

T.C.

AKDENİZ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANA BİLİM DALI

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME

TEZLİ YÜKSEK LİSANS PROGRAMI

**TIMSS 2015 DÖRDÜNCÜ SINIF MATEMATİK TESTİNİN ÖLÇME
DEĞİŞMEZLİĞİNİN ÜLKELERE GÖRE İNCELEMESİ**

YÜKSEK LİSANS TEZİ

Ayşenur TAVLICA

ANTALYA, 2019

T.C.

AKDENİZ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANA BİLİM DALI

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME

TEZLİ YÜKSEK LİSANS PROGRAMI

**TIMSS 2015 DÖRDÜNCÜ SINIF MATEMATİK TESTİNİN ÖLÇME
DEĞİŞMEZLİĞİNİN ÜLKELERE GÖRE İNCELEMESİ**

YÜKSEK LİSANS TEZİ

Ayşenur TAVLICA

Danışman:

Dr. Öğr. Üyesi Güçlü ŞEKERCİOĞLU

ANTALYA, 2019

DOĐRULUK BEYANI

Yüksek lisans tezi olarak sunduĐum bu çalıřmayı, bilimsel etik ve geleneklere aykırı düşecek bir yol ve yardıma başvurmaksızın yazdıĐımı, yararlandıĐım arařtırmaların kaynakçada gösterilenlerden oluřtuĐunu ve bu kaynakları her kullanımında alıntı yaparak yararlandıĐımı belirtir; bunu onurumla doĐrularım. Enstitü tarafından belli bir zamana baĐlı olmaksızın, tezimle ilgili yaptıĐım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara katlanacaĐımı bildiririm.

.../.../.....

Ayřenur TAVLICA

İMZA ONAY SAYFASI

T.C.

AKDENİZ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Ayşenur Tavlıca 'nın bu çalışması 05.07.2019... tarihinde jürimiz tarafından Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Lisans Programında Yüksek Lisans Tezi olarak oy birliği/oy çokluğu ile kabul edilmiştir

İMZA


Başkan : **Dr. Öğr. Üyesi Hakan KOĞAR**
Akdeniz Üniversitesi, Eğitim Fakültesi,
Eğitim Bilimleri Anabilim Dalı



Üye : **Dr. Öğr. Üyesi Neşe ÖZTÜRK GÜBEŞ**
Mehmet Akif Ersoy Üniversitesi, Eğitim Fakültesi,
Eğitim Bilimleri Anabilim Dalı



Üye (Danışman) : **Dr. Öğr. Üyesi Güçlü ŞEKERCİOĞLU**
Akdeniz Üniversitesi, Eğitim Fakültesi,
Eğitim Bilimleri Anabilim Dalı



YÜKSEK LİSANS TEZİNİN ADI: TIMSS 2015 DÖRDÜNCÜ SINIF MATEMATİK TESTİNİN
ÖLÇME DEĞİŞMEZLİĞİNİN ÜLKELERE GÖRE İNCELEMESİ

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

(Doç. Dr. Ramazan KARATAŞ)

Enstitü Müdürü

TEŐEKKÜR

Akademik alıőmalarımın bir baőlangıcı ve ilerleyen yıllarımda bana büyük getirileri olacağına inandığım bu alıőmamda bilgi birikimi, hayat tecrübesi, kiőiliđi ile her zaman örnek alacağım, güvenini hep yanımda hissettiđim deđerli tez danışmanım sayın Dr. Öğr. Üyesi Güçlü ŐEKERCİOĐLU' na yardımlarından ve bu tezin tamamlanmasında gösterdiđi titiz alıőmalarından dolayı őükranlarımı sunarım. Yüksek lisans eđitimim boyunca engin bilgilerinden, tecrübelerinden yararlandığım beni her konuda cesaretlendiren ve desteklerini hep arkamda hissettiđim deđerli hocalarım Dr. Öğr. Üyesi Hakan KOĐAR ve Do. Dr. Bayram BIAK' a sonsuz teőekkür ederim.

Hayatımın her anında ve aldıđım bütün kararlarda her zaman yanımda olan, beni destekleyen, alıőmalarım boyunca bilgisinden ve tecrübesinden yararlandığım hayat arkadaşıma ve daha çok küçük olmasına rağmen beni olgunlukla karşılayan ođluma sonsuz teőekkürlerimi sunarım. Son olarak bugünlere gelmemde en büyük emeđi olan canım annem ve babama sonsuz teőekkür ederim.

ÖZET

TIMSS 2015 DÖRDÜNCÜ SINIF MATEMATİK TESTİNİN ÖLÇME DEĞİŞMEZLİĞİNİN ÜLKELERE GÖRE İNCELEMESİ

Tavlıca, Ayşenur

Yüksek Lisans, Eğitim Bilimleri Ana Bilim Dalı

Tez Yöneticisi: Dr. Öğr. Üyesi Güçlü Şekercioğlu

Temmuz 2019, 64 Sayfa

Bu araştırmada, TIMSS 2015 uygulamasına ait dördüncü sınıf matematik başarı testinden elde edilen puanların ölçme değişmezliğinin OECD üyesi ülkelere göre incelenmesi amaçlanmıştır. Araştırmaya dahil edilen OECD üyesi ülkeler; Almanya, Amerika, Avustralya, Belçika, Çek Cumhuriyeti, Finlandiya, Hollanda, İngiltere, İrlanda, İspanya, İsveç, İtalya, Japonya, Kanada, Kore, Macaristan, Norveç, Polonya, Portekiz, Slovakya, Slovenya, Şili, Türkiye ve Yeni Zelanda'dır.

