.114-125). Washington, DC: American Psychological Association.
- Horn, J. L. and McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117-144.
- Hox, J. J. and Bechger, T. M. (2007). An introduction to Structural Equation Modelling. *Family Science Review*, 11, 354-373.
- Hoyle, R. H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. R. H. Hoyle (Ed.). In *Structural equation modeling: Concepts, issues, and applications* (p. 1-15). Thousand Oaks, CA: Sage Publications, Inc.

- Hu, L. T. and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6 (1), 1-55.
- Hulin, C. and Mayer, L. (1986). Psychometric equivalence of a translation of the job descriptive index into hebrew. *Journal of Applied Psychology* 71 (1), 83-94.
- Husen, T. and Tuijnman, A.C. (1994). Monitoring standards in education: Why and how it came about. In. A.C. Tuijnman, T.N. Postlethwaite (Ed.), *Monitoring standards of education* (p. 1-21). Oxford: Pergamo.
- IEA (2003). PIRLS 2001 International Report. International Study Center, Lynch School of Education, Boston College.
- Jöreskog, K. and Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software International Inc.
- Kankaraš, M. (2010). *Essays on measurement equivalence in cross-cultural survey research: A latent class approach* (Doctoral Thesis). Tilburg University, Tilburg.
- Kankaraš, M. and Moors, G. (2010). Researching measurement equivalence in cross-cultural studies. *Psihologija*, 43 (3), 121-136.
- Keeves, J. P. (2011). From the Beginning in 1958 to 1990. C. Papanastasiou, T. Plomp, E. Papanastasiou, E. IEA 195802088: *50 Years of Experiences and Memories* (p. 1-20). Nicosia (Cyprus): Cultural Centre of the Kykkos Monastery.
- Kehoe, J. (1995). Basic item analysis for multiple-choice tests. *Practical Assessment, Research & Evaluation*, 4 (10), 1-3.
- Kıbrıslıoğlu, N. (2015). *PISA 2012 Matematik Öğrenme Modelinin Kültürlere ve Cinsiyete Göre Ölçme Değişmezliğinin İncelenmesi: Türkiye, Çin (Şangay), Endonezya Örneği* (Yayınlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. New York: Methuen & Co. LTD.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. 3rd edition. New York: The Guilford Press.
- Lissitz, R. W. (2009). The concept of validity: Revisions, new directions, and applications, book review. *Journal of Educational Measurement*, 48(4), 457-461.
- Little, T. D and Slegers, D.W. (2005). Factor analysis: Multiple groups. *Encyclopedia of Statistics in Behavioral Science*, 2, 617-623.
- Little, T. D. (2013). *Longitudinal Structural Equation Modeling*. New York: Guilford Press.
- Malda, M., Van de Vijver, F.J.R., Srinivasan, K., Transler, C., Sukumar, P. and Rao, K. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly*, 50, 451-468.

- Malda, M., Van de Vijver, F. J. R. and Temane, Q. M. (2010). Rugby versus soccer in South Africa: Content familiarity contributes to cross-cultural differences in cognitive test scores. *Intelligence*, 38(6), 582-595.
- Marsh, H. W. (1994). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling*, 1, 5-34.
- MEB (2003). TIMSS 2009 Ulusal Rapor. Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı, Ankara: Earged
- MEB (2005). PISA 2003 Projesi Ulusal Nihai Rapor. Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı, Ankara: Earged
- MEB. (2009). İlköğretim Matematik Dersi 6-8 Sınıflar Öğretim Programı. Milli Eğitim Bakanlığı Talim Terbiye Kurulu Başkanlığı, Ankara.
- MEB (2010). PISA 2006 Ulusal Nihai Rapor. Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı, Ankara: Earged
- MEB (2011). PISA Türkiye. Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü, Ankara: Eğitek.
- MEB (2013). PISA 2012 Ulusal Ön Raporu. Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü, Ankara: Yeğitek
- MEB (2015). PISA 2012 Araştırması Ulusal Nihai Raporu. Millî Eğitim Bakanlığı, Ölçme Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü, Ankara: İşkur Matbaacılık.
- Meredith, W. and Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, 57(2), 289-311.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58,525-543.
- Mertler, C. A. and Vannatta, R.A. (2005) *Advanced and Multivariate Statistical Methods: Practical Application and Interpretation (3rd Edition)*. Pyrczak, Los Angeles: Routledge
- Milfont, T. L. and Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3 (1), 111-121.
- Millsap, R. E. and Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, 9 (1), 93-115.
- Mondal, S., Das, S., Musunuru, K. and Dash, M. (2017). Study on the factors affecting customer purchase activity in retail stores by confirmatory factor analysis. *Revista Espacios*, 38 (11),1-25.

