



**T.C.
EGE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü**

**PISA 2015 OKUMA BECERİLERİ TESTİNİN ANA DİLİ
DEĞİŞKENİNE GÖRE ÖLÇME DEĞİŞMEZLİĞİNİN
İNCELENMESİ**

PELİN BAĞDU SÖYLER

**Eğitim Bilimleri Anabilim Dalı
Ölçme ve Değerlendirme
Yüksek Lisans Programı**

İZMİR

2020

T.C.
EGE ÜNİVERSİTESİ
Eđitim Bilimleri Enstitüsü

PISA 2015 OKUMA BECERİLERİ TESTİNİN ANA DİLİ
DEĐİŐKENİNE GÖRE ÖLÇME DEĐİŐMEZLİĐİNİN
İNCELENMESİ

PELİN BAĐDU SÖYLER

Yüksek Lisans Tezi

Eđitim Bilimleri Anabilim Dalı
Ölçme ve Deđerlendirme Yüksek Lisans Programı

Tez Danıőmanı:
Doç. Dr. Hakan ATILGAN



EĞİTİM BİLİMLERİ ENSTİTÜSÜ
TEZ SAVUNMA SINAV TUTANAĞI

Pelin BAĞDU SÖYLER'in "PISA 2015 Okuma Becerileri Testinin Ana Dili Değişkenine Göre Ölçme Değişmezliğinin İncelenmesi" başlıklı tezi 10/01/2020 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Ege Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği"nin ilgili maddeleri uyarınca, Eğitim Bilimleri Anabilim Dalı (Eğitimde Ölçme ve Değerlendirme Bilim Dalı) Yüksek Lisans Yeterlik tezi olarak oy birliğiyle kabul edilmiştir.


Prof. Dr. Tuncay ÖĞRETMEN

Jüri Başkanı



Doç. Dr. Hakan ATILGAN (Danışman)

Üye



Doç. Dr. İrfan YURDABAKAN

Üye

ETİK KURALLARA UYGUNLUK BEYANI

Ege Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğüne sunduğum “PISA 2015 Okuma Becerileri Testinin Ana Dili Değişkenine Göre Ölçme Değişmezliğinin İncelenmesi” adlı yüksek lisans tezinin tarafımdan bilimsel, ahlak ve normlara uygun bir şekilde hazırlandığını, tezimde yararlandığım kaynakları bibliyografyada ve dipnotlarda gösterdiğimi onurumla doğrularım.


Pelin BAĞDU SÖYLER

ÖNSÖZ

Eđitim sistemlerine yön veren, günümüzün deęişip geliřmekte olan eđitim anlayışını önemli ölçüde etkileyen parçalarından biri de ölçme ve deęerlendirme süreçleridir. Öğretimin etkililięi hakkında karar verme işinin en önemli aşaması ise ölçmedir. Ölçme, günlük hayatın içinde olan bir eylem olduęu gibi bilimsel çalışmaların da ayrılmaz bir unsurudur.

Uluslararası alanda uygulanan, ülkemizin de katıldığı PISA ve TIMMS gibi sınavlar ülkelerin eđitim sistemlerinin ne derece etkili olduęunu ölçmeyi amaçlamaktadırlar. Ana dilleri ve kültürleri birbirinden oldukça farklı olan birçok ülkenin eđitim sistemlerinin karşılaştırılması için kullanılan ölçme aracının tüm ülkeler için eşdeęerlięinin sağlanmış olması son derece önemli bir konudur. Bu nedenle uluslararası sınavlarda ölçme deęişmezlięi çalışmaları dikkatimi çekmiştir. PISA 2015 uygulaması okuma becerileri testinin dil deęişkeni açısından ölçme deęişmezlięi üzerine yapılmış bir çalışma olmadığından tezimin ilerlemesi bu yönde olmuştur.

Bu arařtırmayı yürütürken edindiğim tüm bilgi ve tecrübeler, öğretmenlik mesleęime ve akademik alanda gelişimime önemli katkılar sağlamıştır. Keyifle ve merakla sürdürdüğüm yüksek lisans eđitimim sonrasında, aynı motivasyonla başladığım bu çalışma benim için son derece verimli bir süreç olmuştur.

Arařtırmacılara, uygulayıcılara ve bilime katkı sağlaması dileęiyle...

İzmir, 2020

Pelin BAĞDU SÖYLER

TEŞEKKÜR

Yüksek lisans eğitim sürecimin tüm aşamalarında yol gösterici olan, bilgi ve deneyimlerini her zaman benimle paylaşan, akademik gelişimimde büyük katkıları bulunan değerli tez danışman hocam Doç. Dr. Hakan ATILGAN'a,

Tüm derslerinde sorgulayarak öğrenmemizi sağlayan, kendilerinden çok şey öğrendiğim ve derslerine keyifle katıldığım bölüm hocalarım Prof. Dr. Tuncay ÖĞRETMEN ve Doç. Dr. Tahsin Oğuz BAŞOKÇU'ya,

Tezimin gelişmesinde çok önemli katkıları olan, akademik alanda tüm bilgi ve önerilerini hiç esirgemeyen Dr. Öğr. Gör. Burak AYDIN'a,

Tez çalışma konumuyla yakından ilgilenip, alandaki tecrübelerini içtenlikle paylaşan Arş. Gör. Merve AYVALLI'ya,

Birlikte lisansüstü eğitim sürecini son derece verimli yürüttüğümüzü düşündüğüm ve bir dostluğu paylaştığımız dönem arkadaşlarıma,

Hayatımın her sürecinde bana destek olan, cesaretlendiren ve beni bugünlere getiren canım annem Ayşe BAĞDU ve canım babam Erol BAĞDU'ya; pozitif enerjileri ile her zaman güçlerini hissettiğim canım kardeşlerim Kazım ve Nilsu BAĞDU'ya

Birlikte bir yaşamı paylaştığım, hayatta her yükümü azaltan, tüm başarılarımda yanımda olan ve her türlü desteği sağlayan canım eşim Serhat SÖYLER'e,

Ve tezimin gelişmesinde katkısı olup beni destekleyen herkese çok teşekkür ederim.

Tezimi, yazmaya başladığım ilk zamanlardan beri benimle olan, bu süreçte dünyaya gelip hayatıma anlam katan, her haliyle, enerjisi ve neşesiyle motivasyon kaynağım olan biricik oğlum Bilge Aras SÖYLER'e ithaf ediyorum.

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı	Pelin BAĞDU SÖYLER
Doğum Yeri	Ankara
Doğum Tarihi	08/02/1988

Eğitim Durumu

Lise	Polatlı Anadolu Öğretmen Lisesi	2006
Lisans	Eskişehir Osmangazi Üniversitesi İlköğretim Fen Bilgisi Öğretmenliği	2010
	Eskişehir Osmangazi Üniversitesi İlköğretim Matematik Öğretmenliği	2010
Yüksek Lisans	Ege Üniversitesi Eğitim Bilimleri Enstitüsü, Eğitim Bilimleri Ana Bilim Dalı Ölçme ve Değerlendirme Yüksek Lisans Programı	2020

İş Deneyimi

Çalıştığı Kurumlar	Muş Selçuklu Ortaokulu	2012-2013
	Bergama Gazi Ortaokulu	2013-2014
	Konak Kemal Atatürk Ortaokulu	2014-Halen

Çalıştaylar

CERN Türk Öğretmen Çalıştayı (TTP-4) 2015

ÖZET

Giriş

Eğitim, bireyin davranışlarında kendi yaşantılarından yola çıkarak planlı bir biçimde yapılan davranış değişikliği oluşturma sürecidir. Bu sürecin etkililiği hakkında karar verme işi için davranışların ölçülmesi son derece önemli bir aşamadır. Eğitimde ölçme araçlarının güvenilirlik ve geçerlik sorunu olmaması, değerlendirme sürecinde alınan kararların da güvenilir olmasını sağlar.

PISA gibi uluslararası düzeyde yapılan sınavlarda, ülkelerin eğitim sistemlerinin etkililiği hakkında değerlendirmeler yapılmaktadır. Dolayısıyla, bu uygulamalar için hazırlanan ölçme araçlarının geçerliği incelenirken farklı değişkenlere göre eşdeğerliğinin de sınanması oldukça önemli bir husustur. PISA uygulamasının soruları İngilizce dilinde hazırlanmaktadır. Birçok ülkeden katılan öğrenciler, testin orijinal formunu değil çeviri formunu cevaplamaktadırlar. PISA'nın uygulama dilinden farklı bir dil kökeninden katılım sağlayan öğrenciler ile testi orijinal formda alan öğrenciler arasında dil değişkenine göre bir ölçme değişmezliği sorunu olmaması gerekmektedir. Bu doğrultuda araştırmanın problemini, PISA 2015 okuma becerileri testinin ana dili İngilizce olan ülkeler ile ana dili İngilizce olmayan ülkeler arasında dil değişkenine göre ölçme değişmezliğinin test edilmesi oluşturmaktadır.

Araştırmanın Amacı

Bu çalışmanın amacı Uluslararası Öğrenci Değerlendirme Programı (PISA-2015) okuma becerileri alt testinde ana dili İngilizce olan ülkeler ve İngilizce olmayan ülkelere göre çeviriden kaynaklanan bir ölçme değişmezliği problemi olup olmadığını araştırmaktır.

Yöntem

Araştırmanın amacı doğrultusunda ana dili İngilizce olan ülkelere Kanada, ABD ve İngiltere; ana dili İngilizce olmayan ülkelere ise Japonya, Tayland ve Türkiye örneklemi araştırmaya dahil edilmiştir. Ana dili İngilizce olan ve ana dili İngilizce olmayan ülkelere oluşturulacak iki grup için, ülke seçimleri yapılırken ülkelerin PISA 2015 okuma becerileri testindeki başarı ortalamaları ve sınavı bilgisayar tabanlı uygulama durumları dikkate alınmıştır.

Araştırmaya seçilen ülkelerden toplam 64171 öğrenci katılmıştır. PISA 2015 uygulamasında bilgisayar tabanlı test alan ülkelere 66 farklı form hazırlanmıştır. Bu çalışmada Kanada, İngiltere, ABD, Japonya, Tayland ve Türkiye’de ortak olarak uygulanmış ve en çok cevaplanan form olduğu için 41 numaralı form seçilmiştir. Araştırma bu formu alan 1526 kişilik örneklem üzerinden yürütülmüştür. Veri seti ana dili İngilizce olan ülkeler (Kanada, ABD ve İngiltere) ve ana dili İngilizce olmayan ülkelerden (Tayland, Japonya, Türkiye) oluşan iki gruba ayrılmıştır. İlk etapta doğrulayıcı faktör analizi ile ölçme modeli doğrulanmış, madde parametreleri incelenmiş; daha sonra ölçme değişmezliği ikili değişkenlerde çok gruplu ölçme değişmezliği analizi ile test edilmiştir. Madde faktör yükleri ve eşik değerler analize dahil edilen 28 madde için ayrı ayrı serbest bırakılarak söz konusu gruplar bazında BIC (Bayesian Information Criterion) değerleri ve madde parametrelerindeki değişim incelenmiştir.

Bulgular

Bu araştırmada yapılan analizler sonucunda ulaşılan bulgular araştırmanın alt problemlerine göre incelenmiştir. Araştırmanın birinci alt problemine göre her bir grup için ayrı ayrı DFA yapılmış ve model doğrulanmıştır. Elde edilen model uyum iyiliği indekslerinin kabul edilebilir düzeyde olduğu görülmüştür.

Araştırmanın ikinci alt problemine göre, DFA sonuçlarının her bir grupta maddeler bazında Madde Tepki Kuramına (MTK) göre madde parametreleri incelenmiş ve birbiriyle karşılaştırılmıştır. Madde 4, 9, 12, 18, 21, 27 ve 28’e ilişkin madde faktör yüklerinin ana dili İngilizce olan ülkeler ile ana dili İngilizce olmayan ülkelerden oluşan gruplar arasında birbirinden farklılık gösterdiği görülmüştür. Madde 2, 3, 6, 8, 10, 15, 22, 23, 25, 26 ve 27’ye ait eşik değerler ve a, b parametrelerinin de söz konusu iki grup arasında birbirinden oldukça farklı olduğu anlaşılmıştır.

Araştırmanın üçüncü alt problemine göre ikili puanlanmış maddelerde ölçme değişmezliği analizi yapılmış, her bir maddenin M_0 serbest modeline göre BIC değerlerinin farkı alınmıştır. İkinci alt problemde madde parametreleri arasında farklılık olan maddelerle tutarlı olarak 17 maddenin ölçme değişmezliğini sınırladığı görülmüştür.

Sonuç

Elde edilen bulgular doğrultusunda, PISA 2015 okuma becerileri testindeki maddelerin yarısından fazlasının ana dili İngilizce olan ve olmayan ülkelere göre eşdeğerliğinde büyük ölçüde sınırlamalar olduğu sonucuna ulaşılmıştır. Dolayısıyla PISA 2015 okuma becerileri testinin dil değişkeni açısından ölçme değişmezliğini tam olarak sağlamadığı, bireylerin bu testten aldıkları puanların ülkeler arası karşılaştırılmasının büyük oranda doğru sonuç vermeyeceği yargısına varılabilir.

Anahtar Sözcükler: Okuma becerileri, PISA 2015, Ölçme değişmezliği, Değişen Madde Fonksiyonu, İkili Değişkenlerde Çok Gruplu Ölçme Değişmezliği Analizi

Danışman: Doç. Dr. Hakan ATILGAN, Ege Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı

**INVESTIGATION OF THE MEASUREMENT VARIABILITY OF PISA 2015
READING SKILLS TEST ACCORDING TO THE LANGUAGE
VARIABILITY**

Pelin BAĞDU SÖYLER

ABSTRACT

Introduction

Education is the process of creating a change in the behavior of the individual in a planned manner based on their own experiences. Measuring behaviors for decision-making about the effectiveness of this process is a crucial step. Acceptable reliability and validity of the measurement tools ensures that the decisions taken during the evaluation process are reliable.

International examinations such as International Student Assessment Program(PISA) assess the effectiveness of countries' education systems. Therefore, it is very important to ensure the reliability and validity of the measurement tools prepared for these applications as well as their measurement equivalence with respect to different variables. Questions of the PISA application are prepared in English. Students from many countries cannot take the tests in the original form and they answer the translation form of the test. There should be no problem of measurement invariance for the students who participate in a different language origin than the application language of PISA and the students who took the test in the original form. In this respect, the problem of the study is the test of the measurement invariance according to the language variable between the native English-speaking countries of the PISA 2015 reading skills test and the non-native English-speaking countries.

Purpose

The aim of this study is to investigate whether there is a measurement invariance problem arising from translation in the PISA 2015 reading skills subtest compared to native English and non-English countries.

Method

For the purpose of the study, Canada, the US and the UK as the countries that native speakers of English and Japan, Thailand and Turkey as the non-English speaking countries were included in the study sample. For the two groups which will be composed of native speakers of English and non-native speakers of English, the country averages and computer-based application of the exams in the PISA 2015 reading skills test were taken into consideration.

A total of 64171 students from selected countries participated in the study. In PISA 2015, 66 different forms were prepared for countries that received computer-based tests. In this study, data from the 41st form were used given that it was the most frequently used form for Canada, UK, USA, Japan, Thailand and Turkey. The research was conducted on a sample of 1526 people who took this form. Data set was divided into two groups consisting of countries with native and non-native language English. In the first stage, the measurement model was confirmed by confirmatory factor analysis, item parameters were examined, and then the measurement invariance was tested with measurement invariance analysis for binary variables. Item factor loadings and thresholds were released separately for the 28 items included in the analysis, and the BIC (Bayesian Information Criterion) values and the change in item parameters were examined.

Findings

The findings obtained as a result of the analyzes carried out in this research are handled according to the sub-problems of the research. According to the first sub-problem of the study, CFA was performed separately for each group and the model was confirmed. The model fit indices were found to be acceptable.

According to the second sub-problem of the study, in each group, the parameters of the CFA were analyzed and compared with each other. It was observed that the item factor loadings for items 4, 9, 12, 18, 21, 27 and 28 differed from each other between the native English-speaking countries and the non-native English-speaking groups. The thresholds and other item parameters as to 2, 3,6,8,10,15,22,23,25,26 and 27th items were found to be quite different each other between the two groups.

According to the third sub-problem of the study, measurement invariance analysis was performed for the binary items and the difference of BIC values according to M_0 nested model of each item was found. In the second sub-problem, it was found that 17 items limited the measurement invariance consistent with the items that differed between the item parameters.

Results

Based on the findings, it was concluded that more than half of the items in the PISA 2015 reading skills test were not comparable for those of native and non-English speaking countries. Therefore, it can be concluded that the PISA 2015 reading skills test does not fully provide the measurement invariance in terms of language variable and that the comparison of the scores obtained by this test between countries might not provide accurate results.

Key Words: Reading skills, PISA 2015, Measurement Invariance, Differential item Functioning, Multivariate Measurement Invariance Analysis with Binary Items

İÇİNDEKİLER

ÖNSÖZ	iv
TEŞEKKÜR	v
ÖZGEÇMİŞ	vi
ÖZET	vii
Giriş.....	vii
Araştırmanın Amacı	vii
Yöntem.....	vii
Bulgular.....	viii
Sonuç.....	ix
ABSTRACT	x
İÇİNDEKİLER	xiii
TABLolar LİSTESİ	xiv
SİMGELER VE KISALTMALAR	xv
TANIMLAR	xvi
GİRİŞ	1
1.1 Problem Durumu	14
1.2 Problem Cümlesi	17
1.3 Araştırmanın Amacı ve Önemi.....	17
1.4 Sınırlılıklar	18
1.5 Konu İlgili Ulusal ve Uluslararası Düzeyde Yapılmış Çalışmalar	19
BÖLÜM II	25
YÖNTEM	25
2.1 Araştırmanın Modeli	25
2.2 Çalışma Grubu	25
2.3 Verilerin Elde Edilmesi.....	27
2.4 Verilerin Analizi.....	27
BÖLÜM III	29
BULGULAR	29
3.1 Araştırmanın 1. Alt Problemine Yönelik Bulgular	29
3.2 Araştırmanın 2. Alt Problemine Yönelik Bulgular.....	30
3.3 Araştırmanın 3. Alt Problemine Yönelik Bulgular.....	33
BÖLÜM IV	38
SONUÇ VE ÖNERİLER	38
4.1 Sonuçlar.....	38
4.2 Öneriler	40
KAYNAKÇA	43
EKLER	47

TABLÖLAR LİSTESİ

Tablo 1. Ülkeler Bazında Örneklem Frekansları	26
Tablo 2. PISA Okuma Becerileri Testine İlişkin Doğrulayıcı Faktör Analizi Sonuçları	29
Tablo 3. PISA 2015 Okuma Becerileri Testi Dil Değişkenine Göre Oluşturulan Grupların DFA Sonuçlarına İlişkin Madde Parametreleri	31
Tablo 4. İki Değişkenli Maddelerde Ölçme Değişmezliği Analizi Sonuçları	34



SİMGELER VE KISALTMALAR

KTK: Klasik Test Teoremi

MTK: Madde Tepki Kuramı

DMF: Değişen Madde Fonksiyonu

AFA: Açımlayıcı Faktör Analizi

DFA: Doğrulayıcı Faktör Analizi

ÇGDFA: Çoklu-grup doğrulayıcı faktör analizi

YEM: Yapısal eşitlik modellemesi

Sd: Serbestlik derecesi

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

CFI: Comparative fit index (karşılaştırmalı uyum indeksi)

TLI: Tucker-Lewis Index

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

BIC: Bayesian Criterion Information

TANIMLAR

Ölçme Değişmezliği: Ölçme aracının ölçtüğü psikolojik yapının farklı gruplar arasında karşılaştırılabilme durumunu ifade eder.

Doğrulayıcı Faktör Analizi: Gözlenen değişkenler ile örtük değişkenler arasındaki ilişkiyi araştıran ve faktör analizi üzerine kurulan hipotezlerin test edilmesi amacıyla sıkça kullanılan bir istatistiksel tekniktir.

Çok Gruplu Doğrulayıcı Faktör Analizi: Doğrulayıcı Faktör Analizinin iki veya daha fazla gruplu karşılaştırmaların zamandaş olarak yapılması gerektiği durumlarda işe koşulan yöntemidir.

İki Değişkenli Puanlanmış Maddelerde Ölçme Değişmezliği Analizi: İkili kategorik değişkenli puanlanmış veri setlerinin her bir madde bazında ölçme değişmezliğini test etmeyi sağlayan bir istatistiksel tekniktir.

Değişen Madde Fonksiyonu: Yetenek düzeyi aynı olan bireylerin bir maddeye doğru yanıt verme olasılıklarının değişmesini ifade eder.

Madde Tepki Kuramı: Bireylerin belli bir alandaki örtük özellikleriyle, bu alanı ölçen test maddelerine verdikleri cevaplar arasında ilişki kuran ve bu ilişkiyi matematiksel olarak ifade eden bir test kuramıdır.