TIMSS 2015 uygulamasına 24 ülkeden toplam 132.226 öğrenci katılmıştır. İlişkisel tarama modelinde yürütülen bu araştırma, 7. Kitapçığı alan 9.641 öğrenci verisi üzerinden yürütülmüştür. Araştırma kapsamında, ölçme değişmezliği analizlerine geçmeden önce verilerin sağlaması gereken temel sayıtlara bakılmış ve ardından ülkelerin ölçme modellerinin doğrulanıp doğrulanmadığını belirlemek amacıyla doğrulayıcı faktör analizi yapılmıştır. Her gruba ait tek faktörlü ölçme modelinin elde edilen uyum indekslerine dayanarak doğrulandığı sonucuna ulaşılmıştır. Daha sonra çoklu-grup doğrulayıcı faktör analizi ile ölçme değişmezliği analizleri gerçekleştirilmiştir. Çoklu-grup doğrulayıcı faktör analizi bulguları hiyerarşik modellerin karşılaştırılması ile değerlendirilmiş olup bu modeller sırasıyla; yapısal değişmezlik modeli, zayıf faktöriyel değişmezlik modeli, güçlü faktöriyel

değişmezlik modeli ve katı faktöriyel değişmezlik modelidir. Bu dört modelin karşılaştırması yuvalanmış (nested) modeller arasında yapılmıştır.

Elde edilen bulgulara göre, modeller karşılaştırıldığında en iyi çalışan modelin güçlü faktöriyel değişmezlik modeli olduğu belirlenmiştir. Bu bulgudan yola çıkılarak TIMSS 2015 dördüncü sınıf matematik başarı testi puanlarının OECD üyesi ülkeler arasında ölçme değişmezliğini sağlamadığı, bu nedenle söz konusu ülkeler arasında karşılaştırılamayacağı sonucuna ulaşılmıştır.

Anahtar Sözcükler: Ölçme değişmezliği, Matematik başarıları, TIMSS 2015, Çoklu-grup doğrulayıcı faktör analizi, OECD

ABSTRACT

AN INVESTIGATION OF MEASUREMENT INVARIANCE FOR TIMSS 2015 FOURTH GRADE MATHEMATICS TEST ACCORDING TO COUNTRIES

Tavlıca, Ayşenur

Department of Educational Sciences, Postgraduate Thesis

Supervisor: Assoc. Prof. Dr. Güçlü Şekercioğlu

July 2019, 64 pages

In this study, it is aimed to examine the measurement invariance of the scores obtained from fourth grade mathematics achievement test of TIMSS 2015 among OECD countries. OECD countries included in the research are Germany, America, Australia, Belgium, Czech Republic, Finland, the Netherlands, England, Ireland, Spain, Sweden, Italy, Japan, Canada, Korea, Hungary, Norway, Poland, Portugal, Slovakia, Slovenia, Chile, Turkey and New Zealand.

Total number of students from 24 countries participated in the TIMSS 2015 application is 132,226. This research, which was conducted in relational survey model, was conducted on the data of 9,641 students who received the Booklet 7. Within the scope of the research, the main assumptions that the data should have before the measurement invariance analysis were examined and then the confirmatory factor analysis was performed to determine whether the measurement models of the countries were verified. It was concluded that the single factor measurement model of each group was validated on the basis of the fit indices. Then, multi-group confirmatory factor analysis and measurement invariance analyzes were performed. The findings of multi-group confirmatory factor analysis were evaluated by comparing hierarchical models; structural invariance model, weak factorial invariance model, strong

factorial invariance model and strict factorial invariance model. The comparison of these four models was performed using nested models.

According to the findings when the models are compared, the best working model was the strong factorial invariance model. Based on this finding, it has been concluded that TIMSS 2015 fourth grade math achievement test scores do not have measurement invariance among OECD member countries and therefore; the given countries cannot be compared in terms of students' performances.

Keywords: *Measurement invariance, mathematics achievement, TIMSS 2015, multi-group confirmatory factor analysis, OECD*

İÇİNDEKİLER

TEŞEKKÜR	iii
ÖZET	iv
ABSTRACT	vi
İÇİNDEKİLER.....	viii
TABLolar LİSTESİ	xi
KISALTMALAR LİSTESİ.....	xii

BÖLÜM I

GİRİŞ

1.1	Problem Durumu	1
1.2	Araştırmanın Amacı	5
1.3	Araştırmanın Önemi	5
1.4	Sınırlılıklar.....	6
1.5	Tanımlar.....	6

BÖLÜM II

KURAMSAL ÇERÇEVE İLE İLGİLİ ARAŞTIRMALAR

2.1	Uluslararası Durum Belirleme Çalışmaları	7
2.2	Geçerlilik	9
2.3	Faktör Analizi	10
2.3.1	Açımlayıcı Faktör Analizi.....	10
2.3.2	Doğrulayıcı Faktör Analizi	12

2.4	Ölçme Değişmezliği	14
2.4.1	Yapısal Değişmezlik (Configural Invariance)	15
2.4.2	Zayıf Faktöriyel Değişmezlik (Weak Factorial Invariance)	16
2.4.3	Güçlü Faktöriyel Değişmezlik (Strong Factorial Invariance).....	16
2.4.4	Katı Faktöriyel Değişmezlik (Strict Factorial Invariance).....	17
2.5	İlgili Araştırmalar	17

BÖLÜM III

YÖNTEM

3.1	Araştırma Deseni	22
3.2	Araştırma Verileri.....	22
3.3	Verilerin Düzenlenmesi	26
3.4	Verilerin Analizi	27

BÖLÜM IV

BULGULAR

4.1.	Doğrulayıcı Faktör Analizi Bulguları	32
4.2.	Kovaryans Matrislerinin Eşitliği Testi Bulguları	34
4.3.	Çoklu-Grup Doğrulayıcı Faktör Analizi Bulguları	34