- OECD (2012) PISA 2009 Technical Report, PISA, OECD Publishing.
- OECD (2013a) PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy. Paris: OECD Publications.
- OECD (2013b) PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed (Volume II), PISA, OECD Publishing.
- OECD (2014) PISA 2012 Technical Report, PISA, OECD Publishing.
- Öğretmen, T. (2006). *Uluslar Arası Okuma Becerilerinde Gelişim Projesi (PIRLS) Testinin Psikometrik Özelliklerinin İncelenmesi: Türkiye-Amerika Birleşik Devletleri Örneği* (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Önen, E. (2009). *Ölçme değişmezliğinin yapısal eşitlik modelleme teknikleriyle incelenmesi* (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara.
- Örs, S. (2010). *İlköğretim 6.,7. Ve 8. Sınıf seviye belirleme sınavı fen ve teknoloji alt testlerinin faktör yapılarının belirlenmesi* (Yayınlanmamış Yüksek Lisans Tezi). Ankara Üniversitesi, Ankara.
- Özdemir, O. (2009). Örneklem ve randomizasyon. *IKU Dergisi*, 23,32-41
- Özmuş, M. ve Kaya, A. (2014). Türkiye'nin PISA 2009 ve 2012 sonuçlarına ilişkin karşılaştırmalı bir analiz. *Journal of European Education*, 4 (1), 23-40.
- Reise, S. P., Widaman, K. F. and Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552-566.
- Rensvold, R. B. and Cheung, G. W. (2001). Testing for metric invariance using structural equation models: Solving the standardization problem. C. A. Schriesheim & L. L. Neider (Eds.), *Research in management: Equivalence in measurement* (p. 25-50). Greenwich, CT: Information Age Publishing.
- Rhodes Purdy, M. (2017). *Regime support beyond the balance sheet: participation and policy performance in Latin America*. New York: Cambridge University Press.
- Sargut, S. (2001). *Kültürlerarası Farklılaşma ve Yönetim* (2. Baskı). Ankara: İmge Yayıncılık.
- Sarı, E. (2004). Kültürlerarası iletişim: Temeller, gelişmeler, yaklaşımlar. *Folklor/edebiyat*, 39, 1-31.
- Satorra, A. and Bentler, P.M. (2001). A scaled difference chisquare test statistic for moment structure analysis. *Psychometrika*, 66, 507-514.
- Schermelleh-Engel K., Moosbrugger, H. and Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures, *Methods of Psychological Research Online*, 8 (2), 23-74.