BÖLÜM I

GİRİŞ

Eğitim, bireylerin davranışlarında istendik yönde değişiklikler oluşturmayı amaçlayan bir sistemdir. Bu davranış değişiklikleri yapılan bir takım etkinlikler sonucunda meydana gelir ve ortaya çıkan sonuçların bazıları beklendik yönde olmayabilir (Baykul ve Turgut, 2015). Eğitim sisteminin başarılı ya da başarısız yanlarının bilinip kontrol edilmesi ihtiyacı, değerlendirme kavramının önemini ortaya koymaktadır. Değerlendirme, bir karar verme işidir; dolayısıyla bir öğrencinin başarı ya da başarısızlığına dair doğru ve güvenilir kararlar vermek söz konusu davranışların ölçülmesinde son derece dikkatli olmayı gerektirir (Baykul ve Turgut, 2015).

Ölçme geniş anlamıyla, herhangi bir özelliğin gözlenerek sonucunun sayı, sembol veya sıfatlarla ifade edilmesidir. Farklı çokluklardaki niteliklerin birbiriyle kıyaslanmasından ölçme ihtiyacı doğmuştur. Ölçme terimi farklı bilim dallarında çeşitli biçimlerde tanımlanmıştır. Stevens (1946) ölçmeyi, “Nesnelerin ve durumların gözlenebilen özelliklerine bazı kurallar ile sayı atamaktır.” şeklinde açıklamıştır. Campbell (1928) ise ölçmeyi, “Maddelerin değişen niteliklerine, bu niteliklikleri tanımlayan kurallara göre sayılar vermektir” ifadesiyle açıklamıştır. Bir diğer tanım olarak ise Algina ve Crocker (1986) tarafından “Fiziki olaylar üzerinde bir gözleyen tarafından uygulanan işlemdir” ifadesiyle literatürde yer almaktadır. Ölçme, günlük yaşamımızda ve bilimin bütün dallarında önemli bir yere sahiptir. Bilimde ölçme öyle bir yere gelmiştir ki, ölçme teorileri geliştirilmiştir (Baykul, 1999). Bu teorilerden ölçme biliminde en çok bilinen ve kullanılanları Klasik Test Kuramı (KTK), Madde Tepki Kuramı (MTK) ve Genellenebilirlik (G) kuramıdır.

Geliştirilen ölçme teorileri eğitim ve psikoloji gibi sosyal bilimlerde ölçmeye farklı yaklaşımlar ve yöntemler sunmuştur. Fiziksel bilimlere deneysel yöntemlerle ölçmek, sosyal bilimlere ölçmeye kıyasla daha kesin sonuçlar verir. Dolayısıyla, sosyal bilimlerde ölçme, ölçülecek özelliğin tanımlanmasının zorluğu ve ölçme araçlarının yüksek güvenilirlikte olması gerekliliği ile fiziksel ölçmeden ayrılır. Eğitim ve psikolojide ölçme; boy, ağırlık, yaş gibi doğrudan gözlenebilen fiziksel ölçmelerden

farklıdır. Eğitim sürecinde amaçlanan istendik davranışlar vardır ve bu davranışların ne ölçüde kazanıldığına ortaya çıkarılması gerekmektedir. Bu doğrultuda eğitimde ölçme son derece önemlidir. Bireyin sahip olduğu özelliklerin ve davranışların ölçülmesi için, öncelikle bunların tanımlanabilmesi gerekir. Psikolojik ve davranışsal özellikler çoğu zaman doğrudan tanımlanamaz, o davranışı ya da özelliği ortaya çıkaran işaretçileri yoluyla tanımlanırlar. Bu davranışsal ve psikolojik özelliklerin işaretçileri de bir takım davranışlar zincirinden oluşur. Ölçme, ölçülecek olan özelliğin bir ölçekle karşılaştırarak ölçek birimiyle ifade edilmesidir. Dolayısıyla davranışsal özelliklerin ölçülebilmesi fiziksel özellikler gibi nesnel ve kesin bir ölçütü olmamasından dolayı hatalardan yüksek derecede etkilenir (Özçelik, 2011). Eğitim ve diğer sosyal bilimlerde karşımıza çıkan ölçme hatası sorununun en aza indirilmesi, ölçme aracının birimlerine ve ne kadar yüksek geçerlik ve güvenilirliğe sahip olduğuna bağlıdır.

Ölçme sonuçlarına farklı şekillerde hatalar karışır. Ölçme hatalarından sakınmak için hata kaynaklarının tespit edilmesi gerekir. Ölçmeye karışan hatalar istatistiksel olarak “sabit hata, sistematik hata ve tesadüfi hatalar” dır. Sabit hatalarda hatanın miktarı, yönü ve kaynağı bellidir. Sistematik hatalarda hata miktarı ve yönü bellidir ve bir sistematığe bağlıdır. Tesadüfi hatalarda ise hatanın kaynağı, miktarı ve yönü belirsizdir (Baykul, 1999). Davranış bilimlerinde yapılan ölçme ve değerlendirme çalışmalarında tesadüfi hataların araştırılması bir güvenilirlik çalışması iken sistematik hataların araştırılması yanlılık ile ilgili olup bir geçerlik çalışmasıdır. Sistematik hatalar özellikle yapı geçerliğini olumsuz etkilemektedir (Zumbo,1999). Ölçme aracı ne kadar duyarlı ve hatalardan arınık ise o kadar doğru sonuçlar verir. Bir ölçme aracının sahip olması gereken önemli özellikler güvenilirlik ve geçerliktir (Baykul,1999).

Güvenirlik, art arda yapılan ölçmelerde aynı sonuca ulaşılmasıdır. Bu aynı zamanda ölçme aracının kararlılık derecesini gösterir. Bir ölçme aracında bulunması gereken en temel özelliklerden biri güvenilirliktir (Wainer, 1990). Elde edilen ölçme sonucu tesadüfi hatalardan ne kadar arınıksa o kadar güvenilirdir. Başka bir anlamda güvenilirlik, ölçme sonuçlarının duyarlılığının da bir göstergesidir (Baykul veTurgut, 1992). Ölçme aracının kararlılık ve duyarlılık derecesi onun güvenilirliğinin bir ölçütüdür. Güvenirliği saptamanın çeşitli yolları vardır. Tesadüfi hatalardan

etkilendiği için, hatanın yönünü ve kaynağını belirlemek zordur. Hata varyanslarının kaynağına göre değiştiği için güvenilirliği belirleme yöntemleri farklı isimler alır. Güvenirlik aynı zamanda art arda yapılan ölçmelerin birbiriyle tutarlılığı anlamına geldiğinden, belirli zaman aralıklarında bir testin aynı gruba iki kez uygulanması ile ulaşılan sonuçlar arasındaki korelasyon test-tekrar test güvenilirliğidir. Burada ölçme aracının kararlılığı söz konusudur ve aradan geçen zaman hata kaynağıdır. Bir ölçüğün tamamen farklı versiyonunun bir örneklem grubuna uygulanması ile iki ölçme sonucu arasında elde edilen korelasyon paralel (eşdeğer) formlar güvenilirliğidir ve bu yöntemde tutarlılık anlamında güvenilirlikten söz edilir. Eşdeğer formlar arasında maddelerin güçlüğü, sayısı ve türünün birbirine çok yakın olması beklenmektedir, dolayısıyla hata kaynağı ölçme aracıdır. Paralel test varsayımı her zaman sağlanamayabilir, bu durumda iki yarı güvenilirliği yöntemini kullanmak daha pratik olabilir. Ölçeği oluşturan madde havuzu iki yarıya bölünerek, iki eşdeğer yarı aynı örneklem grubuna uygulanır. İki ölçme sonucu arasındaki korelasyon katsayısı iki yarı (eşdeğer yarılar) güvenilirliğini verir (DeVellis, 1991). Bir diğer güvenilirlik belirleme yöntemi de Puanlayıcı güvenilirliğidir. Bireyin performansının birden fazla puanlayıcı tarafından puanlanması ile elde edilen sonuçlar arasındaki korelasyona bakılır. Burada hata kaynağı olarak puanlayıcı karşımıza çıkar, ölçme sonuçlarının tutarlılığı söz konusudur (Atılğan, Aydın ve Kan, 2017).

Güvenirlik, hatalardan arınlık derecesi olarak da tanımlandığı için gerçek puana ulaşmak en az hatayla ölçme gerektirir. X gözlenen puan ve T gerçek puan olarak ifade edildiğinde, güvenilirlik indeksi gözlenen puan ile gerçek puan arasındaki korelasyonun karesi olarak tanımlanır. Paralel iki ölçme arasındaki korelasyon, güvenilirlik katsayısıdır ve aşağıdaki eşitlikle gösterilir (Baykul, 2015 ; Lord ve Novick, 1968).

$$\sigma^2(X) = \sigma^2(T) \quad (1)$$

Güvenirlik katsayısı hesaplamasının güvenilirlik belirleme yöntemlerine göre çeşitli yolları vardır. Klasik test teorisinde en çok kullanılan güvenilirlik katsayısı hesaplama yöntemleri Kuder-Richardson-20 (KR-20) ve Kuder-Richardson-21 (KR-21) yöntemleridir. Bu yöntemler 1-0 puanlanabilen testler için uygundur. Eğer test 1-

0 puanlanmayıp çok kategorili puanlanabilen maddeler içeriyorsa bu yönteme benzeyen Cronbach-Alfa yöntemi kullanılır (Atılğan, Aydın ve Kan, 2017; Algina ve Crocker, 1986). Güvenirlik katsayısı test uzunluğu, grubun homojenliği, soruların cevaplanma süresi gibi faktörlerden etkilenir (Baykul,2015).

Güvenirlik ve geçerlik ikisi birlikte bir ölçme aracında olması gereken en önemli özelliklerdir. Bir ölçme aracının güvenilir olması geçerli olduğunu garantilemez. Dolayısıyla bir ölçme aracı güvenilir olsa da geçerli olmayabilir. Geçerlik bir ölçme aracının ölçülmek istenen niteliği başka bir değişkenle karıştırmadan ne kadar doğru ölçtüğüyle ilgilidir. Yani bir ölçme aracının amacına hizmet etme derecesi geçerlik olarak ifade edilir. Ölçme aracının kullanılma amacına göre farklı geçerlik türleri vardır (Baykul ve Turgut, 1992; Hagen ve Thorndike, 1959). Eğitimde ölçmede en çok kullanılan geçerlik türleri; kapsam geçerliği, bir ölçüte dayalı geçerlik, yapı geçerliğidir.

Kapsam geçerliği, bir testin ölçmek istediği kazanımları ne kadar içerdiğiyle ilgili bir geçerlik türüdür. Hedef davranışların testte ne kadar mevcut olduğu kapsam geçerliği açısından önemlidir. Kapsam geçerliğini sağlamak için öncelikle yoklanacak davranışlar belirlenmeli ve buna göre belirtke tablosu oluşturulmalıdır. Belirtke tablosundaki hedeflenen davranışların testte ölçülmemesi testin kapsam geçerliğini düşürür. Bir testte kapsam geçerliğinin sağlanması değerlendirmenin de amaçları arasındadır.

Bir ölçüte dayalı geçerlik, terimin de ima ettiği gibi bir maddenin ya da ölçeğin bazı ölçütler ile ampirik ilişkisidir. Ölçütün zamanına göre uygunluk ve yordama geçerliği olarak iki şekilde incelenir. Ölçüt puanları ile yordayıcı puanlar aynı zamanlarda elde edilmişse bu geçerlik uygunluk geçerliğidir (Atılğan, Aydın ve Kan, 2017; Baykul, 2000; Algina ve Crocker, 1986; Hagen ve Thorndike, 1959). Ölçüt puanları, yordayıcı puanlardan sonra elde edilmişse yordama geçerliğinden söz edilir. Ölçüte dayalı geçerlik, klasik test teorisinde sıkça kullanılmıştır (Baykul, 2015).

Yapı geçerliği, genel anlamda bir ölçme aracının ölçmek istenilen teorik yapıyı ölçebilme derecesidir (Anastasi, 1997; Atılğan, Aydın ve Kan, 2017). Ölçmede yapı, ölçülecek birey ya da eşyanın sahip olduğu özellikler arasındaki ilişkileri içeren bir bütündür. Cronbach ve Meehl (1955) psikolojik yapıyı, “Bireylerin var olduğu kabul edilen özellikleridir” olarak tanımlamıştır. Yapı geçerliğini belirlemek için belirli

yöntemler vardır. Bu yöntemlerin başlıcaları: 1) test maddelerinin uzmanlar tarafından incelenmesi, 2) puanların kararlılığı (test-tekrar test), 3) diğer testlerle korelasyon, 4) grup farklılıkları, 5) pratik ölçütlerle korelasyon, 6) faktör analizi gibi yöntemlerdir.

Faktör analizi, testlerin yapı geçerliği çalışmalarında kullanılan en önemli istatistiksel yöntemlerden biridir. Bu analiz bir ölçme aracının yapı geçerliğine dair kanıt elde edilmesini sağlar ve ölçme değişmezliği çalışmalarının temelini oluşturur. Büyüköztürk (2002) tarafından faktör analizi, bir yapıyı ölçen pek çok değişkenden daha az sayıda ve anlamlı değişkenler oluşturmayı amaçlayan bir istatistik olarak tanımlanmıştır. Algina ve Crocker'a (1986) göre ise faktör analizinin üç temel amacı bulunmaktadır: Bunlar; (1) bir modele ait testlerin tüm çiftleri arasındaki korelasyonların hesaplanması için gerekli ortak faktör sayısını belirlemek, (2) testin iç korelasyonlarını bulmak için hesaplanan ortak faktörlerin sayısına karar vermek (3) varyans oranını saptamaktır. Ölçme aracı geliştirme ve uyarlamada kullanılan bir çok teknik, testin tek boyutlu olduğu varsayımına dayanır fakat çok boyutlu ölçme araçlarının geliştirilmesinde de hangi maddenin hangi boyuta gittiğini bilmek açısından faktör analizi oldukça işlevsel kullanılan bir araçtır (Baykul, 2015). Faktör analiziyle ilgili çalışmalar 19. Yüzyılda Galton, Pearson ve Yule'nin korelasyon ve regresyon konularındaki araştırmalarına dayanır. Bunlar, faktör analizinin hazırlık çalışmaları olarak bilinir (Mulaik, 1972). Faktör analizi matematiksel olarak şöyle açıklanabilir; değişkenler birer vektör olarak varsayılırsa, bu vektör uzayının koordinatlarını ve boyutlarını tespit etmede yararlanılan bir yöntemdir. Test geliştirmede, testin boyutlarını belirlemek açısından faktör analizi oldukça önemli bir tekniktir. Testin boyutlarına faktör, boyut sayısına faktör sayısı ve test koordinatlarına faktör yükleri denir (Baykul, 2015). Faktör yükleri belirlenirken çeşitli analiz yöntemleri vardır. Analizler sonucunda faktör matrisleri elde edilir ve ulaşılan yapıda faktör boyutları orijinal faktör matrisindeki yapıyı vermeyebilir. Bu durumda döndürülmüş faktör yüklerinin hesaplanması gerekir.

Faktör analizi Açımlayıcı Faktör Analizi (AFA) ve Doğrulayıcı Faktör Analizi (DFA) ile yapılır. Açımlayıcı faktör analizinde orijinal yapısı bilinmeyen bir yapı, ölçme aracından alınan sonuçlara göre açıklanmaya çalışılır. Psikolojik yapıların belirlenmesinde ve gizil değişkenlerin belirlenmesinde kullanılan bir istatistiksel yöntemdir. Doğrulayıcı faktör analizinde ise gerçek yapısı bilinen bir yapı, ölçme aracı

verilerine göre doğrulanmaya çabalanır (Atılğan, Aydın ve Kan, 2017). Daha önceden yapılandırılmış, kavramsal alt yapısı olan bir modeldeki faktör yükleri ve sayıları belirlenir. Veri seti ile modelin uyumu değerlendirilir. Bu şekilde modelin yapı geçerliği test edilir (Maruyama, 1998).

Psikolojik bir ölçme aracının yapı geçerliği ve güvenilirlik gibi özelliklerinin farklı gruplardaki eşitliği ölçme değişmezliği (eşdeğerliği) olarak tanımlanır (Herdman, 1998). Byrne ve Watkins (2003) tarafından ise ölçme değişmezliği, test maddelerinin farklı gruplar arasında tam olarak aynı şekilde yorumlanma düzeyi olarak tanımlanmıştır. Ölçülecek olan psikolojik yapının farklı kültürel etmenler veya değişkenler açısından gruplar arasında karşılaştırılabilir olup olmadığı ölçme değişmezliği sorununun temelidir. Başka bir deyişle, ölçme aracının ölçülen psikolojik yapıyı farklı gruplar arasında aynı şekilde ölçüyor olması gerekir. Ölçme eşdeğerliği çalışmaları, Klasik Test Kuramında (KTK) madde istatistikleri hesaplarının gruplar arası farklılıklardan etkilendiği sınırlılığı ile gündeme gelmiştir (Algina ve Crocker, 1986). Araştırmacılar, gruplar arası karşılaştırmalar yaparken aynı psikolojik özellik ölçümünde farklı sonuçlara ulaşmanın yalnızca bireyin özelliklerinden değil ölçme aracından da kaynaklanabileceğinden yola çıkarak ölçme değişmezliğini pek çok araştırmada kullanmışlardır.

Ölçme değişmezliği, bir ölçme modelinin çoklu gruplarda aynı yapıda olması, ölçekteki maddelerin faktör yapıları ve hata varyanslarının aynı olması demektir (Başusta, 2010; Bollen, 1989). Çoklu grup karşılaştırmalarında Yapısal Eşitlik Modeli (YEM) hipotezleri test etmek açısından işe koşulan bir modeldir. Yapısal eşitlik modeliyle yapılan uygulamalarda temel olarak; faktör sayısı ve faktör yük değerleri, regresyon katsayısı, artık regresyon varyansı, ortak faktörlerin ortalamaları, varyans ve kovaryansları elde edilir. Ölçme değişmezliği çalışmalarında daha çok, faktör sayıları ve yük değerleri, regresyon varyans ve kovaryansları örtük değişkenlerin dağılımı ile ilgili yeterli bilgiyi verir. Ölçme değişmezliği çalışmaları farklı amaçlar doğrultusunda farklı yöntemlerle yapılmaktadır. Veri setinin yapısına göre ölçme değişmezliği yöntemleri farklılık göstermektedir. Sürekli değişkenlerden oluşan veri setleri için ölçme değişmezliği aşamalı olarak test edilmektedir. Kategorik değişkenlerden oluşan veri setleri için ölçme değişmezliğinin MTK yaklaşımı ile

incelenmesi önerilmektedir (Dimitrov, Li, Marcoulides, Menold ve Raykov 2018). Bu yöntemler aşağıda açıklanmıştır.

Sürekli Değişkenlerle Ölçme Değişmezliği

Sürekli değişkenlerden oluşan veri setleri için ölçeğin ölçme değişmezliği, aşamalı olarak hipotezler oluşturularak test edilir. Her aşamada kurulan hipotez test edildikten sonra bir sonraki aşamaya geçilir. Bu şekilde, faktör yükleri, faktörler arası korelasyonlar ve hata varyansları gibi parametrelerin eşit olması ölçme eşitliğinin sağlandığı anlamına gelir (Wu, Li ve Zumbo, 2007). Doğrulayıcı faktör analizi tekniği ile ölçme değişmezliğinin test edilmesi gereken tüm aşamaları test edilebilir. Doğrulayıcı faktör analizi Yapısal Eşitlik Modellemesi (YEM) analizlerinin bir çeşididir. Modele ilişkin uyum indekslerinden TLI (NNFI), RMSEA, GFI, AGFI, SRMR değerleri dikkate alınmaktadır. Aşağıda bu uyum indeksleri açıklanmıştır.

Ki-kare (χ^2); en temel test istatistiğidir. Ölçme değişmezliği çalışmalarında elde edilen analiz sonuçlarına göre model uyumu kararı verilirken χ^2 istatistiği değeri ve serbestlik derecesi (sd) ile daha az sınırlı modele ait χ^2 değeri ve sd arasındaki fark ($\Delta\chi^2$ ve Δsd) karşılaştırılır. Elde edilen farkın anlamlılık düzeyi χ^2 tablosunda yer alan değerlerin $p < .01$ ve $p < .05$ düzeyinde karşılaştırılmasıyla belirlenir. Bu karşılaştırmada χ^2 tablosundaki değer ulaşılan sonuçtan büyük olduğunda modeller arasındaki farkın anlamlı olduğu kabul edilmektedir (Kline, 2005; Akt. Ayvalli, 2016;). Ki-kare istatistiği örneklem büyüklüğünden etkilenmektedir. Gözlenen değişkenler arasındaki korelasyon arttıkça Ki-kare değeri de artar; büyük örneklerde ise kovaryanslar arasındaki fark çok olsa bile ki-kare değeri anlamlı çıkabilir (Kline, 2011). Dolayısıyla χ^2 istatistiğinde örneklem büyüklüğü yanıltıcı sonuçlara sebep olabilir.