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

5.1.	Sonuç ve Tartışma	37
------	-------------------------	----

5.2. Öneriler.....	40
KAYNAKÇA	41
EKLER.....	48
ÖZGEÇMİŞ.....	48
BİLDİRİM	49



TABLolar LİSTESİ

Tablo 3.1. Ülkelere Göre Örneklem Dağılımı	23
Tablo 3.2. TIMSS 2015 4.Sınıf Bilişsel Alanlara Göre Dağılımı	25
Tablo 3.3. TIMSS 2015 4.Sınıf Düzeyi Matematik Sorularının Öğrenme Alanlarına Göre Dağılımı	25
Tablo 3.4. Analizde Kullanılan Matematik Başarı Testi Maddeleri	26
Tablo 3.5. Ülkelerin Matematik Başarı Testi Puanları için Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları	28
Tablo 4.1. Ülkelerin Matematik Başarı Testi Puanları Doğrulayıcı Faktör Analizi Bulguları	32
Tablo 4.2. OECD Üyesi Ülkelere Ait Matematik Başarı Testi Puanlarının Kovaryans Matrislerinin Eşitliği Testi Analiz Bulguları	34
Tablo 4.3. OECD Üyesi Ülkelerin TIMSS 2015 4.Sınıf Matematik Başarı Testi Puanlarına ait Çoklu Grup Doğrulayıcı Faktör Analizi Bulguları	35

KISALTMALAR LİSTESİ

TIMSS: Trends in International Mathematics and Science Study

OECD: Organisation for Economic Co-operation and Development

IEA: International Association for the Evaluation of Educational Achievement

AFA: Açımlayıcı Faktör Analizi

DFA: Doğrulayıcı Faktör Analizi

ÇGDFA: Çoklu Grup Doğrulayıcı Faktör Analizi

YEM: Yapısal Eşitlik Modellemesi

S-B χ^2 : Satorra-Bentler Ki- Kare İstatistiği

BÖLÜM I

GİRİŞ

Bu bölümde problem durumu, araştırmanın amacı, alt problemler, araştırmanın önemi, sınırlılıklar ve kısaltmalara yer verilmiştir.

1.1 Problem Durumu

Hedefler, yaşam boyu çalışma ve gayretleri doğru biçimde yönlendirmeye yardımcı olur ve başarının hangi alanlarda olacağını bilgisini verir. Bu hedeflerin eğitimde kullanılma amacı ise öğrencilere hangi davranışların kazandırılacağını göstermesidir (Anderson ve Krathwohl, 2001).

Eğitim hedefleri zaman içerisinde özellikle 1950’li yıllardan itibaren büyük değişimlere uğramıştır. Bu değişimlerin en çok hissedildiği ülkelerin başında Amerika Birleşik Devletleri gelmektedir (Korkmaz, 2004). Ülkelerin statüsünü belirleyen en önemli unsurlardan biri olan bilim ve teknoloji toplumsal ilerlemenin başlangıç noktası olarak kabul edilmektedir (Bayır, Çakıcı ve Atalay, 2016). Bulduğumuz çağa damgasını vuran bilgi teknolojileri, dünyanın birçok ülkesinde pozitif bilimlerin eğitime verilen önemin ve kaynakların artırılmasını sağlamış ve özellikle teknolojilerini büyük bir hızla geliştiren ülkeler arasında bir yarış ortamı oluşmasına sebep olmuştur. Öğretim programlarını bu doğrultuda yenileyen ülkeler özellikle fen ve matematik eğitime ağırlık vermişlerdir (Çepni, Ayas, Johnson ve Turgut, 1997). Bu durum günümüzde de artarak devam etmektedir. İnsan kaynakları, ülkelerin gelişmesinin, yaşam standartlarının yükselmesinin; kalite ve verimin merkezinde bulunması sebebiyle bilgi toplumlarında oldukça önemli bir konumdadır (Ergül, 1999). Dolayısıyla insan kaynaklarının bilgi toplumlarında çağın gerektirdiklerine uygun donanımlara sahip olması amaçlandığından

bu durum doğrudan eğitimi ilgilendirmektedir. Teknolojinin sürekli gelişimi ve güncellenen bilgilerle eğitim ve öğretim programlarının da güncellenmesi kaçınılmazdır. Bu nedenle önemli bazı uluslararası kuruluşlar eğitim alanında araştırmalar yapmaktadır.

Pek çok ülke eğitim sistemlerine eleştirel bir bakış açısı geliştirmek ve sisteme dair eksikliklerini tespit ederek bu noktalarda gerekli çalışmaları yürütmek amacıyla çeşitli araştırmalar yürütmektedir (Millî Eğitim Bakanlığı [MEB], 2010). Değerlendiriciler bu araştırmaların sonuçlarını öğrencilerin performansını karşılaştırmak ve katılımcı ülkelerdeki eğitim politikaları ile uygulamaların etkililiğini değerlendirmek için kullanmaktadır (Gierl, 2000).