- Schnabel, D. B. L., Kelava, A., Van de Vijver, F. J. R. and Seifert, L. (2015). Examining psychometric properties, measurement invariance, and construct validity of a short version of the Test to Measure Intercultural Competence (TMIC-S) in Germany and Brazil. *International Journal of Intercultural Relations*, 49, 137-155.
- Serpell, R. (1979). How specific are perceptual skills? *British Journal of Psychology*, 70, 365-380.
- Shivraj, P. (2014). *Validating the inferences made from the 2012 mathematics PISA* (Doctoral Thesis), Southern Methodist University, Dallas, Texas.
- Shulz, W. (2005). Testing parameter invariance for questionnaire indices using confirmatory factor analysis and Item Response Theory. Paper presented at the annual meetings of *American Educational Research Association*, San Francisco.
- Somer, O., Kormaz, M., Dural, S. ve Can, S. (2009). Ölçme eşdeğerliğinin yapısal eşitlik modellemesi ve madde cevap kuramı kapsamında incelenmesi. *Türk Psikoloji Dergisi*, 24 (64), 61-75.
- Sözer, E., Sarıtaş, S. Çavdar, D., Çorbacı, E. C. ve Bağcı, V. (2017). An Investigation of Measurement Invariance of Interest in Science Model Across Different Groups: A Study on PISA Student Questionnaire. Vopava, J., Douda V., Kratochvil, R., Konecki, M. (Ed.), *Proceedings of MAC 2017*, International Conference, (113-122). Czech Technical University: Prague.
- Steenkamp, J. B. E. M. and Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research . *Journal of Consumer Research*, 25, (1), p. 78-107.
- Sticker, L. J and Rock, D.A. (2008). Factor structure of the TOEFL internet-based test across subgroups. *TOEFL iBT Research Report*, ETS, Princeton, NJ.
- Şekercioğlu, G. (2009). *Çocuklar için benlik algısı profilinin uyarlanması ve faktör yapısının farklı değişkenlere göre eşitliğinin test edilmesi* (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara.
- Şekercioğlu, G. (2018). Measurement Invariance: Concept and Implementation. *International Online Journal of Education and Teaching (IOJET)*, 5, (3). 609-634.
- Tabachnick, B. G. and Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). United States of America: Pearson Education Inc.
- Trochim, W. M. and Donnelly, J. P. (2006). *The research methods knowledge base* (3rd ed.). Cincinnati, OH: Atomic Dog.
- Tucker, L. R. and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
- 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.

- 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 2010*, 35(155), 26-35.
- 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 Schoot, R., Lugtig, P. and Hox, J. (2012) A checklist for testing measurement invariance, *European Journal of Developmental Psychology*, 9, (4), 486-492.
- Van de Vijver, F. J. R. and Hambleton, R.K. (1996). Translating tests: some practical guidelines. *European Psychologist* 1, 89-99.
- Van de Vijver, F. J. R. and Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park CA: Sage Publications.
- Van de Vijver, F. J. R. and Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 54, 119-135.
- Wegener, D. T. and Fabrigar, L. R. (2000). Analysis and Design for Nonexperimental Data: Addressing Causal and Noncausal Hypothesis. H. T. Reis, C. M. Judd (Ed.). *Handbook of research methods in social and personality psychology* (p.412-450). Cambridge: Cambridge University Press.
- Wheaton, B., Muthen, B., Alwin, D., F., and Summers, G. (1977). Assessing reliability and stability in panel models. *Sociological Methodology*, 8 (1), 84-136.
- Wicherts, J.M. and Dolan, C.V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice*, 29 (3), 39-47.
- Wu, A. D., Li, Z. and Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, 12, 1-26.
- Yaşar, M. (2014). İstatistiğe yönelik tutum ölçeği: Geçerlilik ve güvenirlik çalışması. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 36, 59-75.
- Yıldırım, H. H. (2006). *Differential item functioning analysis of mathematics items in the international assessment programs* (Yayınlanmamış Doktora Tezi). Orta Doğu Teknik Üniversitesi, Ankara.

Yılmaz, M. (2008). Türkçe'de okuduğunu anlama becerilerini geliştirme yolları. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 5 (9), 131-139.

Yu, C.Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Doctoral Thesis). University of California, California.