RMSEA; yaklaşık hataların ortalama kareköküdür. Parametre kestirimleriyle topluluğun kovaryans matrisine uyumunu gösterir. RMSEA istatistiği için, .06'ya yaklaşık bir kesme noktası iyi uyum göstergesi olarak kabul edilir; .05 ile 0 arasındaki bir değerde ise modelin mükemmel bir uyum gösterdiği anlaşılır (Hu ve Bentler, 1999). SRMR; ne kadar küçük değerde ise model o kadar mükemmel uyum verir anlamına gelmektedir. Kesme noktası .08'den küçük değerler kabul edilebilir uyum

kriteri olarak belirlenmiştir (Kline, 2011). SRMR .05'ten küçük değerlerde ise model mükemmel uyum gösteriyor olarak kabul edilir (Bentler ve Hu, 1999).

GFI uyum iyiliği ve AGFI düzeltilmiş uyum iyiliği istatistikleridir. Açıklanan varyans oranları ile tanımlanmaktadır. GFI kovaryans matrisinde artık madde varyansının değişkenliği ve toplam değişkenlik kullanılarak hesaplanır. AGFI ise GFI'nın modelin serbestlik derecesine göre düzeltilmiş halidir. Bu düzeltme ile modelin aşırı uyumunu engelleme amaçlanmıştır (Akt. Kıbrısoğlu, 2015). GFI ve AGFI 0 ile 1 arasında değer alır ancak kötü olarak tanımlanmış modellerde negatif değerler de olabilir. GFI için .90'a yakın kesme noktası kabul edilse de, örneklem küçük olduğunda .95 değeri uygun bulunmaktadır. AGFI .90 ve üzerinde değer alıyorsa model mükemmel uyum göstermektedir (Miles ve Shevlin, 2007).

CFI; karşılaştırmalı bir uyum iyiliği indeksidir. Araştırmada kurulan modelin bir taban modele göre ne derece iyi uyum gösterdiğinin bir istatistiğidir. Ki-kare değerinin serbestlik derecesine eşit olduğu durumlarda CFI'nın değeri 1'e eşittir. Kesme noktası .95 üzerindeki değerler modelin kabul edilebilir uyum gösterdiği anlamına gelmektedir (Kline, 2011).

Ölçme değişmezliği çalışmalarında örneklem büyüklüğünden etkilenmediği için karşılaştırmalı uyum indeksleri (ΔCFI , ΔTLI , $\Delta RMSEA$) kullanılması önerilmiştir (Cheng ve Rensvold, 2002). Chen (2007) tarafından kesme noktası olarak, modeller arasındaki farkın anlamlı kabul edilebilmesi için söz konusu indekslere ilişkin kesme noktaları; örneklem büyüklüğü 300'den fazla olduğunda metrik değişmezlik için $\Delta CFI \geq -.010$ ve $\Delta SRMR \geq .015$, güçlü ve katı faktöriyel değişmezliğin karşılaştırılmasında ise $\Delta CFI \geq -.010$ ve $\Delta SRMR \geq .010$ olarak kabul edilmektedir. Örneklem büyüklüğü 300'den az olduğunda metrik değişmezlik için $\Delta CFI \leq -.005$ ve $\Delta SRMR \geq .025$, güçlü ve katı faktöriyel değişmezliğe ait karşılaştırmalarda $\Delta CFI \geq -.005$, $\Delta SRMR \geq 0.005$ olarak önerilmektedir (Akt. Ayvalli, 2016).

TLI; NNFI (Non-normed fit index) olarak da adlandırılmaktadır. Bu istatistiğin örneklem büyüklüğünden önemli ölçüde etkilenmemesi iyi bir avantajdır. TLI değeri ne kadar büyük olursa model o kadar iyi uyum göstermektedir. TLI için .95 kesme noktasından büyük değerler kabul edilebilir uyum, .97 ve üzerindeki değerlerde olan kesme noktaları mükemmel uyum anlamına gelmektedir (Moosbrugger ve Schermelleh-Engel, 2003).

Alanda kullanılan pek çok uyum indeksi vardır. Analiz sonuçları raporlarında birden çok uyum indeksine yer verilmelidir. Literatürde en çok raporlanan indeksler CFI ve RMSEA'dır (Fidell ve Tabachnick, 2007).

Kültür, cinsiyet gibi değişkenlere bağlı olarak karşılaştırma yapılan çok gruplu araştırmalarda Çok Gruplu Doğrulayıcı Faktör Analizi (ÇGDFA) tekniği en uygun olanıdır. ÇGDFA, yapısal eşitlik modelinde oldukça sık kullanılmaktadır (Reise ve Widaman, 1997). DFA'nın iki veya daha fazla gruplu karşılaştırmaların zamandaş olarak yapılması gerektiği durumlarda işe koşulan yöntemidir. Dolayısıyla ölçme değişmezliği çalışmalarında ÇGDFA yaygın kullanılmaktadır. ÇGDFA bir ölçme aracının, farklı gruplar için aynı çalışıp çalışmadığını belirlemeyi amaçlar. Aynı zamanda farklı gruplar arasında ölçme sonuçlarının denk olup olmadığı ile ilgili madde yanlılığıyla da ilgilenir (Salzberger ve diğerleri, 1999 ; Önen, 2009). Yapı geçerliliğini kanıtlamak için en güvenilir yollardan biri olarak bilinmektedir. Ölçme değişmezliği çalışmalarında pek çok yöntem kullanılmakla birlikte en çok tercih edilen ve uygulanan yöntem Çok Gruplu Doğrulayıcı Faktör analizi yöntemidir (Akt. Kıbrısoğlu, 2015). Doğrulayıcı faktör analizi yapı geçerliliği ve gruplar arasındaki ölçme değişmezliği bilgilerini vermektedir (Gregoric, 2006). ÇGDFA analizleri ve ölçme değişmezliği aşamaları analiz edilirken gruplardan biri referans grup olarak belirlenir ve bu gruptaki veriler her aşamada sabitlenip diğer grup ya da gruplara ne derecede uyum sağladığı incelenir.

Sürekli yapıdaki veri setleri için ölçme değişmezliği aşamalı olarak test edilmektedir. Her bir aşama, bir önceki aşamanın ön koşulu olarak kabul edilmektedir. Bu sırasıyla aşamalar şu şekildedir:

- 1) Yapı (Şekil) Değişmezliği (Configural invariance)
- 2) Metrik Değişmezlik (Weak factorial invariance= Metric invariance)
- 3) Ölçek Değişmezliği (Strong invariance= Scalar invariance)
- 4) Katı Değişmezlik (Strict invariance) (Wu, Li ve Zumbo, 2007).

Yapı değişmezliği, ölçme değişmezliğinin ilk aşamasıdır. İlk olarak yapı değişmezliği sağlanamamışsa ölçme değişmezliğinden de bahsedilemez. Diğer değişmezlik aşamalarının sağlanabilmesi için yapı değişmezliği ön koşuldur. Yapı değişmezliği, farklı gruplar arasında aynı maddeleri ölçen faktör yapılarının ve maddelere ait parametrelerin aynı olmasını gerektirir. Dolayısıyla modelin gruplar

arasında aynı olup olmadığını test eder (Kline, 2011). Yapı değişmezliği sağlanmışsa bir sonraki aşama olarak metrik değişmezliğin sağlanıp sağlanmadığı test edilmelidir.

Metrik değişmezlik, yapı değişmezliğinde test edilen faktör yapılarına ek olarak faktör yüklerinin de gruplar arasında eşitliğini gerektirir. Bu aşama test edilirken faktör korelasyonları ve hata varyansları sabit, faktör yükleri ise serbest bırakılır. Faktör yüklerinin gruplar arasında eşit olması, maddelerin tüm gruplarda aynı şekilde algılandığını ve herhangi bir yanlılık olmadığı anlamına gelir (Hill ve Knight 1998). Metrik değişmezlik zayıf faktöriyel değişmezlik olarak da bilinir ve gruplar arasında metrik değişmezlik sağlandığında, kovaryans farklılıklarının ortak faktörlerden dolayı olduğu söylenebilse de bu ortalama farklılıkların kaynağı tam olarak bilinemez (Millsap ve Olivera-Aguilar, 2012). Metrik değişmezliğin sağlanması bir sonraki aşama olan ölçek değişmezliği için gerekli bir koşuldur.

Madde yanlılıklarının incelenmesi için yapı değişmezliği ve metrik değişmezliğin sağlanması yeterli olmaz, ölçek değişmezliğinin de test edilmesi gerekir. Metrik değişmezlik şartlarına ek olarak madde sabitlerinin de eşitliği şartı sağlanmalıdır. Madde sabitinin eşit olması örtük değişkenlerin faktör ortalamaları sıfıra eşitlendiğinde diğer grup ortalamalarının serbest bırakılmasıdır (Millsap ve Olivera-Aguilar, 2012). Ölçek değişmezliği, ölçme değişmezliği çalışmalarında güçlü bir aşamadır; madde yanlılığı konusunda daha net bilgi verir. Ölçek değişmezliği sağlanmazsa, gruplar arasında maddelerin yorumlanması açısından farklılık olduğu anlamına gelir ve bu da gruplar arası karşılaştırmalarda yanlış sonuçlara sebep olur. Katı değişmezlik, oldukça güçlü bir ölçme değişmezliği aşamasıdır. Faktör yükleri ve madde sabitlerine ek olarak madde artık varyanslarının da eşitliğini gerektirir (Reise ve Widaman, 1997). Katı değişmezliğin sağlanması diğer aşamalara göre daha zordur. Madde hata varyanslarının gruplar arasında değişmez olduğu varsayımına dayanır.

İkili Kategorik Değişkenlerde Ölçme Değişmezliği

Ölçme değişmezliği çalışmalarında yaygın olarak kullanılan yöntemlerin son zamanlarda bazı sınırlılıklarının olduğu yapılan araştırmalarla ortaya konulmuştur. ÇGDFA yöntemi daha önceki bölümlerde bahsedildiği gibi daha çok sürekli değişkenler ve likert tipi puanlanmış değişkenler için aşamalı olarak uygulanmaktadır. Raykov ve arkadaşları (2018), ikili puanlanmış maddelerde ölçme değişmezliğini test etmeyi sağlayan alternatif bir yöntem önermişlerdir. Bu yöntem, madde faktör yükleri ve eşik değerleriyle ölçme değişmezliğini sağlamayan durumları belirlemeyi amaçlamaktadır. Bir referans değişkeni belirlenmesini gerektirmeyen bu yeni yaklaşım, direkt bir ya da iki parametrelili madde tepki modellemesiyle değişen madde fonksiyonu çalışmaya imkân sağlamaktadır (Dimitrov, Li, Marcoulides, Menold ve Raykov, 2018).

Bir psikometrik ölçeğin bileşenlerini $y = (y_1, y_2, \dots, y_k)$ temsil etmektedir. Ayrıca “y” bileşeninin her biri büyük örneklere sahip gruplarda yapısal değişmezliğin koşullarını yerine getirdiği varsayılmaktadır (Millsap, 2011). Bu doğrultuda her bir grupta faktör yükleri ve eşik parametrelerinin değişken olduğu bir faktör analizi modeli geliştirilmiştir. Bu modele göre g. grubun y bileşeni ve ölçme değişmezliği için gerekli şartlar aşağıda verilmiştir.

$$y_g^* = \Lambda_g \eta_g + \delta_g \quad (2)$$

$$\Lambda_1 = \Lambda_2 = \dots = \Lambda_G \quad (3)$$

$$\tau_1 = \tau_2 = \dots = \tau_G \quad (4)$$

3 ve 4 numaralı eşitlik çifti ayrıca iki parametrelili madde tepki kuramı modeli (MTK) ya da MTK'nın bir özel durumu olan değişen madde fonksiyonunu (DMF) incelemek için yeterli bir şartı temsil etmektedir (Asparouhov, Muthén ve Morin, 2015).

Madde tepki kuramı, bireyin bir testte gösterdiği performansın örtük özellikler veya yetenekleri ve bir maddedeki performansı bu özelliklerin arasındaki ilişkiyi

tanımlayan madde karakteristik eğrisine göre kestirilebileceğini öne sürer (Hambelton ve Swaminathan, 1985). MTK, bireyin testten aldığı toplam puan yerine her bir maddeye verdiği cevabın doğru ya da yanlışlığı ile ilgilenir. Bireyin her bir maddeye vermiş olduğu doğru cevaplar “1”, yanlış cevaplar “0” olacak şekilde ikili madde (binary item) formunda puanlanır (Baker, 2016). MTK'nın temeli olan madde karakteristik eğrisini tanımlamak için iki madde parametresi kullanılır. Bunlardan biri “madde güçlüğü” (b) diğeri ise “madde ayırt edicilik” (b) indeksleridir. Madde güçlüğü, maddenin hangi noktada işlevsel olduğunu ifade eder. Örnek olarak; kolay bir madde daha düşük yetenek düzeyindeki bireyler için daha işlevsel iken, zor bir madde yüksek yetenek düzeyinde bulunan bireyler için daha işlevseldir. Madde ayırt edicilik indeksi ise, maddenin bulunduğu noktanın altında kalan bireyler ile bu noktanın üzerinde yetenek düzeyine sahip bireyleri ne kadar iyi ayırt edebildiğini gösterir (Baker, 2016).

MTK yaklaşımının özel bir durumu olan değişen madde fonksiyonu (DMF), bir test maddesine aynı yetenek düzeyinde olup farklı gruplardan gelen bireylerin maddeye doğru cevap verme olasılığının aynı olmaması durumunu ifade eder (Adams ve Rowe, 1988; Hambleton, Swaminathan ve Rogers, 1991). Dolayısıyla DMF her bir test maddesinin aynı evrenden oluşturulmuş farklı alt gruplarda benzer şekilde çalışıp çalışmadığı sorunu üzerinde durur (Akt. Başokçu ve Öğretmen, 2013). DMF analizleri temelde, test puanlarının farklı gruplardan gelen değişkenliklerden etkilenip etkilenmediğini ve bu değişkenliklerin herhangi bir alt grup için bir fayda sağlayıp sağlamadığını araştırmayı amaçlar (Algina ve Crocker, 1986). Test ile ölçülen özellik farklı alt gruplarda aynı ise, maddelerin aynı değişkenlik durumundan etkilendiği ve aynı yetenek düzeyindeki bireylerin ölçülen yapıda da benzer yetenek düzeyinde olduğu söylenebilir (Algina ve Crocker, 1986). Bu çalışmada kullanılan ikili maddelerde ölçme değişmezliği analizi yöntemi iki parametrelili MTK altında DMF gösteren maddeleri belirleme yolu ile ölçme değişmezliğini test etme imkânı sunmuştur.

Marcoulides, Millsap ve Raykov (2013) prosedüründe tahminlenemeyen ikili puanlanmış maddelerin ölçme değişmezliği daha önceki metoda paralel olarak kolayca elde edilmektedir (Millsap, 2011). Bu prosedürde Raykov ve arkadaşları ilk adım olarak; veri grubunun orijinal kategorik yapısını hesaba katarak, en büyük olabilirlik

(ML) kestirimi ile G grup modelini “k” ikili deęişkenler grubuna sabitlenmektedir (Muthén ve Muthén, 2016). Bu çok gruplu modelde, a) madde faktör yükleri ve eşik parametreleri grup aynılığı için sabitlenir; b) örtük deęişkenlerin ortalama ve varyansları sadece birinci grupta sırasıyla 0 ve 1 olarak belirlenir, ancak dięer gruplarda serbest kalır ve c) eęer varsa örtük kovaryans matrisinin köşegen dıőı öğeleri tüm gruplarda serbest bırakılır. Birinci aşamada denklem 3 ve 4’te ifade edildięi gibi, tüm faktörlerin ortalama ve varyansları birinci grupta 0 ve 1 olarak belirlenip, kalanı tüm gruplarda serbest bırakılır; bu sınırlı G grup model versiyonu M_0 olarak belirtilir (Dimitrov, Li, Marcoulides, Menold ve Raykov, 2018). İkinci aşamada her bir madde için madde faktör yükleri ve eşik deęerleri M_0 ’daki grup eşitliklerinden birer birer serbest bırakılır. Bu durum madde sayısının iki katı kadar farklı model oluşmasını gerektirir (Marcoulides, Millsap ve Raykov, 2013). Her bir modelin p deęeri ve ki-kare deęerleri M_0 serbest (nested) modelindeki p ve ki-kare deęeri ile karşılaştırılır. Üçüncü aşamada BH prosedürü p deęerleri grubuna uygulanır (Benjamini ve Hochberg, 1995). Elde edilen tüm modellerde yokluk hipotezlerinin hangisinin reddedileceęi belirlenir. Bunların $r=0$ olanları reddedilirse, ölçme deęişmezlięi için gerekli koşul veriyle tutarlı olarak kabul edilir ve analizi yapılan veri setinin ölçme deęişmezlięini sağlamadıęıyla ilgili yeterli bir kanıt içermedięi öne sürülür. Eęer $r>0$ hipotezi reddedilirse, ölçme deęişmezlięi sağlanamadıęı anlamına gelir. Bu yaklaşımda, 2 ve 3 eşitliklerinde ifade edilen sınırlamaların DMF incelenmesi için yeterli bir koşulu sağladığı ortaya çıkarılmıştır. Dolayısıyla bu yöntem, bir veya iki parametrelili DMF incelemek için doğrudan kullanılabilir (Dimitrov, Li, ve Marcoulides, Menold ve Raykov, 2018).

1.1 Problem Durumu

Sürekli değişen ve gelişen dünyada, eğitim bireylere bilgi öğretmenin yanı sıra; bu bilgileri yaşam durumlarıyla ilişkilendirip yeni durumlara uyarlama hedefleri doğrultusunda şekil almaktadır. Bu amaca yönelik yapılan her değişim eğitim programlarına, öğretim tekniklerine, ölçme değerlendirme araçlarına yansımaktadır. Ülkelerin eğitim politikalarında önemli rol oynayan pek çok uluslararası sınav yapılmaktadır. Bu sınavlardan biri de OECD (Organization for Economic Cooperation and Development) tarafından yürütülmektedir (MEB, 2016).

OECD dünya halklarının refahını, ekonomik kalkınmalarının ve eğitim politikalarının düzenlenmesinde önemli rol oynayan bir kuruluştur. Amaçları doğrultusunda pek çok çalışmalar yapmaktadır. Bu çalışmalardan biri de ülkemizin de katıldığı uluslararası boyutta uygulanan dünyanın en büyük eğitim araştırmalarından olan Uluslararası Öğrenci Değerlendirme Programıdır (PISA-Programme for International Student Assessment). PISA, 2000 yılından beri OECD ülkeleri vekatılımcı diğer ülkelerdeki 15 yaş grubu öğrencilerin günümüz dünyasında sahip olmaları gereken bilgi ve becerilere ne kadar sahip olduğunu ölçüp değerlendirmektedir. Bu araştırmanın en önemli amaçlarından birisi öğrencilerin bu yaş grubunda öğrenmesi gereken bilgileri ne ölçüde hatırlayabildiklerinden ziyade, günlük hayatta bu bilgileri ne kadar kullanabildiklerini belirlemektir (Eğitek, 2011). PISA sonuçları, ülkelerin eğitim seviyelerini yükseltmeleri ve standartlar oluşturmaları amacıyla kullanılmakta ayrıca dünya çapında öğrencilerin eğitim durumlarını karşılaştırmak için fırsat vermektedir. Bu hedefle bireylerin Dünya ekonomisine ne kadar katkıda bulunacağını da belirlemeyi hedefler. Dolayısıyla PISA sonuçları Türkiye'nin de katılımcı ülkeler arasında olduğu tüm ülkelerin eğitim politikalarına yansımaktadır.

PISA uygulamaları; matematik, fen ve okuma becerileri konu alanlarında her üç yılda bir düzenli olarak yapılmaktadır. Bu araştırmalar yapılırken ana kavram "okuryazarlık" üzerinde durulmaktadır. PISA'da okuryazarlık kavramı, yaşam pratikleri içinde bir işlevi yerine getirme amaçlı kullanılan bireysel bir donanım olarak ele alınır. Okuryazarlık toplumsal katılımı da içeren geniş bir kavramsal niteliğe dayanır. PISA okuma becerilerini günümüz şartları çerçevesinde tanımlamıştır. PISA'ya göre okuma "Bireylerin hedeflerine ulaşmak için, bilgi ve beceri

potansiyelini geliřtirmek ve topluma katılmak amacıyla yazılı metinleri anlaması, kullanması, yansıması ve onlarla ilgilenmesi”dir (OECD, 2009). PISA’nın yaklařımına gre okur, etkin anlamın oluřturulmasında n bilgilerini metne tařıyan, okumayı gerek yařam durumlarıyla iliřkilendiren bireydir. Bu yaklařıma gre okuryazarlık, toplumsal katılımı da gerektiren geniř bir ierięe dayanır.