Türkiye'nin de katıldığı bu araştırmalardan biri Uluslararası Matematik ve Fen Eğilimleri Araştırmasıdır (Trends in International Mathematics and Science Study-TIMSS). Öğrencilerin matematik ve fen alanlarında kazandıkları bilgi ve becerilerin değerlendirilmesine yönelik bir tarama araştırması olan TIMSS (TIMSS, 2015), Hollanda merkezli Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu'nun (International Association for the Evaluation of Educational Achievement-IEA) bir projesidir. 1959'dan bu yana öğrenci eğitim başarıları ile ilgili uluslararası karşılaştırmalı çalışmalar yürütmekte olan IEA, ülkelerin birbirlerinden etkili eğitim yaklaşımları hakkında bilgi edinebilmelerine katkı sağlayan, ülkeler arası eğitim ve öğrenmede kullanılan farklı yöntemlere ilişkin uluslararası başarıya ilişkin çalışmalara öncülük etmiştir. TIMSS, dört yıllık aralıklarla dördüncü ve sekizinci sınıf düzeyindeki öğrencilere uygulanmaktadır. Her uygulamada okul ve sınıflar ülke genelini yansıtacak biçimde rastgele seçilmektedir. Ülkeler 4. ve 8. sınıf düzeyinde uygulanan sınava sadece 4. sınıf veya sadece 8. sınıf düzeyinde, ya da her iki düzeyde katılabilmektedir. Ayrıca, TIMSS' in dördüncü sınıf öğrencileri için çok zor olabileceği ülkelerde, TIMSS, beşinci veya altıncı sınıftaki öğrencilere ve sekizinci sınıf öğrencileri yerine dokuzuncu sınıfta da benzer biçimde verilebilir (TIMSS, 2015).

TIMSS' in genel olarak amacı, araştırmaya katılan ülkelerdeki dört ve sekizinci sınıf öğrencilerinin matematik ve fen alanlarındaki başarılarını ölçmek, eğitim ve öğretimin okullarda nasıl gerçekleştiğini, eğitim sisteminin etkinlik ve verimliliğini, ülkelerin eğitim sistemleri arasındaki farklılıkları belirlemek ve değerlendirmektir (TIMSS, 2015). TIMSS' in bir diğer amacı, matematik ve fen alanında öğretimi ve öğrenmeyi geliştirmek için kullanılacak önemli bir arka plan bilgisi sağlamaktır. Bu amaç doğrultusunda başarı testleri ve çeşitli anketler kullanılarak öğrencilerin fen ve matematik alanındaki performansları, eğitim sistemleri, öğretim programları, öğrenci özellikleri, öğretmen ve okulların karakteristik özellikleri ile ilgili bilgiler toplanmaktadır.

TIMSS ile düzenli aralıklarla öğrenci başarısındaki değişiklikler izlenmekte, yeni veya gözden geçirilmiş eğitim politikalarının başarıyı etkileyip etkilemediğini araştırılmaktadır. İlk uygulamalarının ardından açıklanan TIMSS ve PISA sonuçları katılımcı ülkeler için eğitim sistemlerinde köklü değişikliklerin başlatılmasına öncülük etmiştir. İngiltere'de 1997 yılında açıklanan ilk TIMSS sonuçlarının ardından hükümete bağlı kuruluşlar öğrencilerin güçlü ve zayıf yönlerini araştırmak için TIMSS verileri üzerinde çalışmaya başlamıştır. İsviçre'nin PISA' ya katılmadan önce düzenli olarak öğrenci performanslarını test etme amacıyla yürüttüğü çalışmalar ancak ilk PISA sonuçları açıklandıktan sonra somutlaştırılma aşamasına geçmiştir. Benzer şekilde Almanya'nın 1995 yılında açıklanan ilk PISA sonuçları öğretmenler, bilim insanları ve politikacılar için beklenenin düzeyin altında bulunmuştur. Bunun sonucunda politikacılara, PISA uygulamasından önce standart testlere dayanmayan Alman eğitim sistemini ulusal eğitim standartlarına ulaştırmayı hedefleyen bir reform kararı aldırılmıştır (Rutkowski, von Davier, & Rutkowski, 2013).

TIMSS başarı testlerinde yer alacak maddelerin geliştirilmesi ve teste alınma süreci, merkezi Boston Üniversitesi'nde olan TIMSS&PIRLS Çalışma Merkezindeki uzmanlar tarafından koordine edilmektedir. TIMSS' te yer alan matematik ve fen soruları önceden belirlenen

kazanımlar çerçevesinde ülke temsilcileri tarafından ortak olarak hazırlanmaktadır. Hazırlanan sorular IEA' nın fen ve matematik maddeleri inceleme komitesi tarafından incelenip açık uçlu sorular için puanlama anahtarları hazırlanmaktadır. Daha sonra oluşturulan yedek ve esas sorular taslak bloklar halinde incelenip sorulara son şekli verilmektedir. Hazırlanan sorular katılımcı ülkelerde çeviri ve uyarlama işlemlerinden sonra pilot uygulama ile test edilmektedir. Psikometrik özellikleri açısından yeterli olan sorular önceki uygulama soruları ile birleştirilerek pilot uygulamadan bir yıl sonra nihai uygulamada yer almaktadır. TIMSS 2015'te yer alan maddelerin yaklaşık yarısı çoktan seçmeli, yarısı uzun/kısa cevaplı sorulardan oluşmaktadır. Her iki sınıf düzeyinde de (4. ve 8. sınıf) fen ve matematik maddeleri 28 bloktan oluşmaktadır. Bu bloklardan 14'ü fen, 14'ü matematik bloklarıdır. Bu bloklar 14 test kitapçığına, ikisi fen ve ikisi matematik olmak üzere dörderli bloklar halinde dağıtılmıştır. Formlar arasında test eşitleme yapılabilmesi için fen ve matematik alanlarındaki her iki bloktan biri iki kitapçık arasında ortaktır (TIMSS, 2015).