EKLER

Ek 1: Doğrulayıcı Faktör Analizi Örnek Lisrel Sözdizimi

```
TURKIYE Olcegi Dogrulayici Faktor Analizi
Observed Variables
V1-V28
Covariance Matrix from File TURK.COV
Asymptotic Covariance Matrix from File TURK.ACM
Sample Size: 349
Latent Variable: F1
Relationships:
V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
V20 V21 V22 V23 V24 V25 V26 V27 V28= F1
Path Diagram
End of Problem
```

Ek 2: Kovaryans Matrisleri Eşitliği Testi Örnek Lisrel Sözdizimi

```
INGILIZCE GRUBU KOVARYANS MATRISLERININ ESITLIGI TESTI
Group ENG:
Observed Variables: V1-V28
Covariance Matrix from File ENG.COV
Asymptotic Covariance Matrix from File ENG.ACM
Method of Estimation = Maximum Likelihood
Iterations: 500
Sample Size: 836
Latent Variables: f1-f28
Relationships:
V1 = 1*f1
V2 = 1*f2
V3 = 1*f3
V4 = 1*f4
V5 = 1*f5
V6 = 1*f6
V7 = 1*f7
V8 = 1*f8
V9 = 1*f9
V10 = 1*f10
V11 = 1*f11
V12 = 1*f12
V13 = 1*f13
V14 = 1*f14
V15 = 1*f15
V16 = 1*f16
V17 = 1*f17
V18 = 1*f18
V19 = 1*f19
V20 = 1*f20
V21 = 1*f21
V22 = 1*f22
V23 = 1*f23
V24 = 1*f24
V25 = 1*f25
V26 = 1*f26
V27 = 1*f27
V28 = 1*f28
Set the Error Variances of V1 - V28 to zero
Group AUSD:
Observed Variables: V1-V28
Covariance Matrix from File AUSD.COV
```

Asymptotic Covariance Matrix from File AUSD.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 500
 Sample Size: 1032
 Latent Variables: f1-f28
 Group CAN:
 Observed Variables: V1-V28
 Covariance Matrix from File CAN.COV
 Asymptotic Covariance Matrix from File CAN.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 500
 Sample Size: 1101
 Latent Variables: f1-f28
 Group IRLN:
 Observed Variables: V1-V28
 Covariance Matrix from File IRLN.COV
 Asymptotic Covariance Matrix from File IRLN.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 500
 Sample Size: 358
 Latent Variables: f1-f28
 Group NWZL:
 Observed Variables: V1-V28
 Covariance Matrix from File NWZL.COV
 Asymptotic Covariance Matrix from File NWZL.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 500
 Sample Size: 296
 Latent Variables: f1-f28
 Group SGP:
 Observed Variables: V1-V28
 Covariance Matrix from File SGP.COV
 Asymptotic Covariance Matrix from File SGP.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 500
 Sample Size: 378
 Latent Variables: f1-f28
 Group USA:
 Observed Variables: V1-V28
 Covariance Matrix from File USA.COV
 Asymptotic Covariance Matrix from File USA.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 500
 Sample Size: 389
 Latent Variables: f1-f28
 End of Problem

Ek 3: Ölçme Değişmezliği Modelleri Testi Model A Örnek Lisrel Sözdizimi

Group 1: Testing Equality Of Factor Structures
 ModelA: Factor Loadings, Factor Correlation, Error Variances
 Invariant
 Observed Variables: V1-V28
 Covariance Matrix from File ENG.COV
 Asymptotic Covariance Matrix from File ENG.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 836
 Latent Variables: Factor1
 Relationships:
 V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
 V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1

Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File AUSD.COV
Asymptotic Covariance Matrix from File AUSD.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 1032
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File CAN.COV
Asymptotic Covariance Matrix from File CAN.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 1101
Group 4: Testing Equality Of Factor Correlations
Covariance Matrix from File IRLN.COV
Asymptotic Covariance Matrix from File IRLN.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 358
Group 5: Testing Equality Of Factor Correlations
Covariance Matrix from File NWZL.COV
Asymptotic Covariance Matrix from File NWZL.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 296
Group 6: Testing Equality Of Factor Correlations
Covariance Matrix from File SGP.COV
Asymptotic Covariance Matrix from File SGP.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 378
Group 7: Testing Equality Of Factor Correlations
Covariance Matrix from File USA.COV
Asymptotic Covariance Matrix from File USA.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 389
End of Problem

Ek 4: Ölçme Değişmezliği Modelleri Testi Model B Örnek Lisrel Sözdizimi

Group 1: Testing Equality Of Factor Structures
Model B: Factor Correlation, Error Variances Invariant
Observed Variables: V1-V28
Covariance Matrix from File ENG.COV
Asymptotic Covariance Matrix from File ENG.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 836
Latent Variables: Factor1
Relationships:
V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File AUSD.COV
Asymptotic Covariance Matrix from File AUSD.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 1032
Relationships:
V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
Group 3: Testing Equality Of Factor Correlations

Covariance Matrix from File CAN.COV
 Asymptotic Covariance Matrix from File CAN.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 1101
 Relationships:
 V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
 V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
 Group 4: Testing Equality Of Factor Correlations
 Covariance Matrix from File IRLN.COV
 Asymptotic Covariance Matrix from File IRLN.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 358
 Relationships:
 V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
 V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
 Group 5: Testing Equality Of Factor Correlations
 Covariance Matrix from File NWZL.COV
 Asymptotic Covariance Matrix from File NWZL.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 296
 Relationships:
 V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
 V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
 Group 6: Testing Equality Of Factor Correlations
 Covariance Matrix from File SGP.COV
 Asymptotic Covariance Matrix from File SGP.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 378
 Relationships:
 V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
 V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
 Group 7: Testing Equality Of Factor Correlations
 Covariance Matrix from File USA.COV
 Asymptotic Covariance Matrix from File USA.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 389
 Relationships:
 V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
 V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
 End of Problem

Ek 5: Ölçme Değişmezliği Modelleri Testi Model C Örnek Lisrel Sözdizimi

Group 1: Testing Equality Of Factor Structures
 Model C: Factor Correlation Invariant
 Observed Variables: V1-V28
 Covariance Matrix from File ENG.COV
 Asymptotic Covariance Matrix from File ENG.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 836
 Latent Variables: Factor1
 Relationships:
 V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
 V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
 Group 2: Testing Equality Of Factor Correlations
 Covariance Matrix from File AUSD.COV

Asymptotic Covariance Matrix from File AUSD.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 1032
V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
Set the Error Variances of V1-V28 free
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File CAN.COV
Asymptotic Covariance Matrix from File CAN.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 1101
V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
Set the Error Variances of V1-V28 free
Group 4: Testing Equality Of Factor Correlations
Covariance Matrix from File IRLN.COV
Asymptotic Covariance Matrix from File IRLN.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 358
V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
Set the Error Variances of V1-V28 free
Group 5: Testing Equality Of Factor Correlations
Covariance Matrix from File NWZL.COV
Asymptotic Covariance Matrix from File NWZL.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 296
V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
Set the Error Variances of V1-V28 free
Group 6: Testing Equality Of Factor Correlations
Covariance Matrix from File SGP.COV
Asymptotic Covariance Matrix from File SGP.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 378
V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
Set the Error Variances of V1-V28 free
Group 7: Testing Equality Of Factor Correlations
Covariance Matrix from File USA.COV
Asymptotic Covariance Matrix from File USA.ACM
Method of Estimation: Maximum Likelihood
Iterations: 1000
Sample Size: 389
V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
Set the Error Variances of V1-V28 free
End of Problem