Okuma becerileri PISA’nın ilk uygulaması olan 2000 yılı ve 2009 da aęırlıklı alan olarak deęerlendirilmiřtir. PISA’nın atıncı dngs olan PISA 2015’te aęırlıklı alan olarak fen okuryazarlıęı belirlendięi iin okuma becerileri alanında daha az sayıda (103) soru bulunmaktadır. Birbirlerinden baęımsız olmadıkları iin kesin gruplara ayırmak mmkn olmasa da, PISA okuma becerileri  boyutta incelenebilir: metin, okurun metne yaklařımı, metnin kullanım řekli (MEB, 2016). Arařtırmaya okuma becerileri testinin alt boyutları dahil edilmemiřtir.

Her PISA uygulamasında bir konu alanındaki okur yazarlık aęırlıklı alan olmaktadır. 2009 yılında aęırlıklı konu alanı okuma becerileri iken, 2012’de matematik okuryazarlıęı, en son yapılan uygulama 2015’te ise fen okuryazarlıęı aęırlıklı konu alanı olarak belirlenmiřtir. 2012’den bu yana ise temel alanlar dıřında yeniliki alan uygulaması yapılmaktadır. PISA 2015 uygulamasında yeniliki alan “iřbirliki problem özme” olmuřtur. PISA arařtırmalarında ęrencilerin ęrenme srecindeki psikolojik durumları, motivasyonları ve zyeterlikleri de hazırlanan anketler ile llmektedir. Bu da arařtırmanın nemli bir parasını oluřturur. Uluslararası dzeyde lkelerin eęitim ıktılarını deęerlendiren bu geniř aplı arařtırmada, eęitim seviyelerinin karřılařtırılabilmesi iin lme aracının geerlięinin saęlanması dolayısıyla eřdeęerlik inceleme alıřmaları son derece nemlidir. PISA’nın altıncı dngs olan PISA 2015 uygulamasında 35’i OECD lkesi olmak zere toplam 72 lke katılım gstermiřtir (OECD, 2015). PISA 2015 uygulamasında dięerlerinden farklı olarak ilk kez bilgisayar tabanlı deęerlendirme yapılmıřtır. Katılımcı lkelerden lkemizin de aralarında bulunduęu 57 lke bilgisayar tabanlı deęerlendirme, 15’i kaęıt-kalem tabanlı deęerlendirme ile uygulamaya katılmıřtır. Bu alıřmaya dahil edilen lkeler, sınavı bilgisayar tabanlı deęerlendirme ile uygulayan lkelerden seilmiřtir. Bilgisayar tabanlı deęerlendirme iin 66 farklı kitapık bulunmaktadır. ęrenciler her biri 60 dakika sren iki oturumda, aık ulu ve oktan semeli maddelerden oluřan kitapıkları yanıtlamıřlardır.

PISA soruları İngilizce olarak hazırlanır ve sorular ana dili İngilizce olmayan ülkelerin kendi dillerine çevirilerek kullanılır. Katılımcı ülkelerin bir çoğunun ana dili İngilizce değildir, dolayısıyla ana dili İngilizce olmayan ülkeler testleri kendi dillerine çevrilmiş şekilde kullanmaktadırlar. PISA, ülkelerin eğitim politikalarını önemli ölçüde etkilediğinden, ülkeler ve farklı gruplar arası ölçülen psikometrik yapının karşılaştırılabilir olması son derece önemlidir (Brown, 2006). Farklı kültür ve dilden katılım gösteren bireylerin farklı konu alanlarında, özellikle de okuma becerileri gibi direkt dile bağlı bir alanda anlamlı olarak karşılaştırılabilmesi için testlerin ölçtüğü yapılarda eşdeğerlik sorunu olmaması, testlerin ölçme değişmezliğinin olması gerekmektedir. Geçerliği ve güvenilirliği kanıtlanmış ölçme araçlarının da birden fazla grupta aynı yapıda olup olmadığı, bir diğer deyişle gruplar arası karşılaştırmaların anlamlılığı ölçme değişmezliği analizleriyle test edilmelidir. Eğitimde ölçme ve değerlendirme çalışmalarında ölçme değişmezliği, özellikle kültürlerarası psikolojik yapıların karşılaştırılabilirlik ve genellenebilirliklerini araştırmak için kullanılmaktadır. Günümüzde eğitim ihtiyaçları doğrultusunda, uluslararası ölçme ve değerlendirmeye verilen önem artmaktadır. Dolayısıyla kültürlerarası uygulanacak olan başarı ve kişilik testi gibi ölçeklerin çeşitli dillere çevrilmesi söz konusu olmaktadır (Hambleton, 2005). Farklı kültürlerdeki bireyleri uluslararası ölçme-değerlendirme araçlarıyla karşılaştırmak için, ölçme aracı başka dillere çevrildiğinde farklı dillerdeki formlarının eş değerliğinin sağlanması önemlidir. Bu nedenle çoklu grupların kültürler arası karşılaştırmalarında ölçme değişmezliği en çok ihtiyaç duyulan çalışmadır. Alanyazında PISA uygulamalarının ölçme değişmezliğine ilişkin çeşitli çalışmalar mevcuttur.

Bu çalışmada ana dili İngilizce olan ülkeler ile ana dili İngilizce olmayan ülkelere oluşturulan örnekleme PISA 2015 okuma becerileri testlerinin ölçme değişmezliğinin test edilmesi için aşağıdaki problem cümlesi ve alt problem cümlelerine yanıt aranmıştır.

1.2 Problem Cümlesi

PISA 2015 okuma becerileri testinin ana dili İngilizce olan ülkeler ile ana dili İngilizce olmayan ülkelere göre ölçme değişmezliği var mıdır?

Alt Problemler

- 1) Ana dili İngilizce olan ülkeler ve ana dili İngilizce olmayan ülkelere oluşturulan gruplar arasında faktöriyel yapı geçerliliği sağlanmakta mıdır?
- 2) Ana dili İngilizce olan ülkelere oluşan grup ile ana dili İngilizce olmayan ülkelere oluşan grupların a, b parametreleri ve madde faktör yükleri farklılıklar göstermekte midir?
- 3) Ana dili İngilizce olan ülkelere oluşan grup ile ana dili İngilizce olmayan ülkelere oluşan gruplar maddeler bazında karşılaştırılabilir midir?

1.3 Araştırmanın Amacı ve Önemi

Bu araştırmanın genel amacı, PISA 2015 okuma becerileri alt testinin ana dili İngilizce olan ülkeler ile ana dili İngilizce olmayan ülkelere için dil becerileri açısından eş değerliğinin sağlanıp sağlanmadığının incelenmesidir.

PISA uygulamaları dünya çapında katılımcı ülkelerin eğitim sistemlerinin güçlü ve zayıf yönlerini tespit etmeyi amaçlar (OECD, 2016). Bu doğrultuda farklı kültür ve dillere sahip katılımcı ülkelerin bilgi ve beceri düzeyleri birbiriyle karşılaştırılır. PISA uygulamalarının orijinal dili İngilizce'dir. Dolayısıyla ana dili İngilizce olan ülkelere testleri orijinal formlarda, diğer ülkelere katılan öğrenciler ise testleri çeviri dil formlarında cevaplamaktadır. Karşılaştırma ve değerlendirmelerin anlamlı olabilmesi için, kültürler ve diller arası ölçme eşdeğerliğinin olması, yani ana dili İngilizce olan ülkelere için herhangi bir yanlılık olmaması gerekmektedir.

PISA uygulamalarında ölçme araçları başka dillere çevirilirken olası çeviri hatalarının belirlenmesi ve düzeltme önerileri için uzman çevirmenlerde kontrol listeleri bulunmaktadır. Bu uyarılama işlemi yapıldıktan sonra psikometrik özellikleri belirlenmektedir (OECD, 2010). Belli bir kültürde ve dilde geliştirilmiş ölçekler, o

kültüre ve dile özgü nitelik ve kavramsallaştırmaları yansıtır. Bir ölçme aracının bir dilden başka bir dile tercüme edilmesi bu iki ölçeğin eş değer olduğunun garantisini vermez (Berberoğlu ve Sireci, 2000). Başka bir dile çevirilecek ya da uyarlanacak ölçme araçlarının orijinal formlarından farklı olacakları bilinmelidir. Söz konusu farklılıkların psikometrik, dil ve anlamlılık açılarından kabul edilebilir düzeyde olmaları sağlanmalıdır (De Jong ve Hambleton, 2003). Bu durumda kültürler arası bir karşılaştırmada, ölçme değişmezliği varsayımından yola çıkarak geçerlik kanıtlarının sunulması kaçınılmaz bir gereklilik olmaktadır. Ülkelerin eğitim politikalarında büyük rol oynayan böyle bir araştırmada testlerin dil becerileri açısından kültürler arası eş değerliğinin sağlanmış olması, doğru kararlar alınmasının ön koşullarından biridir. Tüm bu gereklilikler ışığında yapı geçerliği çalışmalarının titizlikle yapılması, ölçme aracının geçerlik kanıtı için son derece önemlidir. Bu nedenlerle bu çalışmada PISA 2015 uygulamasının okuma becerileri testinin çeviri dil ve orijinal dil arasında herhangi bir ölçme değişmezliği sorunu olup olmadığını istatistiksel analiz yöntemleriyle incelenmiştir.

1.4 Sınırlılıklar

Yapılan araştırmanın kapsamında PISA 2015 uygulamasında bilgisayar tabanlı değerlendirme yapılan ülkelerden altısı bulunmakta olup hepsinde ortak olarak uygulanan 41 numaralı form seçilmiştir. Seçilen bu formda okuma becerileri alanında 28 madde ele alınmıştır. Araştırma PISA 2015 değerlendirmesine katılan ülkelere kapsam doğrultusunda seçilen altıülke ve bunlardan 41 numaralı formu alan 1526 katılımcı ile formda yer alan 28 madde ile sınırlıdır. Bu araştırma için belirlenen İngiltere, Kanada, ABD, Japonya, Tayland ve Türkiye örneklemelerinden elde edilen verilerin tüm evreni temsil ettiği varsayılmıştır.

PISA 2015 araştırmasında ilk defa bilgisayar tabanlı uygulama yapılmıştır. Katılım gösteren 72 ülkeden 57'si uygulamayı bilgisayar tabanlı, 15'i ise kağıt-kalem tabanlı olarak değerlendirmeye dahil olmuştur. İki farklı değerlendirme yöntemi arasında farklılık olmadığı varsayılmıştır. Bilgisayar tabanlı değerlendirmede 66 farklı kitapçık kullanılmıştır, bu kitapçıklar aynı sorular olduğu gibi farklı soruları da içermektedir. Kitapçık türlerinin eşdeğer kazanımlarda sorular içerdiği kabul edilmiştir.

PISA 2015'e kadar olan uygulamalarda kâğıt-kalem tabanlı değerlendirme yapılmış olup, PISA 2015 uygulamasına katılan birçok ülkede bilgisayar tabanlı değerlendirme yapılmıştır. Okuma becerileri alanı uzun metinlerden oluşan sorular içerdiğinden bilgisayar tabanlı değerlendirme yönteminin ülke sonuçlarını etkileyip etkilemeyeceği sonuçları değerlendirme sürecinde dikkate alınmalıdır. Okuma becerileri alanında yer alan sorulara ilişkin uzun metinlerin olmasından dolayı uygulama yönteminin öğrencilerin sahip olduğu bilgisayar okuryazarlığı veya becerisinden dolayı ülke sonuçlarına etkisinin olup olmayacağı ülke sonuçlarını değerlendirirken göz önünde bulundurulmalıdır. Bu iki yöntem arasındaki farklılıkların da ayrıca ele alınması gerekecektir (MEB, 2016).

1.5 Konu İlgili Ulusal ve Uluslararası Düzeyde Yapılmış Çalışmalar

Bu kısımda öncelikle ulusal düzeyde uygulanan sınavların ölçme değişmezliği ile ilgili çalışmalara ardından uluslararası düzeyde yapılmış olan ölçme değişmezliği çalışmalarına yer verilmiştir.

Ulusal ve uluslararası uygulanan sınavlarda bir çok ölçme aracının geçerliğini test etmek amacıyla ölçme değişmezliği çalışmaları yapılmıştır. Ulusal kapsamda yapılan çalışmalarda daha çok demografik özelliklere göre karşılaştırmalar yapılmıştır (Akt. Kıbrıslıoğlu, 2015).

Uluslararası sınavların ölçme değişmezliği çalışmaları dünyada olduğu gibi Türkiye'de de yapılmıştır. Öğretmen (2006), ülkemizin de katıldığı PIRLS 2001 kapsamında uygulanan okuma parçaları testlerinin psikometrik özelliklerini ABD ve Türkiye örneklemelerinde karşılaştırmalı olarak incelemiştir. Araştırma iki aşamada gerçekleştirilmiş olup, birinci aşamada okuma parçaları testlerinin ölçtüğü düşünülen yapıların kültürlere göre eşdeğer olup olmadığını ÇGDFA yöntemi ile test edilmiştir. İkinci aşamada ise test maddelerinin kültürlere göre DMF içerip içermediği MTK bağlamında parametreleri karşılaştırarak ve olabilirlik oran testi karşılaştırma yöntemleri kullanılarak araştırılmıştır. ÇGDFA sonuçlarına göre bu testlerin yapılarının kültürler arası bir eşdeğerliğinin olmadığı, parametre karşılaştırma ve olabilirlik oran testi ile yapılan analizler sonucunda maddelerinin çoğunun DMF içerdiği gözlenmiştir.

Akyıldız (2009), PIRLS 2001 başarı testinin 35 ülke arasında faktör yapılarının birbirine denkliğini ÇGDFA ile incelemiştir. Faktör yapıları çalışmaya dahil edilen tüm ülkelerde uygulanmış olan 4 ayrı alt ölçek kapsamında ayrı ayrı analiz edilmiştir. Ulaşılan bulgulara göre faktör yapılarının araştırmaya katılan tüm ülkelerde orta düzeyde denkliği sağladığı belirtilmiştir.

Öğretmen ve Uzun (2010), TIMSS-R 1999 uygulamasında öğrencilerin fen başarısına etki eden duyuşsal değişkenlerin ve Türkiye örnekleminde cinsiyetin gruplar arasında ölçme değişmezliği şartını sağlayıp sağlamadığını araştırmışlardır. Modelde yer alan öz yeterlik, tutum, önem ve sınıf içi öğrenme etkinlikleri gibi alt boyutlar yapısal eşitlik modellemesi yöntemleri ile analiz edilmiştir. Araştırma sonuçlarına göre, tüm değişkenlerin metrik değişmezliği sağladığı fakat cinsiyet gruplarının katı değişmezliği sağlayamadığı görülmüştür.

Asil ve Gelbal (2012), PISA 2006 kapsamında uygulanan öğrenci anketinin kültürel ve dil bakımından ölçme değişmezliğini incelemişlerdir. Çalışmaya Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye örneklemi dahil edilmiştir. Çok gruplu doğrulayıcı faktör analizi yöntemiyle öğrenci anketi maddelerinin kültürler ve dillere göre değişen madde fonksiyonu (DMF) gösterip göstermediği araştırılmıştır. Ulaşılan sonuçlara göre, ülkeler arasında kültürel ve dilsel açıdan farklılıklar arttıkça doğru orantılı olarak değişen madde fonksiyonu gösteren madde sayısının da arttığı ortaya çıkmıştır.

Doğan ve Uyar (2014), PISA 2009 Türkiye örnekleminde öğrenci anketinde yer alan öğrenme stratejileri modelinin okul türü, cinsiyet ve bölgelere ölçme değişmezliğini test etmiştir. Değişmezlik analizleri aşamalı olarak yapılmıştır ve çok gruplu doğrulayıcı faktör analizi yöntemi kullanılmıştır. Bulgulara göre öğrenme stratejileri modelinin cinsiyet ve okul türü gruplarında yalnızca yapısal ve metrik değişmezliği sağladığı, istatistiksel bölgelere göre ise tam değişmezlik şartlarını sağladığı sonucuna ulaşılmıştır.

Kıbrısoğlu (2015), PISA 2012 değerlendirmesinde matematik öğrenme modelinin ülkeler cinsiyet grupları arasındaki ölçme değişmezliğini incelemiştir. Araştırmaya Türkiye, Çin-Şangay ve Endonezya örneklemi dahil edilmiştir. Matematik öğrenme modeli 55 madde ve dokuz faktör ile açımlayıcı faktör analizi kullanılarak doğrulanmış, sonrasında çok gruplu doğrulayıcı faktör analizi yapılmıştır.

Ölçme değişmezliği aşamalı olarak test edildikten sonra, modelin ülkeler arasında yalnızca şekil değişmezliğini sağladığı; cinsiyet grupları arasında ise ölçme değişmezliğinin tüm aşamalarını sağladığı sonucu elde edilmiştir. Dolayısıyla matematik öğrenme modelinin ülkeler arası karşılaştırmalarının yanıtıcı olabileceği, cinsiyet grupları bazında ise karşılaştırılabilir olduğu yargısına varılmıştır.

Başusta ve Gelbal (2015), çalışmalarında PISA 2009 öğrenci anketinde bulunan fen bilimleri ile ilgili maddelere ait faktörleri belirlemeyi ve Türkiye örneğinde bu faktörlerin cinsiyet açısından ölçme değişmezliğini test etmeyi amaçlamışlardır. Yapısal eşitlik modellemesi ile veriler analiz edilmiştir. Ölçme değişmezliği de aşamalı olarak çok gruplu doğrulayıcı faktör analizi ile test edilmiştir. Bulgulara göre, cinsiyet bazında grupların ölçme değişmezliğini sağladığı neticesine ulaşılmıştır.

Ayvallı (2016), PISA 2012 uygulamasında matematik okuryazarlığının OECD üyesi ülkeler ve Türkiye örneğinde bölgeler ve cinsiyet grupları arasında ölçme değişmezliğini incelemeyi amaçlamıştır. Bu doğrultuda öncelikle bölgeler ve cinsiyet grupları ayrı ayrı doğrulayıcı faktör analizi yöntemi ile doğrulanmıştır. Bir sonraki aşamada çok gruplu doğrulayıcı faktör analizi ile ölçme değişmezliği aşamalı olarak test edilmiştir. Sonuçlara göre, OECD ülkeleri için ölçme değişmezliğinin sağlanamadığı; sadece güçlü faktöriyel değişmezlik şartını yerine getirdiği, Türkiye örneğinde ise cinsiyet ve gruplar bazında ölçme değişmezliğinin sağlandığı tespit edilmiştir.

Çetin ve Ölçülüoğlu (2016), TIMSS 2011 uygulamasının Türkiye örneğinde 8. sınıf öğrencilerinin matematik başarısını etkileyen faktörler modellenmiş ve modelin coğrafi bölgelere göre ölçme değişmezliği incelenmiştir. Açıklayıcı faktör analizi ile (AFA) öğrenci anketinde matematik başarısını etkileyen maddeler belirlenmiş, daha sonra matematik başarısı ile modeldeki değişkenler arasındaki ilişkiyi incelemek için yapısal eşitlik modellemesi ile bir model oluşturulmuştur. Bir sonraki aşamada modelin bölgeler arası değişmezliğini test etmek için çok gruplu doğrulayıcı faktör analizi yapılmıştır. Ulaşılan bulgulara göre, bölgeler bazında yalnızca yapısal ve metrik değişmezlik şartlarının yerine sağlandığı ve sadece bu aşamalarda yapılan karşılaştırmalarda gruplar arası farklılıkların ölçme aracından kaynaklanamayacağı sonucuna varılmıştır.

Dolan, Hessen, Wicherts ve arkadaşları (2004) standartlaştırılmış zeka testlerinin ölçme değişmezliğini incelemişlerdir. Araştırmanın amacı beş zeka testinin grup açısından ölçme değişmezliğini test etmektir. Wechsler Yetişkin Zeka Ölçeği(WAIS), Dutch Kademeli Yetenek Testi (DAT) ve Dutch Çocuk Zeka Testi'nin (RAKIT) yıllara göre ölçme değişmezliğini ÇGDFA yöntemi ile incelemişlerdir. Analiz sonuçları ölçme değişmezliğinin sağlanmadığını açıkça göstermiştir.

Glanville ve Wildhagen (2007) yaptıkları araştırmada, ulusal boylamsal eğitim çalışmaları (NELS, 1988) kapsamında kullanılmış okula katılım ölçeğini değerlendirmeyi ve sistematik olarak geliştirilmiş bu ölçeğin ırk-etnik gruplar açısından ölçme değişmezliğini incelemeyi amaçlamışlardır. Doğrulayıcı faktör analizlerinden elde edilen sonuçlar, okul katılımının çok boyutlu bir kavram olarak ölçülmesi gerektiğini göstermiştir. Bir dizi çoklu grup analizinden elde edilen bulgular ise ölçeğin Beyaz, Afrika kökenli Amerikalı, Latin ve Asyalı öğrenciler için ölçme değişmezliği sağladığını ortaya koymuştur.