Psikolojik araştırmalar, grupları genellikle psikolojik değişkenlerle karşılaştırır. Bir kültürel grupta gösterdiği psikometrik özellikler yeterli kabul edilen ölçme aracı başka bir kültürel gruba uyarlanarak uygulanır. Bu, kültürlerarası yürütülen çalışmalarda benimsenen temel yaklaşımlardan biridir. Araştırmacılar genellikle, ölçme aracının tüm kültürel gruplarda aynı yapıyı ölçtüğünü varsayar, ancak bu varsayımın test edilmesi gerekmektedir (Milfont ve Fischer, 2010). Kültürler arası karşılaştırmaların geçerliliği, psikoloji ve eğitimle ilgili araştırmalardaki birçok uygulama için hayati öneme sahiptir. Kültürler arası araştırmanın geçerlilik kanıtı farklı ülkelere elde edilen test puanlarının aynı yapıyı ölçmesidir (Wu, Li and Zumbo, 2007). TIMSS ülkelerin fen ve matematik başarılarını tek bir ölçme aracıyla değerlendirip karşılaştıran bir uygulamadır. Yapılan karşılaştırmaların ve puan analizlerinin anlamlılığında bahsedebilmek için kullanılan ölçme aracının farklı gruplarda aynı özelliği ölçtüğünün doğrulanması gerekmektedir (Uzun ve Öğretmen, 2010). Gruplar arasında ölçme

değişmezliğinin sağlanması, çapraz grup karşılaştırmaları yapmak için mantıksal bir gerekliliktir, ancak ölçme değişmezliği örgütsel araştırmalarda nadiren test edilmiştir (Vandenberg ve Lance, 2000). Mark ve Wan'a (2005) göre ölçme değişmezliğinin varlığının kanıtlanmadığı durumlarda yapılan çıkarımlar bilimsel bir nitelik taşımayacağı gibi, gruplar arası farklılıkları yorumlayabilmek mümkün olmamaktadır.

1.2 Araştırmanın Amacı

Bu araştırmanın amacı, TIMSS 2015 dördüncü sınıf matematik başarı testinden elde edilen puanların ölçme değişmezliğinin Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) üyesi ülkeler için test edilmesidir.

1.3 Araştırmanın Önemi

Eğitim sistemindeki düzeltilmesi gereken alanlara dair önlemlerin alınmasını sağlayan değerlendirme programları ülkeler için büyük bir anlam ifade etmektedir. Bu da söz konusu değerlendirme programları üzerinden yürütülen araştırmaların çoğalmasına neden olmaktadır.

TIMSS dört yılda bir gerçekleştirilmektedir ve dördüncü sınıf öğrenci grubuna dört yıl sonra sekizinci sınıfta da değerlendirme fırsatı sunmaktadır. Dördüncü sınıf öğrencilerinin değerlendirilmesine ihtiyaç duyulan öğretim programları için oldukça erken bir uyarı niteliği taşıyabilmekte ve bu yeniliklerin sekizinci sınıfa gelindiğinde etkililiği değerlendirilebilmektedir. Ayrıca, TIMSS öğrencilerin öğrenim faaliyetlerini etkileyen faktörleri belirlemek ve öğrenci hakkında bilgi toplamak için uyguladığı anketlerden elde ettiği verilerle ülkelerin öğretim programlarına yön vermektedir. Ancak TIMSS ve PISA gibi geniş ölçekli test uygulamalarına katılan ülkelerin öncelikli olarak dikkate aldığı veriler ülke bazlı gerçekleştirilen sıralamalar olmaktadır. Karşılaştırmaların yapıldığı puanların elde edildiği başarı testlerinin her bir grup için eş ölçekler olduğuna dair kanıt elde edilmeden

yapılan sıralamanın geçerli olduğunu söylemek mümkün değildir. Dolayısıyla bu araştırma, katılımcı ülkelerin tamamında yapılan ölçme işleminin denk olup olmadığının belirlenmesi açısından ölçme değişmezliği çalışmalarının yapılması gerekliliğini ortaya koymaktadır.

İlgili araştırmalar incelendiğinde, ölçme değişmezliği çalışmalarının daha çok PISA uygulamaları üzerinde yapıldığı, TIMSS uygulamalarının incelendiği ölçme değişmezliği çalışma sayısının nispeten daha az olduğu görülmüştür. Ayrıca araştırmanın yürütüldüğü veri setinin büyüklüğü göz önünde bulundurulduğunda, bu çalışmanın alanyazına önemli bir katkı sağlayacağı öngörülmektedir.

1.4 Sınırlılıklar

Bu araştırma TIMSS 2015 uygulamasında yer alan testlerden yalnızca dördüncü sınıflara uygulanan matematik başarı testi ile sınırlandırılmıştır.

Model karşılaştırmaları TIMSS 2015 uygulamasında yer alan 7 numaralı kitapçık (matematik başarısına ilişkin 10'u açık uçlu, 15'i çoktan seçmeli olmak üzere 25 adet soru) ile sınırlandırılmıştır.

Araştırma, TIMSS 2015 uygulamasına katılan 26 OECD ülkesinden 141.109 katılımcı ile başlatılmış, ancak doğrulayıcı faktör analizi aşamasında analiz sonuçlarında uyum indeksleri açısından sayıtların gerektirdiği ölçütlerin elde edilememesi sebebiyle Danimarka ve Fransa analizinin dışında bırakılmıştır.

1.5 Tanımlar

Ölçme değişmezliği, yapılan ölçme işleminin her bir alt grup için eşit olmasıdır.

BÖLÜM II

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde araştırmanın kuramsal temeli kapsamında uluslararası durum belirleme çalışmaları, geçerlilik, faktör analizi, açımlayıcı faktör analizi, doğrulayıcı faktör analizi, ölçme değişmezliği ile ilgili bilgilere ve ölçme değişmezliği ile ilgili araştırmalara yer verilmiştir.