Ek 6: Ölçme Değişmezliği Modelleri Testi Model D Örnek Lisrel Sözdizimi

Group 1: Testing Equality Of Factor Structures
Model D: Factor loadings and Factor Correlation Invariant
Observed Variables: V1-V28
Covariance Matrix from File ENG.COV
Asymptotic Covariance Matrix from File ENG.ACM
Method of Estimation: Maximum Likelihood

Iterations: 1000
 Sample Size: 836
 Latent Variables: Factor1
 Relationships:
 V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
 V20 V21 V22 V23 V24 V25 V26 V27 V28=Factor1
 Group 2: Testing Equality Of Factor Correlations
 Covariance Matrix from File AUSD.COV
 Asymptotic Covariance Matrix from File AUSD.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 1032
 Set the Error Variances of V1-V28 free
 Group 3: Testing Equality Of Factor Correlations
 Covariance Matrix from File CAN.COV
 Asymptotic Covariance Matrix from File CAN.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 1101
 Set the Error Variances of V1-V28 free
 Group 4: Testing Equality Of Factor Correlations
 Covariance Matrix from File IRLN.COV
 Asymptotic Covariance Matrix from File IRLN.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 358
 Set the Error Variances of V1-V28 free
 Group 5: Testing Equality Of Factor Correlations
 Covariance Matrix from File NWZL.COV
 Asymptotic Covariance Matrix from File NWZL.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 296
 Set the Error Variances of V1-V28 free
 Group 6: Testing Equality Of Factor Correlations
 Covariance Matrix from File SGP.COV
 Asymptotic Covariance Matrix from File SGP.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 378
 Set the Error Variances of V1-V28 free
 Group 7: Testing Equality Of Factor Correlations
 Covariance Matrix from File USA.COV
 Asymptotic Covariance Matrix from File USA.ACM
 Method of Estimation: Maximum Likelihood
 Iterations: 1000
 Sample Size: 389
 Set the Error Variances of V1-V28 free
 End of Problem

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Eda CEYHAN
Doğum Yeri ve Tarihi : Çay-Afyonkarahisar / 07.07.1988

Eğitim Durumu

Lisans Öğrenimi : Orta Doğu Teknik Üniversitesi, Eğitim Fakültesi,
İngilizce Öğretmenliği
Yüksek Lisans Öğrenimi : Akdeniz Üniversitesi, Eğitim Bilimleri Anabilim Dalı,
Eğitimde Ölçme ve Değerlendirme Bilim Dalı
Bildiği Yabancı Diller : İngilizce, İspanyolca
Bilimsel Faaliyetler : IATEFL Testing, Evaluation & Assessment Special
Interest Group Pre-Conference Event, Liverpool, UK, 1
April 2019.

İş Deneyimi

Çalıştığı Kurumlar : Mehmet Akif Ersoy Üniversitesi Yabancı Diller
Yüksekokulu

İletişim

E-Posta Adresi : edaceyhan@mehmetakif.edu.tr

Tarih : 19.06.2019

BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kâğıt ve elektronik kopyalarının Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

Tezimin/Raporumun tamamı her yerden erişime açılabilir.

02.08.2019

Eda CEYHAN

İNTİHAL RAPORU

EDA CEYHAN TEZ TESLİM SON BENZERLİK RAPORU

ORIJİNALLIK RAPORU

% 10	% 7	% 4	% 7
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	acikarsiv.ankara.edu.tr İnternet Kaynağı	% 1
2	Submitted to Hacettepe University Öğrenci Ödevi	% 1
3	Submitted to Ankara University Öğrenci Ödevi	% 1
4	readgur.com İnternet Kaynağı	<% 1
5	www.ices-uebk.org İnternet Kaynağı	<% 1
6	acikerisim.deu.edu.tr İnternet Kaynağı	<% 1
7	pisa.meb.gov.tr İnternet Kaynağı	<% 1
8	www.turkpsikolojiyazilari.com İnternet Kaynağı	<% 1
9	dergiler.ankara.edu.tr İnternet Kaynağı	<% 1