Uluslararası geniş çaplı sınavlardan biri olan TIMMS (Third International Mathematics and Science Study), kültürlerarası güçlü bir bakış açısına sahip olduğundan kültürlerarası ölçme değişmezliğini test etmek önemlidir (Hau ve Marsh , 2004). Ercikan ve Koh (2005), 1995 yılında uygulanan TIMSS verilerini kullanarak İngilizce ve Fransızca versiyonlarının ölçme değişmezliğini araştırmışlardır. Araştırmaya Kanada'dan katılan öğrenciler dahil edilmiştir. Madde tepki kuramı ve çok gruplu doğrulayıcı faktör analizi ile karşılaştırmalar yapılmış; matematik ve fen sınavlarının İngilizce ve Fransızca dilindeki versiyonlarında önemli farklılıklar olduğu sonucuna ulaşılmıştır. Değişen madde fonksiyonu analizleri ile elde edilen yüzdelerle iki karşılaştırma grubu arasında büyük farklılıklar olduğu ortaya çıkmıştır. Analizlerde gözlemlenen yapılarıdaki farklılıklar, TIMSS sonuçlarını matematik ve fen alanındaki genel performansları karşılaştırmak için kullanmada ciddi sınırlamaların olduğunu göstermiştir.

Marsh ve arkadaşları (2006), PISA 2000 uygulamasındaki öğrenme yaklaşımları ölçeğinden elde edilen verilerle ölçme değişmezliği çalışması yapmışlardır. 25 ülkeden 4000 kişilik bir örneklem araştırmaya dahil edilmiştir. Kültürlerarası ölçme değişmezliğini test etmek amacıyla ÇGDFA yöntemi

kullanılmıştır. Elde edilen bulgulara göre öğrenme yaklaşımları ölçeğinin kültürlerarası eş değeri sağladığı sonucuna ulaşılmıştır.

Li, Wu ve Zumbo (2007), TIMSS 1999 uygulamasındaki matematik testlerinin ülkeler bazında ölçme değişmezliğini incelemişlerdir. Araştırma için 7 ülke seçilmiştir. Bu ülkeler benzer kültürel özelliklere göre iki gruba ayrılmış ve çok gruplu doğrulayıcı faktör analizi yöntemi ile ölçme değişmezliği analizleri yapılmıştır. Ayrıca benzer ve farklı kültürlere sahip ülkeler arasında ikili karşılaştırmalar da yapılarak bulgulara ulaşılmıştır. Elde edilen veriler ile benzer kültüre sahip ülkeler arasında matematik testlerinin ölçme değişmezliği sorunu olmadığı, farklı kültürler arasında ise tam olarak ölçme değişmezliğinin sağlanamadığı sonucuna varılmıştır.

Abduljabbar, Ebu-Hilal, Marsh ve arkadaşları (2013) TIMSS 2007 değerlendirmesinin matematik ve fen bilimleri motivasyon ölçeklerinde faktör, yöntem etkileri, cinsiyet farklılıkları ile yakınsak ve diskriminant geçerliği incelemişlerdir. Araştırmaya TIMSS 2007 uygulamasına katılmış ülkelerden 4'ü Arapça konuşan (Umman, Mısır ve Ürdün, Suudi Arabistan) 4'ü İngilizce konuşan (ABD, Avusturya, İngiltere ve İskoçya) olmak üzere 8 ülke dahil edilmiştir. Çalışmadan elde edilen verilere göre matematik ve fen bilimleri motivasyon ölçeğinin cinsiyetler bazında ölçme değişmezliğinin sağladığı fakat ülkeler arası değişmezliği sağlayamadığı sonucu ortaya çıkmıştır.

Segeritz ve Pant (2013) PISA 2003 Almanya örneğinde Öğrencilerin Öğrenme Yaklaşımları (SAL) ölçeğinin bir ülkedeki etnik-kültürel gruplar arasındaki ölçme değişmezliğini incelemişlerdir. Araştırmanın amacı doğrultusunda Almanya oldukça büyük bir göçmen nüfusuna sahip olduğundan, kendisi veya ebeveynleri SSBC veya Türkiye'de doğmuş göçmen kökenli öğrencileri ve ayrıca göçmen kökenli olmayanları da çalışmalarına dahil etmişlerdir. Faktör yapıları her bir öğrenci grubu için doğrulayıcı faktör analizi ile doğrulanmış, ölçme değişmezliğinin aşamaları test edilmiştir. ÇGDFA yöntemi kullanılmıştır. Analiz sonuçları ile elde edilen bulgular, Öğrenme Yaklaşımları ölçeğinin faktör yapısının Almanya ve iki göçmen kökenli öğrenci grubu arasında karşılaştırılabilir olduğunu göstermiştir.

Buerger, Goldhammer ve Kroehne (2016), PISA 2015 uygulamasında bilgisayar tabanlı testler ile kâğıt-kalem tabanlı testler arasındaki eşdeğerlik kriterlerini ele alarak yöntem etkileri çalışması yapmışlardır. Bu amaçla maddeler üzerindeki

yöntem etkilerini analiz etmek için çok gruplu MTK modelini kullanmışlardır. Elde edilen bulgulara göre, olası yöntem etki kaynakları tam olarak saptanamadığı için bilgisayar tabanlı testlerde zaman kullanımı ile ilgili daha çok analiz yapılması gerektiği sonucuna ulaşılmıştır.

Raykov ve arkadaşları (2018), ölçme değişmezliği çalışmalarında sürekli ve sıralı kategorik değişkenlerden oluşan veri setleri için uygulanan fakat ikili puanlanmış maddelerin ölçme değişmezliğini test etmek için uygun olmayan yöntemlere alternatif bir yöntem sunmuşlardır. Araştırma 9 sorudan oluşan bir matematik yetenek testinin 771'i erkek, 744'ü kız olan 1515 kişilik bir örnekleme uygulanmasından elde edilen veriler üzerinden yürütülmüştür. Her bir madde için eşik değerler ve faktör yükleri serbest bırakılarak, ÇGDFA yapılmıştır. Referans model olarak belirlenen serbest modele (M_0) göre değişen madde fonksiyonu incelenmiştir. İkili puanlanmış (kategorik) maddeler içeren ölçeklerin ölçme değişmezliği analizleri, sürekli değişkenlerde olduğu gibi referans değişkeni seçmeye gerek olmaksızın, MTK temellerine dayanan bir yöntem ön görülmüştür. Bu yöntem araştırmacıların ölçme değişmezliğini DMF inceleme yoluyla test etmesine imkân vermektedir.

BÖLÜM II

YÖNTEM

Bu bölümde araştırmanın modeli, çalışma grubu, verilerin elde edilmesi ve verilerin analizi ele alınmıştır.

2.1 Araştırmanın Modeli

Araştırmanın genel amacı, PISA 2015 okuma becerileri alt testinin dil becerilerine göre testi orijinal formda alan ülkeler ile çeviri formunu alan ülkeler arasında eşdeğerliğinin sağlanıp sağlanmadığını belirlemektir. PISA 2015 uygulamasının mevcut verileriyle geçerlik belirleme çalışması yapıldığı için bu yönüyle betimsel bir araştırma modelidir.

2.2 Çalışma Grubu

PISA'nın altıncı döngüsü olan PISA 2015 uygulamasının evrenini 15 yaş grubu öğrenciler oluşturmaktadır. 35 OECD ülkesi olmak üzere toplam 72 ülke katılım göstermiş, toplamda 29 milyon öğrenciyi temsilen 540.000'e yakın öğrenciye uygulanmıştır. Bu çalışmanın örneklemini İngiltere, Kanada, ABD, Japonya, Tayland ve Türkiye'den katılan öğrenciler oluşturmaktadır. Ülkelerin seçimi yapılırken ana dilleri ve okuma becerileri testindeki başarı seviyeleri göz önünde bulundurulmuştur. Araştırmaya dahil edilen altı ülkeden üçü testi orijinal formda alan diğer üçü ise ana dili İngilizce olmayıp testi çeviri formda alan ülkeler arasından seçilmiştir. Kanada ve Japonya okuma becerileri testinde başarılı olan ülkeler arasında olduğu için araştırmaya dahil edilmiştir. İngiltere ve ABD okuma becerileri testinde OECD ortalamasına yakın başarı sağlayan ve Tayland ile Türkiye de başarı sıralamasında OECD ortalamasının altında yer alan ülkelerdir. PISA 2015 uygulamasına İngiltere'den 14.157, ABD'den 5712, Kanada'dan 20.058, Japonya'dan 6647, Tayland'dan 8249 ve Türkiye'den 5.895 öğrenci katılmıştır. PISA uygulamalarında tüm öğrencilere aynı test uygulanmaz, test formları ortak sorular olduğu gibi farklı sorular da içermektedir (OECD, 2016). Bu çalışmada, bilgisayar tabanlı değerlendirmede uygulanan 66 farklı form arasından, araştırmanın örneklemine dahil

edilen ülkeler bazında en çok cevaplanan form olduğu için 41 numaralı form seçilmiştir. Dolayısıyla araştırmanın örneklemini PISA 2015 uygulamasında 41 numaralı formu alan öğrenciler oluşturmaktadır. Okuma becerileri başarıları bu formda 29 madde ile ölçülmüştür. Bunların arasından “CR404Q07S” kodlu madde, Tayland’dan katılan öğrenciler tarafından uygulanabilir olmadığı gerekçesiyle araştırmaya dahil edilmemiştir. Örnekleme 41 numaralı formu alan öğrencilerin ülkeler bazında frekansları Tablo 1’de verilmiştir.

Tablo 1. Ülkeler Bazında Örneklem Frekansları.

ÜLKE	N	%
Kanada	524	34.4
İngiltere	384	25.2
Tayland	176	11.5
Japonya	145	9.5
Türkiye	159	10.4
ABD	136	8.9
Toplam	1524	100

Tablo 1 incelendiğinde, Kanada’da 41 numaralı formu alan 524, İngiltere’de 384, Tayland’da 176, Japonya’da 145, ABD’de 136 ve Türkiye örnekleminde 159 öğrenci olduğu görülmektedir. Katılım oranlarına bakıldığında araştırmanın örnekleminde en fazla katılımcıya sahip ülke % 34.4 ile Kanada ve en az katılımcıya sahip ülke ise %8.9 oranı ile ABD olduğu anlaşılmaktadır. Diğer ülkelerin katılım oranlarında %25.2 ile %9.5 arasında değişim göstermektedir. Araştırmanın örneklemini PISA 2015 uygulamasına katılan ülkelere seçilmiş altı ülkeden 41 numaralı formu alan toplam 1524 öğrenci oluşturmaktadır.

2.3 Verilerin Elde Edilmesi

PISA 2015 uygulaması ilk kez bilgisayar tabanlı değerlendirme ile yapılmıştır. Bilgisayar tabanlı değerlendirmeyi seçen ülkelerde 66 farklı form, kağıt-kalem tabanlı değerlendirmeyi seçenlere ise 30 farklı form uygulanmıştır. Kitapçıkların her biri 30 dakika süreli 4 bölümden oluşmaktadır. PISA-2015 uygulaması her biri 60 dakika süren iki oturumda yapılmıştır. Öğrencilerin toplam puanı testlerin içeriğindeki ortak maddeler üzerinden hesaplanmıştır (MEB, 2016). Araştırma için tüm ülkelerde uygulanmış olan rastgele seçilecek bir kitapçığın maddelerinden elde edilen verilere OECD'nin resmi web sitesinin PISA ile ilgili bölümünden <http://www.oecd.org/pisa/data/2015database/> adresinden indirilmiştir.

Araştırmaya dahil edilen ülkeler PISA 2015 uygulamasına bilgisayar tabanlı değerlendirme ile katılan ülkeler arasından seçilmiştir. Dolayısıyla 66 farklı form arasından tüm ülkelerde uygulanmış olan ve en çok yanıtlanan 41 numaralı formdaki 28 madde analize dahil edilmiştir. Bu formda açık uçlu ve çoktan seçmeli maddeler bulunmaktadır. Soru çeşidine göre maddeler yanlış yanıtlar "0", kısmi doğru yanıtlar "1" doğru yanıtlar ise "2" ile kodlanmıştır. Diğer cevap koduna sahip yanıtlar ise (5,6,7,8,9) PISA 2015'in kod kitabında yer aldığı şekilde kayıp veri koduyla analize dahil edilmiştir. Verilerin analizi aşamasında modelin tahminlenememesinden dolayı 5. ve 6. maddelerde "2" kodlu cevaplar "1" olacak şekilde düzenlenmiş, veri seti "1-0" ikili kategorik forma dönüştürülmüştür.

2.4 Verilerin Analizi

Veriler düzenlendikten sonra kayıp verileri silme işlemi yapılarak veri seti analizlere uygun hale getirilmiştir. Pek çok araştırmada toplanan verilerde eksikliklerle karşılaşmaktadır. Bu kayıp değerler çoğu araştırmacının karşılaştığı önemli bir problemdir (Çokluk ve Kayri, 2011). Kayıp veri ile karşılaşılan durumlarla baş etmek için pek çok farklı yöntem bulunmaktadır. Kayıp değerler araştırma sonuçlarını farklı şekilde etkilediği için, araştırmacının örneklem büyüklüğü, kayıp değer oranı gibi etmenleri göze alarak bir karar vermesi gerekmektedir (Cheema, 2012). Kayıp veri sorunu ile ilgili çözüm yolları; kayıp veri ile analize devam etme, eksik verileri analiz dışında bırakma, kayıp veri yerine değer atama ve diğer istatistiksel metodlarla eksik verinin tamamlanması gibi uygulamalardır (Downey ve

King, 1998). Bu arařtırmada rneklem byklğne oranla kayıp deęerlerin arařtırmanın sonucunu nemli lde etkilemeyeceęi dřnldę iin eksik verilerin analiz dıřı bırakılmasına karar verilmiřtir. Kayıp deęerler veri setinden ıkarılmadan nce rneklem sayısı 1625 iken kayıp veriler ıkarıldıktan sonra analize dahil edilen rneklem sayısı 1524 olmuřtur.

Bu alıřmada ncelikle model her bir grup iin yapılan DFA ile doęrulanmıř, yapı geerlilięine dair kanıt elde edilmiřtir. Sonra dil deęiřkeni aısından lkeler arası lme deęiřmezlięini test etmek iin ařamalı olarak GDFA yntemi kullanılmıřtır. Arařtırmaya dahil edilen lkeler ana dili İngilizce olanlar (İngiltere, Kanada, ABD) ve ana dili İngilizce olmayanlar (Japonya, Tayland ve Trkiye) olarak iki gruba ayrılmıřtır. Bu iki gruptan oluřan veri setinde GDFA yapılmıřtır. Ana dili İngilizce olmayan lkelerden oluřan veri grubu referans grup olarak belirlenmiřtir. GDFA alıřmalarında her bir faktr iin deęiřkenlerden bir tanesi de sabit tutularak referans deęiřken olarak belirlenmektedir. Referans deęiřkene gre dięer deęiřkenlerin nasıl deęiřtięi gzlenmektedir. Referans deęiřkenin belirlenmesi iin her bir grup iin yapılan aımlayıcı faktr analizi yapılmıř, madde faktr ykleri incelenmiřtir. En ok faktr yk veren maddenin referans deęiřken olarak belirlenmesi tavsiye edilmektedir (Bykztrk, okluk ve řekercioęlu, 2010). Aımlayıcı faktr analizi sonularına gre her bir faktrde en ok yk veren (DR406Q01C) madde referans deęiřken olarak belirlenmiřtir. Analizler Mplus 8.0 ve R programları kullanılarak yapılmıřtır. lme deęiřmezlięi ařamalarını test etmek iin yapılan GDFA sonularına gre, veri setinde 5 ve 6. Soruların 0-1-2 kategorik yapısından dolayı modelin tahminlenemedięi grlmřtir. Dolayısıyla kategorik yapıdaki bu veri setinde lme deęiřmezlięi ařamalı olarak test edilememiřtir. Son zamanlarda yapılan lme deęiřmezlięi alıřmalarında ikili kategorik deęiřkenli veri setleri iin madde tepki kuramına gre lme deęiřmezlięi analizi yntemi yaygın olarak kullanılmıřtır ve yeni bir yaklařım olarak n grlmřtir (Dimitrov, Li, Marcoulides, Menold ve Raykov 2018). Tm veri setinde “2” ile kodlanmış maddeler “1” olacak řekilde tekrar dzenlenerek “1-0” kategorik forma dnřtrlmřtir. Mplus 8.0 programı ile veri setinin kategorik yapısı dikkate alınarak, iki deęiřkenli puanlanmış maddelerde GDFA yntemi kullanılarak lme deęiřmezlięi analizi yapılmıřtır. Yapılan analizlerle elde edilen sonulara Bulgular blmnde yer verilmiřtir

BÖLÜM III

BULGULAR

Bu bölümde araştırmanın amaçları doğrultusunda düzenlenen veri seti için PISA 2015 okuma becerileri testinin ana dili İngilizce olan ve ana dili İngilizce olmayan ülkelere göre dil değişkeni açısından söz konusu gruplara ait test istatistiklerine yer verilmiştir. Araştırmanın bulguları alt problemlere göre rapor edilmiştir.

3.1 Araştırmanın 1. Alt Problemine Yönelik Bulgular

Araştırmanın birinci alt problemine ilişkin “Ana dili İngilizce olan ülkeler ve ana dili İngilizce olmayan ülkelere oluşturulan gruplar arasında faktöriyel yapı geçerliliği sağlanmakta mıdır?” sorusuna cevap aranmaktadır. Bu bağlamda öncelikle her bir grup için ikili değişkenlerin yapısına uygun olarak DFA tamamlanmış, faktör yükleri belirlenmiş ve model uyumu incelenmiştir. DFA ile elde edilen bulgular Tablo 2’de verilmiştir.

Tablo 2. PISA Okuma Becerileri Testine İlişkin Doğrulayıcı Faktör Analizi Sonuçları

Grup (Ülkeler)	P	χ^2/sd	RMSEA	CFI	TLI
Anadili İngilizce Olan	.00	1,81	.03	.96	.97
Anadili İngilizce Olmayan	.00	1.40	.03	.97	.98

Tablo 2’deki model uyum indeksleri incelendiğinde ki-kare değerinin iki grupta da anlamlı çıktığı görülmektedir ($p < .05$). Ki-kare istatistiği örneklem büyüklüğünden önemli ölçüde etkilendiği için genellikle anlamlı çıkmaktadır (Kline, Principles and Practices of Structural Equation Modelling., 2011). χ^2/sd değerinin ise ana dili İngilizce olan grup ($\chi^2/sd=1.81$) ve ana dili İngilizce olmayan grup ($\chi^2/sd=1.40$) iyi uyum aralığında ($\chi^2/sd \leq 2$) olduğu tablodan anlaşılmaktadır. RMSEA değerlerine bakıldığında, her iki grup için de .03 değerinde olduğundan modelin her

iki grupta da mükemmel uyum gösterdiğini söylenebilir. CFI ve TLI uyum indeksleri incelendiğinde ise, ana dili İngilizce olan grup için CFI değeri .96 ve TLI değeri ise .97 ile kabul edilebilir uyum sağlamaktadır. Ana dili İngilizce olmayan gruptan CFI değerinin .97 ile kabul edilebilir ve TLI değeri .98 ile iyi model uyumu verdiği görülmektedir.

DFA sonuçları PISA 2015 okuma becerileri testinden oluşan yapının her iki grup için de ayrı ayrı model uyumu gösterdiğini ortaya koymaktadır. Ana dili İngilizce olan ülkelerden oluşan grup ile ana dili İngilizce olmayan ülkelerden oluşan grup için ayrı ayrı model doğrulanmıştır.

3.2 Araştırmanın 2. Alt Problemine Yönelik Bulgular

Araştırmanın 2. alt probleminde “Ana dili İngilizce olan ülkelerden oluşan grup ile ana dili İngilizce olmayan ülkelerden oluşan grupların a, b parametreleri ve madde faktör yükleri farklılıklar göstermekte midir? sorusuna cevap aranmaktadır. Bu bağlamda her bir grup için yapılan DFA analizi sonucunda elde edilen madde faktör yükleri, eşik değerler, a ve b parametreleri Tablo 3’te verilmiştir.