2.1 Uluslararası Durum Belirleme Çalışmaları

Gelişen teknoloji ile birlikte çağın gerekliliklerine ayak uydurmaya çalışan ülkelerde, farklı rekabet alanları ortaya çıkmaktadır. Özellikle eğitim alanında kendini gösteren küresel rekabet, ülkelerin uyguladıkları veya uygulamayı tasarladıkları eğitim sistemlerini sürekli olarak gözden geçirmeyi zorunlu hale getirmektedir (MEB, 2014). Bu kapsamda temel eğitim seviyesinde yapılan değerlendirmelerin geleceğe yönelik daha sağlıklı yordamalar yapılmasına yardımcı olacağı düşünülmekte ve bu seviyede eğitim gören öğrenciler için ölçme değerlendirme çalışmalarına ağırlık verilmektedir.

Eğitim sistemlerinin faydalı ve verimli olup olmadığının kararı verilirken, çoğu zaman öğrencilere uygulanan standart testlerden elde edilen başarı puanları esas alınmaktadır (Çakan, 2003). Bu amaçla, uluslararası bazı kuruluşlar periyodik olarak her kıtadan birçok ülkenin katılımının sağlandığı geniş ölçekli test uygulamaları gerçekleştirmektedir. Bunlar arasında sonuçları en çok göze çarpan ve ülkelerin eğitim reformlarını şekillendiren sınavlardan bazıları aşağıda açıklanmıştır.

PISA (Uluslararası Öğrenci Değerlendirme Programı): Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından, üçer yıllık periyotlarla 15 yaş grubu öğrencilere uygulanmaktadır.

PISA, öğrencilerin okulda öğrendikleri bilgi ve becerilerin günlük hayatta kullanma yetilerini ölçmeyi amaçlamaktadır. Öğrencilere bu doğrultuda; matematik okuryazarlığı, fen bilimleri ve okuma becerileri konu alanlarında başarı testleri uygulamanın yanı sıra, öğrenme ortamları ve aileleriyle ilgili bilgiler toplamak için anketler uygulamaktadır.

Ülkemiz de bir OECD üyesi olarak, 2000 yılından itibaren gerçekleştirilen PISA uygulamalarına ilk kez 2003 yılında katılmıştır. Elde edilen sonuçlar ulusal rapor halinde düzenlenip yayınlanmaktadır.

PIRLS (Uluslararası Okuma Becerilerinde Gelişim Projesi): Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu'nun (IEA) bir projesi olan PIRLS, her beş yılda bir 10 yaş grubu 4.sınıf öğrencilerine uygulanmaktadır. Bu uygulamanın amacı standart test ve anketlerle öğrencilerin dil becerilerini, okuma becerilerini ve okuma alışkanlıklarını belirlemektir. 2001 yılından itibaren yapılan uygulamalara Türkiye yalnızca 2001 yılında katılmıştır.

TIMSS (Uluslararası Matematik ve Fen Eğilimleri Araştırması): Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu'nun (IEA) bir diğer önemli projesi olan TIMSS, her dört yılda bir 4. ve 8.sınıf öğrencilerine uygulanmaktadır. 1995'ten itibaren uygulanan TIMSS' in temel amacı, öğrencilerin matematik ve fen bilimleri alanlarında kazandıkları bilgi ve becerileri değerlendirmektir. Matematik ve fen alanlarında çoktan seçmeli ve açık uçlu sorulardan oluşan standart başarı testlerinin kullanılmasının yanı sıra; öğrenme ortamları, öğrenci ve öğretmen özellikleri hakkında veri toplamak amacıyla da anketler uygulanmaktadır. 60'dan fazla ülkenin katılım gösterdiği uygulamaya Türkiye ilk kez 1999 yılında katılmıştır.

Geniş çaplı katılımın gerçekleştiği uluslararası durum belirleme çalışmaları hem ülkelerin mevcut eğitim sistemlerinin verimliliğinin gözden geçirilmesi, hem de geliştirilen yeni eğitim yaklaşımlarının yeterliliğinin değerlendirilebilmesi için önem arz etmektedir.

Geniş ölçekli test uygulamalarının standart olması yapılan karşılaştırmaların anlamlılığı açısından zorunludur. Standart testlerde bulunması gereken en önemli özellik ise güvenilirlik ve geçerliliğinin kanıtlanmış olmasıdır. Geçerlilik ile ilgili alanyazına aşağıda yer verilmiştir.

2.2 Geçerlilik

Geçerlilik, testin ölçülmek istenilen özelliği ne kadar doğru ölçtüğüyle ilgilidir. Bir başka ifadeyle geçerlilik, ölçmeyi gerçekleştirebilme derecesidir (Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demirel, 2008). Aynı zamanda geçerlilik, ölçmeden elde edilen sonuçların evrene genellenebilir olmasıdır (Şencan, 2005). Bu bağlamda, bir evreni temsil eden örneklem üzerinden yapılan çalışmalarla evren hakkında genellemelere ulaşma amacındaki araştırmalarda geçerliliğin sağlam kanıtlara dayandırılması zorunluluğu ortaya çıkmaktadır. LoBiondo-Wood ve Haber (2014)'e göre geçerlilik, bir ölçme aracının bir kavramın niteliklerini ne kadar doğru ölçtüğünün bir ölçüsüdür.

İlgili alanyazın incelendiğinde geçerlilik türleriyle ilgili farklı gruplamalara rastlansa da genellikle geçerlilik türlerinin üç başlık altında toplandığı görülmektedir. Bunlar, ölçüt geçerliliği, kapsam geçerliliği ve yapı geçerliliğidir. Ölçme aracının türüne göre bazı geçerlilik türlerinin daha kullanışlı olacağı görüşü yaygın olsa da, yapı geçerliliğinin diğer geçerlilik türlerini kapsayıcı nitelikte olduğu görüşü de araştırmacılar tarafından savunulmaktadır (Messick, 1995). Dolayısıyla yapı geçerliliği kanıtlarının elde edilmesi, ölçme aracının geçerliliğinin sağlandığı yönündeki bulguları güçlendirmektedir. Yapı geçerliliğini incelemek için en sık başvurulan tekniklerden biri faktör analizidir. Aşağıdaki bölümde faktör analizi ilgili bilgiler yer almaktadır.