Tablo 3. PISA 2015 Okuma Becerileri Testi Dil Değişkenine Göre Oluşturulan Grupların DFA Sonuçlarına İlişkin Madde Parametreleri

Madde	Anadili İngilizce Olan Ülkeler				Anadili İngilizce Olmayan Ülkeler			
	λ	t	a	b	λ	t	a	b
1	1.00	-0.81	.64	-1.50	1.00	-0.38	.95	-0.56
2	1.01	-1.13	.65	-2.07	.99	-0.35	.93	-0.51
3	1.06	-1.26	.70	-2.20	.88	-0.64	.76	-1.06
4	1.10	-1.07	.74	-1.80	.65	-0.62	.51	-1.37
5	1.23	-0.90	.88	-1.36	.61	-0.41	.46	-0.97
6	1.25	-0.92	.91	-1.36	1.03	-0.51	1.02	-0.71
7	1.18	.67	.82	1.06	1.03	.76	1.01	1.06
8	1.00	.33	.64	.61	.83	.69	.71	-1.24
9	1.31	-1.15	.99	-1.64	.84	-0.72	.71	-1.24
10	.88	.79	.54	1.68	.98	.98	.93	1.45
11	1.17	-0.06	.82	-0.09	.83	.22	.70	.39
12	.99	-0.08	.63	-0.14	.51	-0.13	.38	-0.36
13	1.19	-0.79	.84	-1.23	.97	-0.54	.90	-0.81
14	.90	-0.06	.56	-0.12	.70	.08	.55	.17
15	.50	.37	.28	1.36	.45	.53	.32	1.73
16	.63	-0.24	.36	-0.70	.60	.08	.46	.20
17	1.07	-0.86	.71	-1.48	.76	-0.76	.62	-1.26
18	1.31	-0.88	.99	-1.25	.76	-0.69	.61	-1.33
19	.91	-0.07	.56	-0.14	.67	.10	.53	.22
20	1.22	-0.41	.87	-0.62	1.02	.17	.99	.24
21	.20	-0.14	.85	-0.22	1.06	.30	1.07	.41
22	.99	-0.26	.63	-0.49	.85	-0.47	.72	-0.81
23	.87	-0.83	.53	-1.77	1.02	-0.34	.98	-0.48
24	.81	.16	.48	.36	.84	.39	.71	.68
25	.79	-0.10	.47	-0.21	.73	.53	.58	1.06
26	.77	-0.94	.46	-2.26	.60	-1.15	.45	-2.78
27	1.02	-0.32	.66	-0.57	.53	-0.27	.40	-0.64
28	1.26	.58	.92	.86	.96	.73	.88	1.10

λ = madde faktör yükü; t =madde eşik değeri; a =madde güçlük parametresi; b =madde zorluk parametresi

Tablo 3'te verilen ana dili İngilizce olan ülkelerden oluşan grup bazında faktör yükleri incelendiğinde, .20 ile 1.31 değerleri arasında değiştiği anlaşılmaktadır. En düşük faktör yüküne sahip 21. maddeyken ana dili İngilizce olmayan ülkelerden oluşan grupta ise en fazla faktör yüküne sahip olduğu görülmektedir. Tam ölçme değişmezliği yaklaşımı karşılaştırmalı gruplar arasında madde faktör yükleri ve eşik değerlerinin aynı, yaklaşık olarak tanımlanan ölçme değişmezliği yaklaşımı ise karşılaştırılan gruplar arasında söz konusu parametrelerde yalnızca küçük farklılıkların olmasına izin verir (Cao, Kim, Wang ve Nguyen, 2017). Bu doğrultuda her iki grup için de faktör yüklerinin birbirine yakın olması beklenirken, bu durum maddenin her iki grup için farklı çalıştığını göstermektedir. Yine Tablo 3'ten ana dili İngilizce olan grupta 9. ve 18. maddelerin en çok faktör yüküne sahip maddeler olduğu anlaşılmaktadır. Ana dili İngilizce olmayan grupta ise faktör yük değerleri .45 ile 1.06 arasında değişim göstermektedir. 15. madde ana dili İngilizce olmayan grupta en az faktör yüküne (.45) sahip maddedir ve diğer grubun 15. maddede verdiği faktör yükü (.50) ile oldukça yakın değerdedir. Her iki grubun faktör yükleri maddeler bazında karşılaştırıldığında 4, 9, 12, 18, 21, 27 ve 28. maddelerin faktör yüklerinin iki grupta birbirinden oldukça farklı değerler aldığı görülmektedir. Buradan bazı maddelerin her iki grup için de aynı şekilde çalışmadığı söylenebilir.

Madde eşik değerleri ile a ve b parametrelerini incelediğimizde 2. maddenin ana dili İngilizce olan grupta eşik değerinin -1.13 iken ana dili İngilizce olmayan grupta -.35 olduğu, a parametresinin ana dili İngilizce olan grupta .65, diğer grupta .93 ve b parametrelerinin ise ana dili İngilizce olan grupta -2.07; ana dili İngilizce olmayan grupta ise -0.51 olduğu görülmektedir. Bu değerler her iki grupta da aynı özelliği ölçmesi gereken bir madde için birbirinden oldukça farklı değerlerdir. Benzer şekilde 3. ve 23. maddelerin gruplar arasında b parametreleri karşılaştırıldığında, ana dili İngilizce olan grupta 3. maddenin b parametresi -2.20 iken ana dili İngilizce olmayan grupta -1.06; 23. maddede ise ana dili İngilizce olan grupta b -1.77, diğer grupta -0.48 değerini aldığı anlaşılmaktadır. Tablo 3'te 6, 8, 10, 15, 22, 25, 26 ve 27. maddelere ait eşik değerleri ve a, b parametrelerinde farklılaşmalar olduğu görülmektedir.

Her bir grup bazında yapılan DFA sonrasında ölçme değişmezliğini test etmek için Mplus 8.0 programı kullanılarak iki değişkenli kodlanmış maddelerde ÇGDFA yapılmıştır.

3.3 Araştırmanın 3. Alt Problemine Yönelik Bulgular

Araştırmanın 3. alt probleminde “Ana dili İngilizce olan ülkelerden oluşan grup ile ana dili İngilizce olmayan ülkelerden oluşan gruplarda maddeler bazında ölçme değişmezliğini sağlamakta mıdır? sorusuna cevap aranmaktadır. Bu doğrultuda yapılan iki değişkenli maddelerde ölçme değişmezliği analizi ile her bir madde için madde faktör yükleri ve eşik parametreleri serbest bırakılmıştır. 28 maddeden oluşan veri seti için 56 farklı modelde madde parametreleri ve BIC değerleri değişimi incelenmiştir. Ana dili İngilizce olan ve ana dili İngilizce olmayan her iki gruba ait iki değişkenli puanlanan maddelerde ölçme değişmezliği analizi sonuçları Tablo 4’ te verilmiştir.

Tablo 4. İki Değişkenli Maddelerde Ölçme Değişmezliği Analizi Sonuçları

Model	Par	BIC	Δ BIC	Grup 1		Grup 2	
				λ	t	λ	t
M1	t ₁	44748.62	3.28	-	-1.69	-	-1.27
M2	t ₂	44708.93	36.41*	-	-2.35	-	-1.25
M3	t ₃	44737.70	7.64*	-	-2.69	-	-2.02
M4	t ₄	44748.02	2.68	-	-2.21	-	-1.87
M5	t ₅	44746.98	1.64	-	-1.84	-	-1.48
M6	t ₆	44752.57	7.23*	-	-2.14	-	-2.19
M7	t ₇	44739.41	5.93	-	1.66	-	1.00
M8	t ₈	44751.38	6.04*	-	.71	-	.87
M9	t ₉	44752.39	7.05*	-	-2.76	-	-2.66
M10	t ₁₀	44752.62	7.28*	-	1.65	-	1.61
M11	t ₁₁	44751.68	6.34*	-	-0.10	-	-0.24
M12	t ₁₂	44731.95	13.39*	-	-0.14	-	-.72
M13	t ₁₃	44749.29	3.95	-	-1.78	-	-2.10
M14	t ₁₄	44748.84	3.50	-	-0.10	-	-0.34
M15	t ₁₅	44752.65	7.31*	-	.64	-	.65
M16	t ₁₆	44748.44	3.10	-	-0.42	-	-0.17
M17	t ₁₇	44749.87	4.43	-	-1.79	-	-2.05
M18	t ₁₈	44745.93	1.59	-	-2.01	-	-2.44
M19	t ₁₉	44751.14	5.80	-	-0.13	-	-0.29
M20	t ₂₀	44742.71	2.63	-	-0.84	-	-0.37
M21	t ₂₁	44751.50	6.16*	-	-0.21	-	-0.71
M22	t ₂₂	44691.77	53.57*	-	1.13	-	-1.58
M23	t ₂₃	44745.36	.02	-	-1.66	-	-1.30
M24	t ₂₄	44752.56	7.22*	-	.31	-	.27
M25	t ₂₅	44722.68	22.66*	-	-1.15	-	.58
M26	t ₂₆	44725.27	20.07*	-	-1.75	-	-2.66
M27	t ₂₇	44743.97	1.37	-	-0.59	-	-0.97
M28	t ₂₈	44743.49	1.85	-	1.32	-	.82
M29	λ_1	44747.36	2.02	1.15	-	1.57	-
M30	λ_2	44722.44	22.90*	1.09	-	2.09	-

Tablo 4 devam

Model	Par	BIC	Δ BIC	Grup 1		Grup 2	
				λ	t	λ	t
M31	λ_3	44741.84	3.50	1.20	-	1.73	-
M32	λ_4	44752.67	7.43*	1.19	-	1.20	-
M33	λ_5	44751.06	5.72	1.36	-	1.17	-
M34	λ_6	44752.55	7.21*	1.68	-	1.74	-
M35	λ_7	44749.53	4.19	1.30	-	1.74	-
M36	λ_8	44752.60	7.26*	1.09	-	1.14	-
M37	λ_9	44752.26	6.92*	1.75	-	1.65	-
M38	λ_{10}	44746.56	1.12	.92	-	1.49	-
M39	λ_{11}	44748.68	4.34	1.40	-	1.05	-
M40	λ_{12}	44731.21	24.13*	1.06	-	.44	-
M41	λ_{13}	44751.05	5.71	1.52	-	1.33	-
M42	λ_{14}	44752.08	6.74*	.89	-	.78	-
M43	λ_{15}	44752.64	7.30*	.50	-	.51	-
M44	λ_{16}	44750.76	5.42	.59	-	.77	-
M45	λ_{17}	44749.77	4.43	1.28	-	1.05	-
M46	λ_{18}	44736.14	9.20*	1.75	-	1.15	-
M47	λ_{19}	44751.27	5.93	.94	-	.77	-
M48	λ_{20}	44749.49	4.15	1.46	-	.86	-
M49	λ_{21}	44751.57	6.23*	1.46	-	1.69	-
M50	λ_{22}	44736.79	8.55*	1.18	-	.63	-
M51	λ_{23}	44734.46	10.88*	.97	-	1.62	-
M52	λ_{24}	44748.02	2.68	.75	-	1.11	-
M53	λ_{25}	44747.32	1.98	.78	-	1.17	-
M54	λ_{26}	44739.60	5.71	.95	-	.44	-
M55	λ_{27}	44736.40	8.64*	1.12	-	.58	-
M56	λ_{28}	44752.64	7.30*	1.43	-	1.40	-

λ = madde faktör yükü; t=madde eşik değeri;

Grup 1: Ana dili İngilizce Olan Ülkeler **Grup 2:** Ana dili İngilizce Olmayan Ülkeler

Tablo 4 incelendiğinde her bir madde için madde faktör yükleri ve eşik değerlerinin serbest bırakılması ile 56 farklı modelden elde edilen BIC değerleri, bunların M_0 serbest modelindeki (nested) BIC değerinden farkları (ΔBIC) ve madde faktör yükleri ile eşik değerleri yer almaktadır. M_0 serbest modelinin BIC değerleri ile her bir modelin BIC değerleri ayrı ayrı karşılaştırılmıştır. BIC değeri ne kadar küçükse, modelin o kadar mükemmel uyum verdiği anlamına gelmektedir. ΔBIC değerinin büyüklüğü de modelin iyiliği hakkında kanıt oluşturmaktadır. $6 < \Delta BIC < 10$ arasındaki değerler modelin iyi olmadığına dair güçlü, $\Delta BIC > 10$ olan değerler ise çok güçlü bir kanıt sunar (Fabozzi, Frank J. ve Wiley, 2014). Bu doğrultuda $\Delta BIC > 10$ olan maddeler oldukça kötü bir model uyumu veriyor anlamına gelir, dolayısıyla ölçme değişmezliği sağlanamadığına dair bir kanıt oluşturmaktadır (Dimitrov, Li, Marcoulides, Menold ve Raykov, 2018).

M_0 serbest modelinde BIC değeri 44745.34 olarak bulunmuştur. Her bir modeldeki BIC değerinin bu değerle farkı hesaplanmıştır. Elde edilen bulgulara göre, 2. maddenin Model 2’de ΔBIC değerinin 36.41 ($\Delta BIC > 10$) olduğu görülmektedir. Yine 2. Maddeye ait eşik değerlerinin de ana dili İngilizce olan ülkelerden oluşan grup (Grup 1) için -2.35 ve ana dili İngilizce olmayan ülkelere oluşan grup için (Grup 2) -1.25 değerleriyle birbirinden oldukça farklı olduğu anlaşılmaktadır. Bu doğrultuda 2. maddenin model uyumu göstermediği ve her iki grup için de karşılaştırılabilir olmadığı söylenebilir. Madde 3’ün ΔBIC değerinin 7.64 ($6 < \Delta BIC < 10$) ile yine kötü model uyumu veren aralıkta olduğu anlaşılmaktadır. Madde eşik değerleri ve a,b parametrelerinin de birbirinden farklı değerler aldığı Tablo 3’ten görülmektedir. Madde 4’ün Model 32’deki ΔBIC değerlerinin 7.43 ($6 < \Delta BIC < 10$) olduğu görülmektedir.

Tablo 4’te madde faktör yüklerine bakıldığında ana dili İngilizce olan ülkeler için 1.10; ana dili İngilizce olmayan ülkeler için.65 değerleriyle birbirinden oldukça farklı olduğu görülmektedir. Madde parametreleri ve BIC değerlerindeki farka göre 4. maddenin de ölçme değişmezliğini tam olarak sağlamadığına dair bir kanıt oluşturduğu söylenebilir. Madde 6,8,9 ve 10’da da BIC değerleri arasındaki farklar 6 ile 10 arasında değer almaktadır. Araştırmanın 2. alt probleminin bulgularında bu maddelere ait parametreler arasında da farklılıklar bulunmaktadır. Benzer şekilde 12. maddenin de ΔBIC değeri 13.39 ($\Delta BIC > 10$) olduğu tablodan anlaşılmaktadır. Model

40'ta da 12. maddenin ΔBIC deęerinin 24.13 ($\Delta BIC > 10$) ve madde faktör yüklerinin ana dili İngilizce olan grupta 1.06; ana dili İngilizce olmayan grup için ise .44 olduęu bulunmuştur. Elde edilen bu deęerler 12. maddenin de söz konusu gruplar arasında karşılaştırılabilir olmadığına dair güçlü bir kanıt oluşturmaktadır (Dimitrov, Li, Marcoulides, Menold ve Raykov, 2018).

Madde 15, 18, 21'in ΔBIC deęerleri de 6 ile 10 arasında olup, ölçme deęişmezliğini kısıtlayan maddeler arasında oldukları söylenebilir. Bu maddelere ilişkin parametreler de birbirinden farklılıklar göstermektedir. Madde 22'nin ΔBIC deęerine bakıldığında 53.57 ($\Delta BIC > 10$) olduęu, madde faktör yüklerinin sabit, eşik deęerlerin serbest bırakıldığı modelde eşik parametrelerinin de gruplar bazında oldukça farklı deęerlerde ($t_{22} = 1.13_1$ ve $t_{22} = -1.58_2$) olduęu görülmektedir. Madde 23'ün de Model 51'de ΔBIC deęerinin 10.88 ($\Delta BIC > 10$) olduęu anlaşılmaktadır. Ayrıca b parametrelerinin de birbirinden uzak deęerlerde olduęu araştırmanın 2. alt problemine dair bulgularda açıklanmıştır. 25. maddenin BIC deęerlerindeki farka bakıldığında deęerin 22.66 ($\Delta BIC > 10$) ile ölçme deęişmezliğinin tam olarak sağlanamadığı maddeler arasında olduğunu söylenebilir. Benzer şekilde b parametrelerinin birbirinden çok farklı olması (ana dili İngilizce olan ülkeler için -0.21, ana dili İngilizce olmayan ülkeler için 1.06) da bu durumu desteklemektedir. 26. maddenin de BIC farkları incelendiğinde, 20.07 deęeriyle ($\Delta BIC > 10$) ölçme deęişmezliğinin kabul edilebilir sınırları dışında olduęu anlaşılmaktadır. Son olarak 28. maddenin madde faktör yüklerinin serbest bırakıldığı Model 56'ya göre ΔBIC deęeri 7.30 ($6 < \Delta BIC < 10$) olduęu görülmektedir. Bu maddeye ait eşik deęer ve faktör yükleri de birbirinden oldukça farklı deęerler olup ölçme deęişmezliği problemi içeren maddeler arasında gösterilebilir.

Elde edilen bulgulara göre 28 maddeden oluşan PISA-2015 okuma becerileri testinin yarısından fazlasının (17 madde) ölçme deęişmezliğini tam olarak sağlamadığı söylenebilir. İkili puanlanmış maddelerde ölçme deęişmezliği analizi ile ulaşılan sonuçlarda en az bir maddenin ölçme deęişmezliğini tam olarak sağlamaması, testin ölçme deęişmezliğini sınırlandırarak testin bütününde bir ölçme deęişmezliği sorun olabileceğini ortaya koymaktadır (Dimitrov, Li, Marcoulides, Menold ve Raykov, 2018).

BÖLÜM IV

SONUÇ VE ÖNERİLER

Bu araştırmada PISA 2015 okuma becerileri testinin ana dili İngilizce olan ülkeler ile ana dili İngilizce olmayan ülkeler arasında dil değişkeni açısından ölçme değişmezliği ikili puanlanmış maddelerde çok gruplu doğrulayıcı faktör analizi yöntemi ile test edilmiştir. Araştırmanın bulgularına göre elde edilen sonuçlar aşağıda özetlenmiştir. Bu sonuçlara göre ölçme değişmezliği çalışmaları için uygulamaya ve araştırmaya dönük önerilere yer verilmiştir.

4.1 Sonuçlar

Araştırma ile ulaşılan bulgulardan elde edilen sonuçlar aşağıda araştırmanın alt problemlerinin sırasına göre verilmiştir.

Bu araştırmada veri setinin kategorik yapısına göre MTK'ya dayalı ölçme değişmezliği analizi yöntemi kullanılmıştır. Sürekli değişkenler için kullanılan aşamalı çok gruplu doğrulayıcı faktör analizi yönteminin veri setinin kategorik yapısına uymadığı görülmüştür. İkili puanlanmış maddelerde ÇGDFA ile elde edilen bulgulara göre ulaşılan sonuçlar aşağıdaki gibidir.

Ana dili İngilizce olan ve ana dili İngilizce olmayan ülkelere oluşan iki grup için ayrı ayrı DFA yapılarak model uyumu incelenmiştir. Yapılan DFA ile elde edilen bulgulara göre her iki grubun da model uyumunun kabul edilebilir aralıklarda olduğu anlaşılmıştır. Faktör yapıları her bir grup için doğrulanmıştır. DFA maddelerin yapı geçerliliğine ilişkin bilgi vermektedir (Gregoric, 2006). Buradan grupların ayrı ayrı yapı geçerliliğini sağladığına ilişkin bir sonuca varılabilir.

Ölçme değişmezliğinin sağlanabilmesi için yapıların her bir grupta aynı olması gerekmektedir. DFA ile elde edilen bulgularla her bir maddenin ana dili İngilizce olan ve ana dili İngilizce olmayan gruplarda a ve b parametreleri karşılaştırılmıştır. PISA 2015 okuma becerileri testinin her iki grupta aynı özelliği ölçtüğü varsayılan maddelerinin bazılarının faktör yükleri ve eşik parametrelerinin birbirinden oldukça farklı olduğu anlaşılmıştır. Bazı maddelerin gruplardan birinde diğerine göre oldukça fazla faktör yükü verdiği ya da aynı yetenek düzeyine sahip olduğu düşünülen

bireylerin her iki grupta soruyu doğru cevaplama olasılıkları arasında farklılıklar olduğu bulunmuştur. Dolayısıyla bu durum, ana dili değişkenine göre oluşturulan grupların birbirleriyle karşılaştırılabilirlikleri için bir sınırlılık olabileceği sonucuna varılmıştır.

Gruplar arasında ölçme değişmezliğini incelemek için ikili puanlanmış maddelerde ÇG DFA yapılmıştır. Elde edilen bulgulara göre DFA sonuçlarına paralel olarak, ölçme değişmezliğini sağlamadığı düşünülen maddelerde BIC değerleri arasındaki farkın 6 ile 10 arasında ya da 10'dan büyük olduğu belirlenmiştir. Bu maddelere ilişkin eşik değerleri ve faktör yükleri karşılaştırıldığında ise farklılaşma olduğu görülmüştür. Bu durumda PISA 2015 okuma becerileri testinin 41 numaralı formunda yer alan 28 maddeden 17'sinin ana dili değişkenine göre oluşturulan gruplar arasında karşılaştırma yapılmasını ciddi şekilde sınırladığı sonucuna varılmıştır. En az bir maddede böyle bir sınırlılık olması testin bütünü için ölçme değişmezliğinin tam olarak sağlanamadığı anlamına gelmektedir (Dimitrov, Li, Marcoulides, Menold ve Raykov, 2018). Dolayısıyla bu testte ölçme değişmezliğinin, ana dili İngilizce olan ülkeler ile ana dili İngilizce olmayan ülkelere oluşan gruplar arasında karşılaştırılabilir olmasına kısıtlamalar getiren kaynaklar tanımlanmadan tam olarak yorumlanamayacağı yargısına varılabilir.

Alanyazında bu çalışmanın bulgularından ulaşılan sonuçlarla aynı doğrultuda olan ölçme değişmezliği çalışmaları bulunmaktadır. Peterson ve arkadaşları (2003), Avrupa'da konuşulan 10 farklı dili temsil eden katılımcılardan elde edilen verilerde EORTCQLQ-C30 duygusal fonksiyon ölçeğinin orijinal versiyonu (İngilizce) ve dokuz çevirisinin DMF analizleri ile karşılaştırmasını yapmışlardır. İki farklı DMF yöntemi kullanılmış ikisinde de benzer şekilde dokuz çeviriden yedisinin DMF gösterdiği sonucuna varılmıştır. Tercümelemlerin hiçbir zaman orijinali ile aynı olamayacağı göz önüne alındığında, DMF analizi yönteminin testlerin çevirilerinin birebir doğruluğunu kanıtlamak için iyi bir yol olduğu ön görülmüştür (Bjorner, Groenvold ve Peterson, 2003).

Ercikan ve Koh (2005), TIMMS 1995 verilerini kullanarak matematik ve fen testlerinin dil değişkeni açısından testin İngilizce ve Fransızca versiyonlarında ölçme değişmezliğini incelemişlerdir. MTK ve ÇG DFA ile karşılaştırmalar yapmış ve iki farklı dilde uygulanan formlar arasında önemli farklılıklar olduğu sonucuna

ulaşmışlardır. DMF analizleri ile de iki grubun karşılaştırılmasında önemli kısıtlılıklar olduğu yargısına varmışlardır.

Raykov ve arkadaşları (2018), ikili kategorik puanlanmış maddelerde ölçme değişmezliğini incelediği çalışmasında ÇGDFA sonuçları ile DMF sonuçlarının birbiriyle paralel olduğu sonucuna ulaşmıştır. İkili maddelerde MTK yaklaşımı ile yapılan ÇGDFA analizi sonuçlarının DMF analizleriyle doğrulanabileceği ön görülmüştür. Dolayısıyla bu çalışmada DMF analizi yapılırsa sonuçların ÇGDFA bulguları ile örtüşeceği beklenilmektedir (Dimitrov, Li, Marcoulides, Menold ve Raykov, 2018).

4.2 Öneriler

Bu bölümde araştırma süreci ve ulaşılan sonuçlar doğrultusunda, uygulayıcı ve araştırmacılara yönelik önerilere yer verilmiştir.

Uygulayıcılara Yönelik Öneriler

PISA uygulamalarından elde puanların katılımcı ülkeler arasında karşılaştırılabilir olması için, yapı ve kapsam geçerliğinin yanı sıra bir çok değişken açısından ölçme değişmezliğinin de test edilmesine ihtiyaç vardır. Ölçme değişmezliği ve DMF analizleri sonuçları ile yanlılık gösteren ya da ölçme değişmezliğini kısıtlayan maddeler için uygulama yapılmadan önce gerekli düzeltmeler dikkatle yapılmalıdır. Ölçek geliştirme ve uyarlama çalışmalarında ölçme aracının gruplar arasındaki eşdeğerliği de göz önünde bulundurulmalıdır.

Sorular erişime açık olsaydı, ölçme değişmezliğine sınırlılık getiren maddeler incelenerek sonuçlar arasındaki farklılıklar ayrıntılı olarak incelenebilirdi. PISA'nın resmi web sitesinden bu maddelere ulaşamadığı için araştırmaya dahil edilememiştir. PISA uygulamasında maddelerin erişime açık olması ve bir çok değişken açısından içerik olarak incelenmesi, ölçme değişmezliğini sınırlayan faktörlerin araştırılması açısından önemlidir. Bu doğrultuda PISA'nın resmi web sitesinde yalnızca uygulandığı yıla özgü alanın soruları değil, tüm soruların erişime açık olması önerilebilir.

Farklı diller ve kültürlerarasında yapılan ölçmelerde bu değişkenlere bağlı yanlılık gösteren ya da eşdeğerliği sağlamayan madde içerikleri mutlaka göz önünde bulundurulmalıdır. Belli bir dile özgü olan ve ayrıcalık sağlayacak anlatımlar içeren maddeler testin dışında bırakılmalıdır.

PISA gibi bir çok dile çevirisi yapılan uluslararası uygulamalarda orijinal hali İngilizce dilinde yazılan sorular, uzman çevirmenler tarafından başka bir dile çevrilip ardından orijinal versiyonu ile eşdeğerliğini sağlamak için tekrar İngilizceye çevrilmektedir. Bu uygulamalarda yalnızca çeviriden kaynaklı değil, eşdeğerlik sorunu oluşturan veya yanlılık kaynağı olarak gösterilen etmenler, “karşılaştırılabilirlik” kavramı iyi tanımlanarak belirlenmelidir. Bu etmenleri dikkatlice incelemek için ölçme araçlarında anlayışlardaki kültür farklılıklarından kaynaklanan ve dile yansıyan etkileri hakkında bilgi edinilmelidir (Goldstein, 2017).

Araştırmacılara Yönelik Öneriler

Bu araştırmanın analizleri seçilen ülkelerde ortak olarak uygulanan ve en çok cevaplanmış olan 41 numaralı forma ait veriler üzerinden yapılmıştır. Seçilen formun araştırmaya dahil edilen ülkelerde bilgisayar tabanlı uygulamada kullanılan formlardan biri olmasına dikkat edilmiştir. Bilgisayar tabanlı uygulamada bilgisayar okur yazarlığı becerisi de gerektirdiğinden söz konusu formların sınırlılıkları olabilmektedir. Dolayısıyla kağıt- kalem tabanlı uygulanan farklı formlarda da ölçme değişmezliği çalışmaları yapılmalıdır.

Araştırmamıza ana dili İngilizce olmayan ülkelere Japonya (Japonca), Türkiye (Türkçe) ve Tayland (Tayca) dahil edilmiştir. PISA okuma becerileri testlerinin farklı dil gruplarında da ölçme değişmezliği incelenmelidir. PISA uygulamasının orijinal dili İngilizce olduğundan İngiltere, ABD ve Kanada bu grupta çalışmaya dahil edilmiştir. Ana dili İngilizce olan başka ülke gruplarının ölçme değişmezliği çalışmaları yapılabilir.

Bu çalışmada veri setinin kategorik yapısı dikkate alınarak iki farklı yaklaşımdan yola çıkılarak ölçme değişmezliği analizi yapılmıştır. İki kategorili

puanlanmış veri setlerinin yapısına uygun GDFA analizi yapıldıktan sonra, DMF analizleri de yapılarak yanlılık gösteren maddeler ile ölçme değışmezliğini kısıtlayan maddelerin tutarlılığı araştırılmalıdır. Söz konusu maddelerde eşdeğerliği kısıtlayan yanlılık kaynakları detaylıca incelenmeli, uzman görüşleri alınarak nedenleri anlaşılmalıdır.



KAYNAKÇA

- Adams, R., & Rowe, K. (1988). *Educational research, methodology, and measurement: An international handbook*. Oxford: Pergamon Press.
- Akyıldız, M. (2009). PIRLS 2001 testinin yapı geçerliliğinin ülkelerarası karşılaştırılması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 18-47.
- Anastasi, A. (1997). *Psychological Testing*. New Jersey.
- Asil, M., & Gelbal, S. (2012). PISA Öğrenci Anketinin Kültürler Arası Eşdeğerliği. *Eğitim ve Bilim*, 236-249.
- Asparouhov, T., Muthén, B., & Morin, A. J. (2015). Bayesian Structural Equation Modeling With Cross-Loadings and Residual Covariances: Comments on Stromeier et al. *Journal Of Management*.
- Asparouhov, T., & Muthén, B. (2014). Auxiliary Variables in Mixture Modeling: Three-Step Approaches Using Mplus. *Structural Equation Modeling*, 329-341.
- Atılğan, H., Kan, A., & Aydın, B. (2017). *Eğitimde Ölçme ve Değerlendirme*. Ankara : Anı Yayıncılık, 10. Baskı.
- Atılğan, H., Kan, A., & Doğan, N. (2016). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Anı.
- Ayvalı, M. (2016). Pisa 2012 Matematik Okuryazarlığının OECD Üyesi Ülkeler ve Türkiye'deki Cinsiyet ve Bölgelere Göre Ölçme Değişmezliğinin İncelenmesi.
- Baker, F. B. (2016). *Madde Tepki Kuramının Temelleri*. Ankara: Pegem Akademi .
- Başusta, N. (Kış 2010). Ölçme Eşdeğerliği. *Eğitim ve Psikolojide Ölçme Ve Değerlendirme Dergisi*.
- Başusta, N., & Gelbal, S. (2015). Gruplararası Karşılaştırmalarda Ölçme Değişmezliğinin Test Edilmesi: PISA Öğrenci Anketi Örneği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 80-90.
- Baykul, Y. (1999). *Eğitimde ve Psikolojide Ölçme*. Ankara: Pegem Akademi.
- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Baykul, Y. (2015). *Klasik Test Teorisi ve Uygulaması*. Ankara: Pegem Yayıncılık 3. Baskı.
- Baykul, Y., & Turgut, M. (2015). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Akademi.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Bryne, B., & Watkins, D. (2003). *The Issue of Measurement Invariance*. Journal Of Cross-Cultural Psychology.
- Buerger, S., Kroehne, U., & Goldhammer, F. (2016). The transition to computer-based testing in large-scale assessments: Investigating (partial) measurement invariance between modes. *Psychological Test and Assessment Modeling*, 597-616.
- Campbell, M. (1928). *An Account of Measurement and Calculation*. London: McKay.
- Cheema, J. (2012). Handling missing data in educational research using SPSS. Unpublished doctoral dissertation. *George Mason University*.

- Cheung, G., & Rensvold, R. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling*, 233-255.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston.
- Çetin, S., & Ölçülüoğlu, R. (2016). TIMSS 2011 Sekizinci Sınıf Öğrencilerinin Matematik Başarısını Etkileyen Değişkenlerinin Bölgelere Göre İncelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 202-220.
- Çokluk, Ö., & Kayri, M. (2011). Kayıp değerlere yaklaşık değer atama yöntemlerinin ölçme araçlarının geçerlik ve güvenilirliği üzerindeki etkisi. *Kuram ve Uygulamada Eğitim Bilimleri*, 289-309.
- Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2010). *Sosyal Bilimler için Çok Değişkenli İstatistik SPSS ve LISREL Uygulamaları*. Ankara: Pegem Akademi.
- DeVellis, R. (1999). *Scale Development Theory and Applications*.
- Development), O. f.-o. (2016). <https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf>.
- Development), O. f.-o. (2016). <https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf>.
- Downey, R., & King, C. (1998). Missing Data in Likert Ratings: A Comparison of Replacement Methods. *The Journal of General Psychology*, 175-191.
- Eğitek. (2011). *PISA TÜRKİYE*. Ankara.
- Ercikan, K., & Koh, K. (2005). Examining the Construct Comparability of the English and French Versions of TIMSS. *International Journal Of Testing*, 23-35.
- Frank J. Fabozzi, S. M., & Wiley, J. (2014). Model Selection Criterion: AIC and BIC. *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*.
- Glanville, J. L., & Wildhagen, T. (2007). The Measurement of School Engagement: Assessing Dimensionality and Measurement Invariance Across Race and Ethnicity. *Educational and Psychological Measurement*.
- Goldstein, H. (2017). Measurement and evaluation issues with PISA. *Routledge*.
- Gregoric, S. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups?: Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, 78-94.
- Hambelton, R. K., & Swaminathan, H. (1985). *Item Response Theory*. Nijhoff Publishing.
- Hambleton, R. K. (2005). Issues, Designs, and Technical Guidelines for Adapting Tests into Multiple. *Adapting educational and psychological tests for cross-cultural assessment*, 3-38.
- Hambleton, R. K., & De Jong, J. A. (2003). Advances in translating and adapting educational and psychological tests. *Language Testing*, 127-134.
- Hau, K., & Marsh, H. (2004). The use of item parcels in structural equation modelling: Nonnormal data and small sample sizes. *British Journal of Mathematical and Statistical Psychology*, 327-351.
- Hu, L., & Bentler, P. (1999). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 117-144.
- Kıbrısoğlu, N. (2015). PISA 2012 Matematik Öğrenme Modelinin Kültürlere ve Cinsiyete Göre Ölçme Değişmezliğinin İncelenmesi: Türkiye-Çin(Şangay)-Endonezya Örneği.

- Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement Invariance Testing with Many Groups: A Comparison of Five Approaches. *Structural Equation Modeling: A Multidisciplinary Journal*.
- Kline, R. (2011). *Principles and Practices of Structural Equation Modelling*. New York: The Guilford Press.
- Kline, R. (2011). *Principles and Practices of Structural Equation Modelling*. New York: The Guilford Press.
- Knight, G., & Hill, N. (1998). Measurement equivalence in research involving minority adolescents. *Studying Minority Adolescents: Conceptual, Methodological, And Theoretical Issues*, 183-210.
- Marsh, H. W., Ebu-Hilal, M. M., & Abduljabbar, A. S. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: a comparison of arab and anglo-saxon countries. *Journal of Educational Psychology*, 108-128.
- Marsh, H., Hau, K., Artelt, C., Baumert, J., & Peschar, J. (2006). OECD's Brief Self-Report Measure of Educational Psychology's Most Useful Affective Constructs: Cross-Cultural, Psychometric Comparisons Across 25 Countries. *International Journal Of Testing*, 311-360.
- Maruyama, G. (1998). *Basics of structural equation modeling*. CA: Sage Publications.
- MEB. (2016). *PISA 2015 Ulusal Raporu*. Ankara.
- MEB. (2016). *PISA 2015 Ulusal Raporu*. Ankara.
- MEB. (2016). *PISA 2015 Ulusal Raporu*. Ankara.
- Miles, J., & Shevlin, M. (2007). A time and a place for incremental fit indices. *Personality and individual differences*, 869-874.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: US: Routledge/Taylor & Francis Group.
- Millsap, R. E., & Olivera-Aguilar, M. (2012). *Investigating Measurement Invariance Using Confirmatory Factor Analysis*. New York: The Guilford Press.
- Mulaik, A. (1972). *The Foundations of the Factor Analysis*. New York: McGraw-Hill Book Company.
- Muthén, L., & Muthén, B. (2016). *Mplus User's Guide: Statistical Analysis with Latent Variables*. *Psychology*.
- OECD. (2010). *Translation and adaptation guidelines for PISA 2012*.
- OECD. (2015). *(Organisation for Economic Co-operation and Development)*.
- Öğretmen, T. (2006). *Uluslararası okuma becerilerinde gelişim projesi (PIRLS) 2001 testinin psikometrik özelliklerinin incelenmesi: Türkiye-Amerika Birleşik Devletleri örneği*. Ankara.
- Öğretmen, T., & Başokçu, T. (2013). Öğretmen Öz-yeterlilik Ölçeğinde Değişen Madde Fonksiyonlarının Ağırlıklandırılmış Cevap Modeli ile Belirlenmesi. *Ege Eğitim Dergisi*, 63-78.
- Öğretmen, T., & Uzun, B. (2010). Fen başarısı ile ilgili bazı değişkenlerin TIMSS-R Türkiye örneğinde cinsiyete göre ölçme değişmezliğinin değerlendirilmesi. *Eğitim ve Bilim*, 26-35.
- Önen, E. (2009). *Ölçme Değişmezliğinin Yapısal Eşitlik Modellemesi Teknikleri ile İncelenmesi*. Ankara: Ankara Üniversitesi, Doktora Tezi.
- Özçelik, D. (2011). *Ölçme ve Değerlendirme*. Ankara: Pegem Akademi.
- Pant, H., & Segeritz, M. (2013). Do They Feel the Same Way About Math? Testing Measurement Invariance of the PISA "Students' Approaches to Learning"

- Instrument Across Immigrant Groups Within Germany. *Educational and Psychological Measurement*, 601-630.
- Peterson , M. A., Groenvold, M., & Bjonner, J. (2003). Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. *Quality of Life Research*, 373-385.
- Raykov, T., Dimitrov, D., Marcoulides, G., Li, T., & Menold, N. (2018). Examining Measurement Invariance and Differential Item Functioning With Discrete Latent Construct Indicators: A Note on a Multiple Testing Procedure. *Educational and Psychological Measurement*, 343-352.
- Raykov, T., Millsap, R. E., & Marcoulides, G. A. (2013). Factorial Invariance in Multiple Populations A Multiple Testing Procedure. *Educational and Psychological Measurement*, 713-727.
- Salzberg, T., Sĭnkovics, R., & Schlgelmich, B. (1999). Data equivalence in cross-cultural research: a comparison of classical test theory and latent trait theory based approaches. *Australasian Marketing Journal*, 23-38.
- Schermelleh-Engel, K., & Moosbrugger, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 23-74.
- Sireci, S. G., & Berberođlu, G. (2000). Using bilingual respondents to evaluate translated-adapted items. *Applied Measurement in Education*, 229-248.
- Tabachnick, B., & Fidell, L. (2013). Using multivariate statistics. *New Jersey: Pearson*.
- Thorndike, L., & Hagen, E. (1959). *Measurement and Evaluation in Psychology and Education*. . New York: Chapman and Hall, Limited.
- Turgut, M., & Baykul, Y. (1992). *Ölçekleme Teknikleri*. Ankara: ÖSYM Yayınları.
- Uyar, Ş., & Dođan , N. (2014). PISA 2009 Türkiye Örnekleminde Öğrenme Stratejileri Modelinin Farklı Gruplarda Ölçme Deđişmezliđinin İncelenmesi. *Uluslararası Türk Eğitim Bilimleri Dergisi*, 30-43.
- Wainer, H. (1990). *Computerized Adaptive Testing*.
- Wicherts, J., Dolan, C., Hessen, D., Oosterveld, P., Baal, G., Boomsma, D., & Span, M. (2004). Are intelligence tests measurement invariant over time? Investigation the nature of the Flynn effect. *Intelligence*, 509-537.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in substance use domain. *Methodological advances from alcohol and substance abuse*.
- Widaman, K., & Reise, S. (1997). Exploring the measurement invariance of psychological instruments: Applications in substance use domain. The science of prevention. *Methodological advances from alcohol and substance abuse research*, 281-324.
- Wu, D., Li, Z., & Zumbo, B. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assesment, Research & Evaluation*, 1-26.