2.3 Faktör Analizi

Faktör analizi, ölçeklerin yapı geçerliliği kanıtlarını ortaya koymak için arařtırmacıların en sık başvurduđu tekniklerden biridir (Çokluk, Şekerciođlu ve Büyüköztürk, 2010). Ölçeklerin faktör yapısını ortaya koyan bu teknik diđer analizlere kıyasla daha geniş kapsamlı bilgi vermektedir. Faktör analizi birbiriyle ilişkili, bir diđer ifadeyle benzer yapıları ölçen deđişkenleri bir araya getirerek gruplama işlemini yapan ve daha az sayıdaki bu gruplarla ölçme işlemini açıklamaya çalışan çok deđişkenli bir istatistiksel yöntemdir (Büyüköztürk, 2002).

Byrne (1998) veri analizinde faktör analizi kullanımını, altta yatan gizli yapılar hakkında bilgi toplamak ve gözlenen deđişkenler arasındaki kovaryansı incelemek için kullanılan en eski tekniklerden biri olarak tanımlamıştır. Faktör analizi çok deđişkenli bir veri kümesinin karmaşıklığını azaltmayı sağlayan bir teknik olup, arařtırmacılara uygulama yapma ve teori geliřtirmede kolaylıklar sağlamaktadır (Huck, 2007).

Faktör analizinde açımlayıcı (exploratory) ve dođrulayıcı (confirmatory) olmak üzere iki çeřit yaklaşım vardır. Açımlayıcı faktör analizinde veriler üzerinde kesin bir model oluřturmadan ölçme aracının karakteristik özelliklerini tespit etmek amaçlanırken, dođrulayıcı faktör analizinde göreceli olarak az sayıda parametre açısından deneysel verileri açıklamak ve tanımlamak için bir model oluřturulması amaçlanmaktadır (Jöreskog and Sörbom, 1993).

2.3.1 Açımlayıcı Faktör Analizi

Açımlayıcı faktör analizi (AFA) sosyal bilimlerde ölçeklerin yapı geçerliliğini incelemeye yaygın olarak kullanılan ve uygulanan istatistiksel bir tekniktir. Temelde faktör analizi maddelerin faktör adı verilen yapılar altında gösterdikleri ilişkiye dair bilgi verir (Çokluk ve diđerleri, 2010). Açımlayıcı faktör analizi gözlenen deđişkenlerle gizil deđişkenler arasındaki bađlantıların bilinmediđi durumlar için tasarlanmıştır. Analizin amacı gözlenen deđişkenlerin

altta yatan faktörlerle nasıl ve ne derecede ilişkili olduğunu ortaya koyma ve gözlenen değişkenlerin ortaklaştığı asgari faktör sayısını belirlemektir (Byrne, 1998).

Açımlayıcı faktör analizi bir grup değişken arasındaki ortak kovaryansı daha anlaşılır kılar çünkü faktörler ölçülen değişkenlerden daha az sayıdadır (Brown, 2006). Bu yöntem araştırmanın başında karşılaşılan; gözlenen değişkenlerin belirli yapılar altında toplanıp toplanmadığı, kaç adet yapının ortaya çıkacağı ve bu yapıların niteliklerinin nasıl olacağı gibi sorulara yanıtlar bulunmasına yardımcı olmaktadır.

Faktör analizinde, faktör belirleme veya faktör çıkarma için kullanılan birden çok teknik bulunmaktadır:

- a) Temel bileşenler analizi: Faktör analizinde gözlenen değişkenleri temsil eden ve faktör gruplanmalarını açıklayan bağımsız bileşenlerin ortaya çıkarılmasına dayanan bir istatistiki tekniktir.
- b) Ortak faktör analizi: Üç veya daha çok değişkenin etkilendiği ortak faktörlerin altında yatan gizli yapıyı açıklamak ve daha önceden belirlenen faktör yapısını doğrulamak amacıyla kullanılan bir yöntemdir.
- c) Maksimum olasılık yöntemi: Uygun teknikle faktör analizi yapıldıktan ve faktör yapılarına dair bilgi edinildikten sonra başvurulan bu yöntem, değişkenleri daha iyi açıklayan farklı bir modelin kurulup kurulamayacağı hakkında bilgi verir. Bu yönüyle teyit edici faktör analizini andırmaktadır.
- d) Ağırlıklandırılmamış en küçük kareler yöntemi: Gözlemlenen ve yeni oluşturulan korelasyon matrisleri arasındaki farkların karesini en düşük seviyeye getirmeyi amaçlar.
- e) Genelleştirilmiş en küçük kareler yöntemi: Bu yöntem de gözlemlenen ve yeni oluşturulan korelasyon matrisleri arasındaki farkların karesini en düşük seviyeye

getirmeyi amaçlar, ancak bu yöntemde ağırlıklandırılmamış en küçük kareler yönteminden farklı olarak her bir değişken asıl faktör değeri ile ağırlıklandırılarak işleme girer.

f) İmaj faktör analizi yöntemi: Olası faktörlerin birbirleriyle zıt olduklarının tahmin edildiği çalışmalarda faktör çıkarmada kullanılması önerilen alternatif bir yöntemdir.

g) Alfa faktör analizi: Bu yöntem faktörlerin güvenilirlik katsayılarını en üst düzeye getirmeyi amaçlar. (Şencan, 2005).