EKLER

EK 1. PISA 2015 Okuma Becerileri Testinde Ölçme Değişmezliğini Sınırlayan Maddelerin Kodları

Madde	Kod
2	DR219Q01C- Employment (Q01)
3	DR219Q02C-Employment (Q02)
4	CR067Q01S-Aesop (Q01)
6	DR067Q5C-Aesop (Q05)
8	DR102Q05C-Shirts(Q05)
9	CR102Q07S-Shirts (Q07)
10	CR220Q01S-South Pole (Q01)
12	CR220Q04S-South Pole (Q04)
15	CR424Q02S- Fair Trade (Q02)
18	CR404Q03S-Sleep (Q03)
21	DR404Q10BC- Sleep (Q10B)
22	DR406Q01C- Kokeshi Dolls (Q01)
23	DR406Q05C- Kokeshi Dolls (Q05)
25	DR455Q02C- Chocolate and Healts (Q02)
26	DR455Q03C -Chocolate and Healts (Q03)
27	CR455Q04S- Chocolate and Healts (Q04)
28	CR455Q05S- Chocolate and Healts (Q05)

EK 2. İKİLİ PUANLANMIŞ MADDELERDE ÇOK GRUPLU DOĞRULAYICI FAKTÖR ANALİZİ İÇİN KULLANILAN MPLUS 8.0 KODLARI

M₀ Serbest Modeli İçin Yazılan Kod:

TITLE: Raykov M0

DATA: FILE = multicfaALL1.dat;

!LISTWISE=ON;

VARIABLE: NAMES = g u1-u28;

CATEGORICAL = u1-u28;

KNOWNCLASS = C(g = 1 g = 2); !g=1 ING, g=2 NOing

CLASSES = C(2);

MISSING=ALL(999);

ANALYSIS: ESTIMATOR = ML;

TYPE = MIXTURE;

ALGORITHM = INTEGRATION;

MODEL:

%OVERALL%

f1 BY u1* (L1)

u2-u28 (L2-L28);

[u1\$1-u28\$1](T1-T28);

[f1@0];

f1@1;

%C#2%

f1 BY u1* (L1)

u2-u28 (L2-L28);

[u1\$1-u28\$1](T1-T28);

[f1*];

f1*;

Madde Eşik Değerlerinin Serbest Bırakıldığı Bir Model Örneği (M₁-M₂₈):

TITLE: Raykov M1 (relase first threshold)

!LISTWISE=ON;

DATA: FILE = multicfaALL1.dat;

VARIABLE: NAMES = g u1-u28;

CATEGORICAL = u1-u28;

KNOWNCLASS = C(g = 1 g = 2); !g=1 ING, g=2 NOing

CLASSES = C(2);

MISSING=ALL(999);

ANALYSIS: ESTIMATOR = ML;

TYPE = MIXTURE;

ALGORITHM = INTEGRATION;

MODEL:

%OVERALL%

f1 BY u1* (L1)

u2-u28 (L2-L28);

[u1\$1-u28\$1](T1-T28);

[f1@0];

f1@1;

%C#2%

f1 BY u1* (L1)

u2-u28 (L2-L28);

[u2\$1-u28\$1](T2-T28);

[u1\$1*];

[f1*];

f1*;

Madde faktör Yüklerinin Serbest Bırakıldığı Bir Model Örneği(M29-M56):

TITLE: Raykov M29 (relase first loading)

!LISTWISE=ON;

DATA: FILE = multicfaALL1.dat;

VARIABLE: NAMES = g u1-u28;

CATEGORICAL = u1-u28;

KNOWNCLASS = C(g = 1 g = 2); !g=1 ING, g=2 NOing

CLASSES = C(2);

MISSING=ALL(999);

ANALYSIS: ESTIMATOR = ML;

TYPE = MIXTURE;

ALGORITHM = INTEGRATION;

MODEL:

%OVERALL%

f1 BY u1* (L1)

u2-u28 (L2-L28);

[u1\$1-u28\$1](T1-T28);

[f1@0];

f1@1;

%C#2%

f1 BY u1*

u2-u28 (L2-L28);

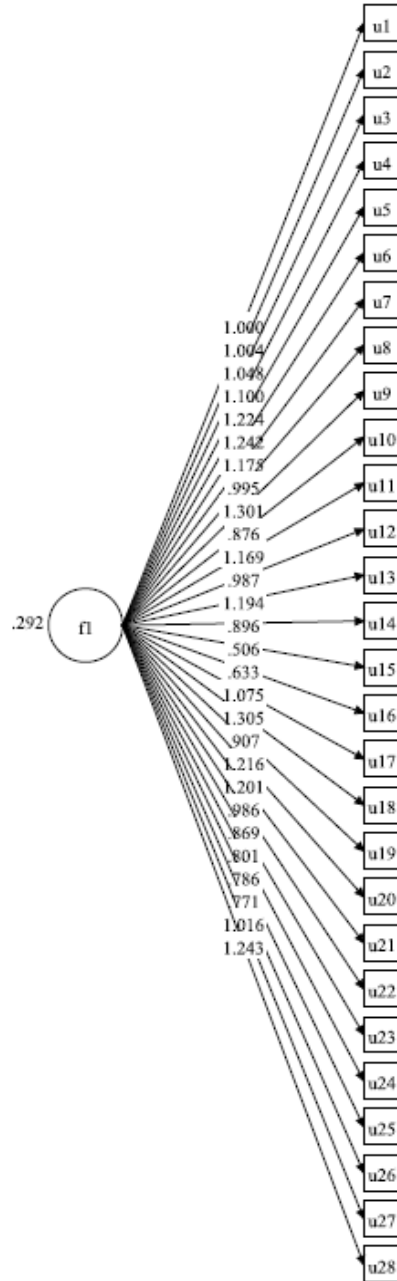
[u1\$1-u28\$1](T1-T28);

[f1*];

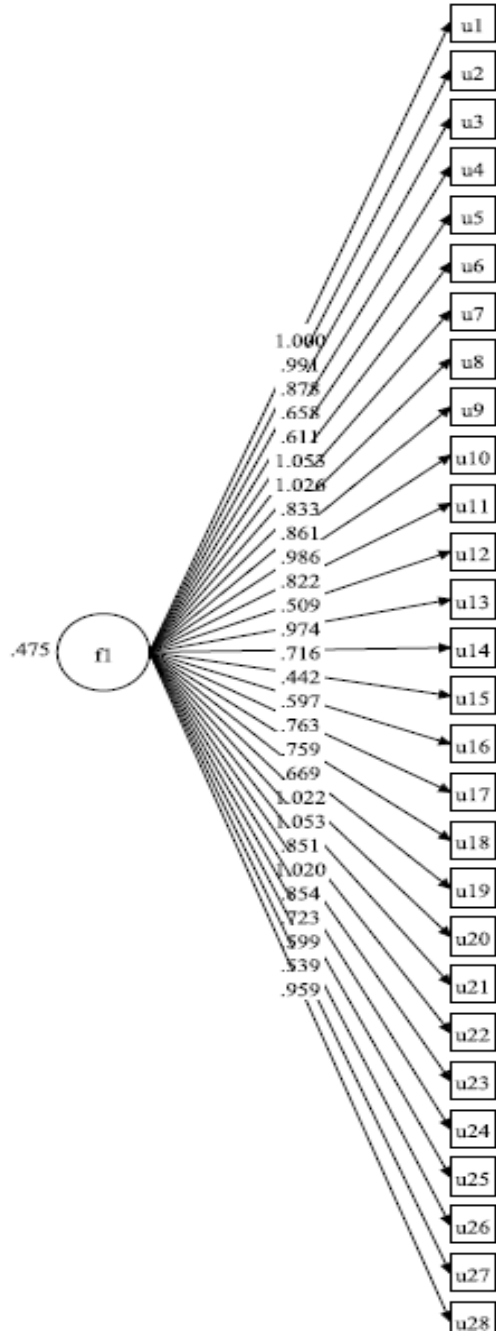
f1*;

EK 3. DOĞRULAYICI FAKTÖR ANALİZLERİNE İLİŞKİN YOL DİYAGRAMLARI

Ana Dili İngilizce Olan Grup

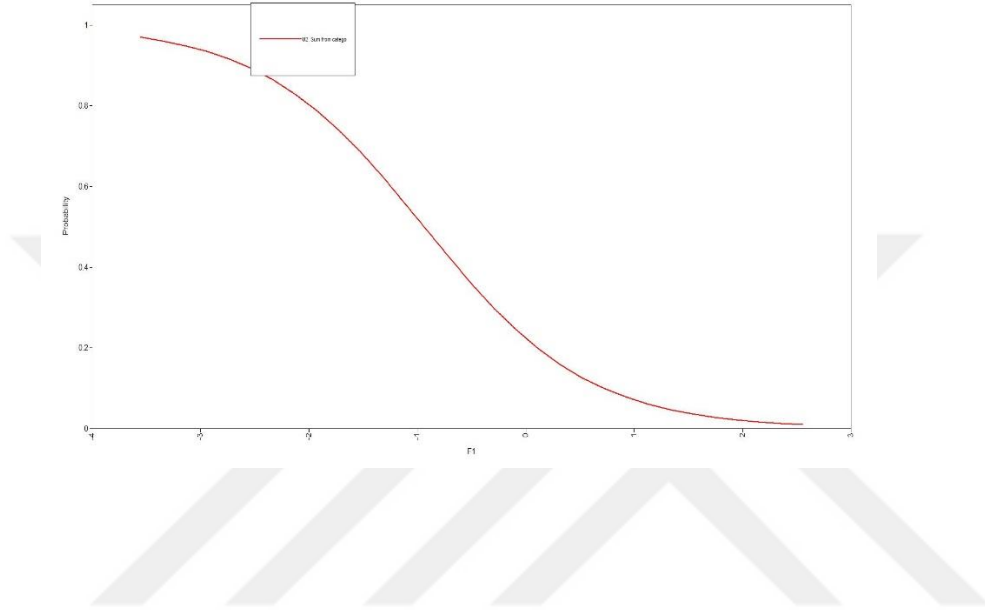


Ana Dili İngilizce Olmayan Grup

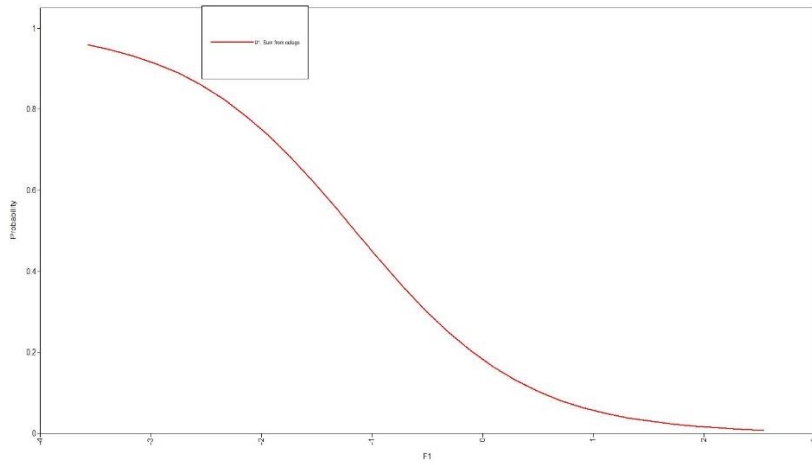


EK 4. ÖLÇME DEĞİŞMEZLİĞİNİ SINIRLAYAN BAZI MADDELERİN KARAKTERİSTİK EĞRİLERİ

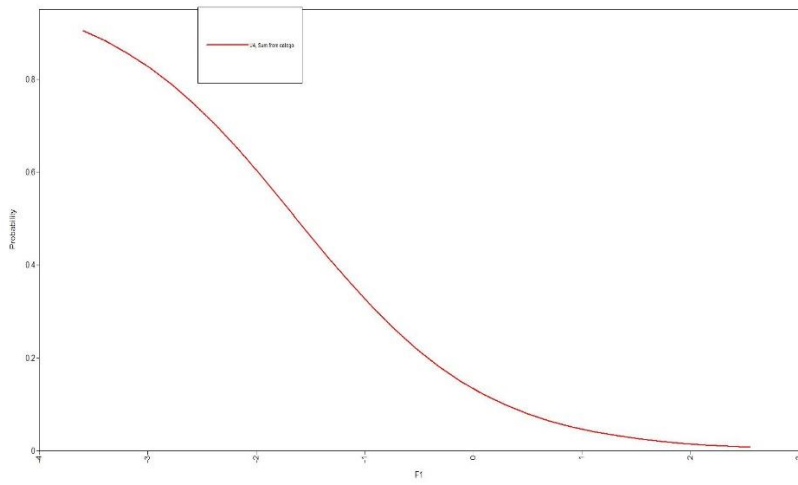
Madde 2.



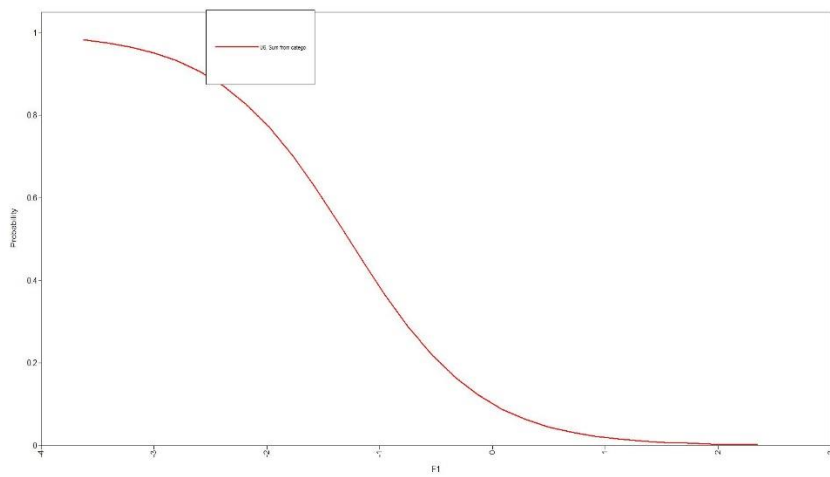
Madde 3.



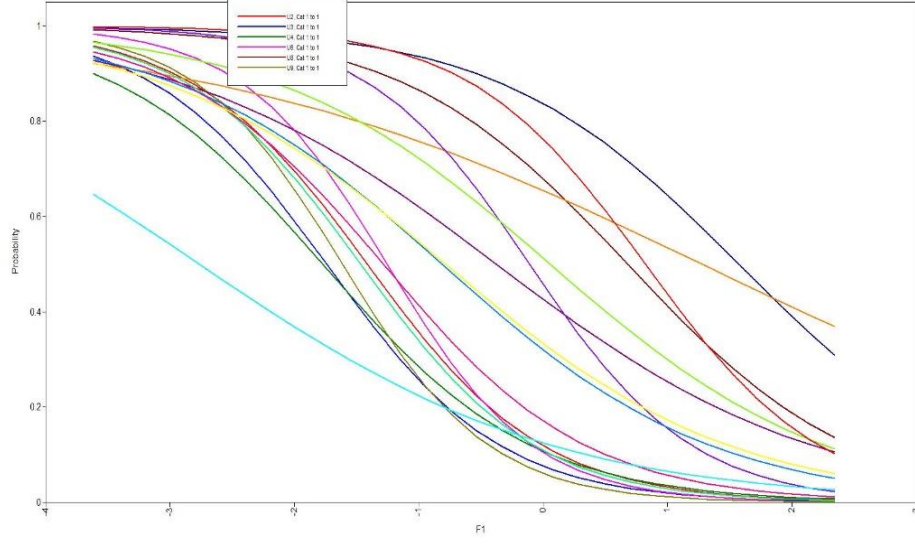
Madde 4.



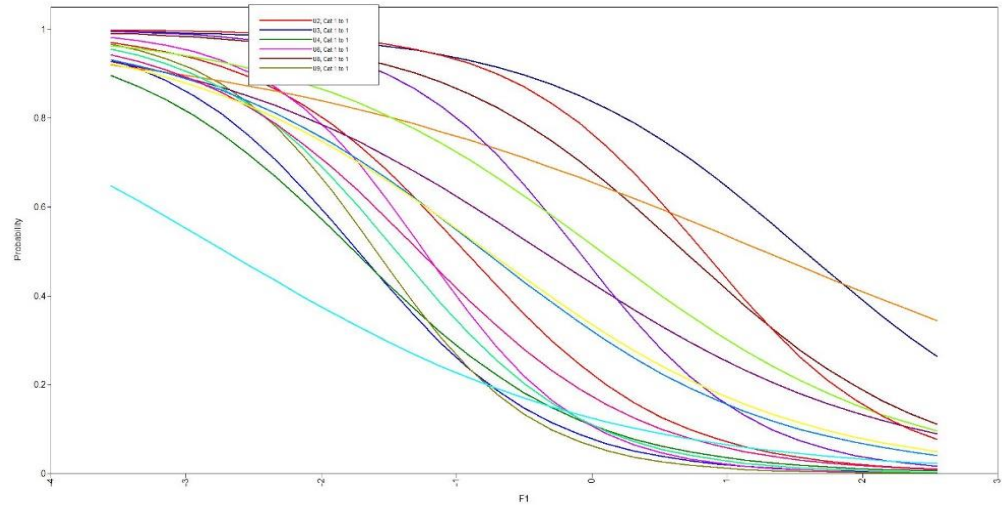
Madde 6.



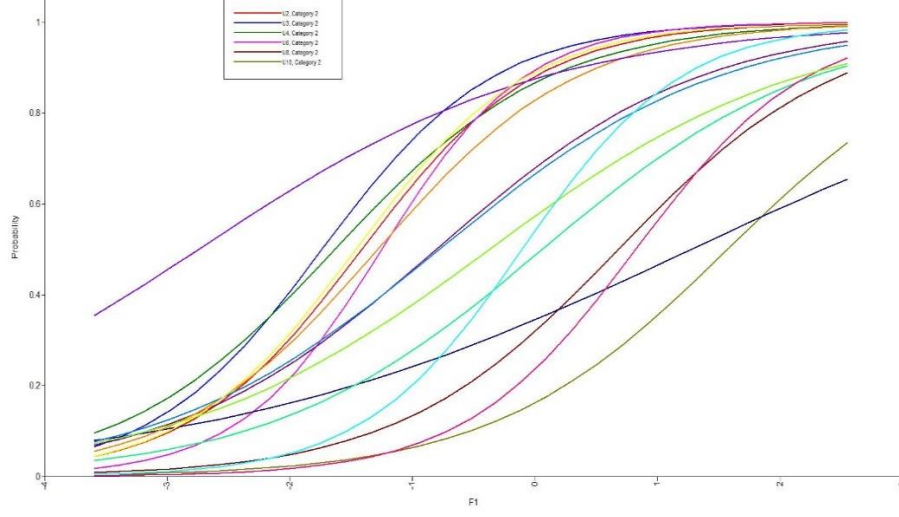
Nested Modele (M₀) Göre Ölçme Değişmezliğini Sınırlayan Maddelerin Karakteristik Eğrisi



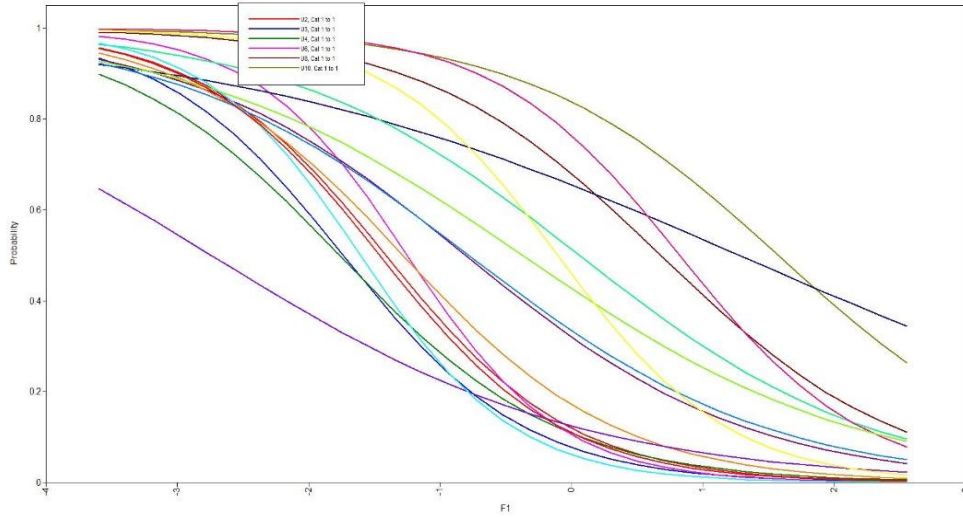
Model 2'ye Göre Ölçme Değişmezliğini Sınırlayan Maddelerin Karakteristik Eğrisi



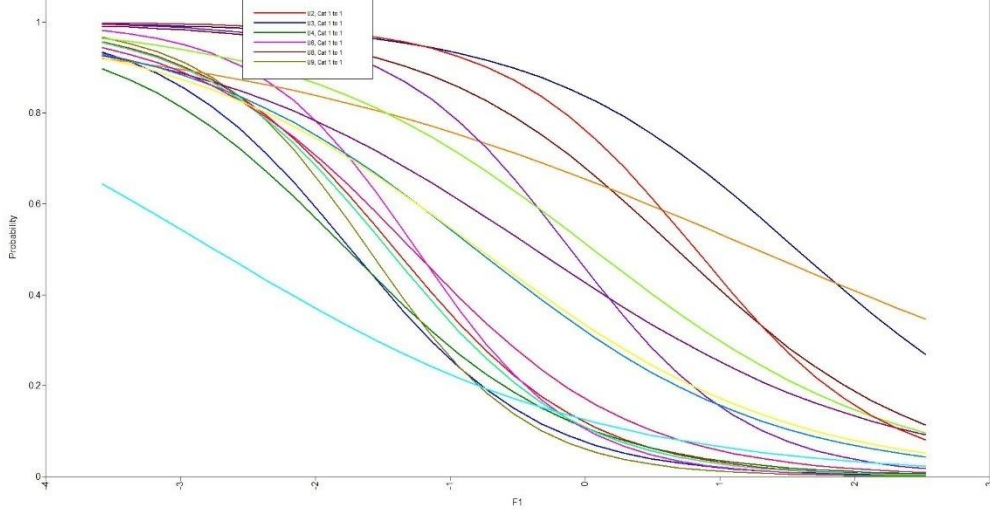
Model 4'e Göre Ölçme Değişmezliğini Sınırlayan Maddelerin Karakteristik Eğrisi



Model 5'e Göre Ölçme Değişmezliğini Sınırlayan Maddelerin Karakteristik Eğrisi



Model 29'a Göre Ölçme Değişmezliğini Sınırlayan Maddelerin Karakteristik Eğrisi



Model 30'a Göre Ölçme Değişmezliğini Sınırlayan Maddelerin Karakteristik Eğrisi