2.3.2 Doğrulayıcı Faktör Analizi

Doğrulayıcı faktör analizi (DFA), gözlemlenen bir kümenin faktör yapısını doğrulamak için kullanılan istatistiksel bir tekniktir. Ayrıca, gözlenen değişkenler ile altta yatan gizli yapılar arasında ilişki olduğuna dair hipotezin test edilmesini sağlar (Suhr, 2006). Açımlayıcı faktör analizi veride açıkça görülemeyen değişkenler arasındaki korelasyonların bile görülmesini sağlar, bu yönüyle bir keşif analizidir. Diğer yandan, teorik açıdan olması beklenen ilişkilerin gerçekte de verilerde görünüp görünmediğini test etmeyi sağlar (Jöreskog and Sörbom, 1993). Şencan'a göre (2005), doğrulayıcı faktör analizinde gözlenen değişkenlerin gizil yapılarla ve gizil yapıların da birbirleriyle olan ilişkisi ile alakalı tüm öngörüler araştırmacının dayandırdığı kuramsal temelle bağlantılıdır. Bu doğrultuda elde edilen verilerin kurama bağlı kalınarak geliştirilen modelle uyum sağladığı gösterilir. Bir diğer deyişle doğrulayıcı faktör analizi teyit edici faktör analizidir.

Doğrulayıcı faktör analizi, modelin tüm özelliklerinin önceden bilinmesini gerektiğinden açımlayıcı faktör analizinden farklıdır, dolayısıyla faktör modelinin değerlendirilmesi için güçlü bir ampirik temel gereklidir. Bu nedenle, çoğunlukla açımlayıcı faktör analizi uygulaması yapı geçerliliği doğrulaması sürecinde önceki aşamalarda kullanılır; doğrulayıcı

faktör analizi ise daha önce kurulan deneysel ve teorik temeller üzerine uygulanır (Brown, 2006).

Brown'a (2006) göre doğrulayıcı faktör analizinin en yaygın kullanım alanlarından bazıları şunlardır:

- a) Testlerin Değerlendirilmesi: Doğrulayıcı faktör analizi, bir ölçme aracının gizil yapısını incelemek için kullanılmaktadır. Ayrıca testin nasıl puanlanacağına da yardımcı olmaktadır. Tek faktörlü yapınının yanı sıra çok faktörlü yani, iki ya da daha fazla faktörlü yapılarda alt ölçeklerin nasıl puanlanması gerektiğini gösterir. Psikometrik değerlendirmenin birçok alanında kullanılabilen doğrulayıcı faktör analizi, ölçek geçerliliği hakkında da bilgi vermektedir.
- b) Yapı Doğrulama: Doğrulayıcı faktör analizi sosyal bilimlerde yapı doğrulama için vazgeçilmez bir tekniktir. Analiz, yapı geçerliliğini değerlendirmek için kullanılan yöntemlerden biri olarak kabul edilen yakınsak (convergent) ve ıraksak (discriminant) geçerlilik adına sağlam kanıtlar sağlayabilmektedir. Yakınsak geçerlilik maddelerin ait oldukları faktörle ve faktöre ait diğer değişkenlerle yüksek korelasyona sahip olduğunu, ıraksak geçerlilik ise maddelerin ait oldukları faktör haricindeki faktörlerle düşük korelasyona sahip olduğunu öngörmektedir (Yaşlıoğlu, 2017). Yapı doğrulamada doğrulayıcı faktör analizinin temel rolü, ölçme hatalarının geçerlilik sonuçlarında hesaba katılmasıdır. Bu yönüyle doğrulayıcı faktör analizi geleneksel yöntemlerden daha güçlü bir kanıt sağlar.
- c) Yöntem Etkileri: Çoğu zaman, gözlenen ölçümlerin bazı kovaryansları gizil değişkenlerden farklı sebeplerle ortaya çıkmaktadır. Ölçme yaklaşımları ile göstergeler arasında ilave bazı değişkenler görüldüğü zaman yöntem etkisinden bahsedilebilir. Yöntem etkisi genellikle olumlu ve olumsuz olarak ifade edilen cümlelerin kombinasyonlarını içeren araçlarda görülmektedir. Açımlayıcı faktör

analizi yöntem etkilerini ortaya çıkarmada yetersiz kalırken doğrulayıcı faktör analizi, yöntem etkilerini ölçme modeline ait hata teorisinin bir parçası olarak ele alır.

- d) Ölçme Değişmezliği Değerlendirmesi: Ölçme değişmezliğini araştırmak test geliştirmenin ve uygulamanın önemli bir parçasıdır, ölçme özelliklerinin değişmezliğinin cinsiyet, ırk vs. gibi etkenler için tespit edilmesi gereklidir. Değişmezliğin tespit edilememesi durumunda o testin yanlılığından şüphelenilir. Bu tip sorunlar çoklu grup doğrulayıcı faktör analizinde ele alınabilir. Bu analizde çeşitli alt gruplarda ölçme modeli aynı anda kestirilebilmektedir.

2.4 Ölçme Değişmezliği

Ölçme değişmezliği testinin önemi 50 yıldan uzun bir süre önce alanyazına girmiştir ancak, değişmezlik testi için kullanılan istatistiksel teknikler daha çok son yıllarda ulaşılabilir hale geldiğinden, araştırmacıların ölçme değişmezliğini test etmesi günümüzde daha beklenir hale gelmiştir (Putnick and Bornstein, 2016). Faktör yapılarının eşitliği hakkında yazan ilk araştırmacı Jöreskog (1971) olmuştur. Ölçme değişmezliği kavramı ise, Byrne, Shavelson ve Muthe'n (1989) tarafından tanıtılmış ve daha sonra test edilmiştir.

Ölçme değişmezliği, gözlenen değişkenlerle gizil değişkenlere yani yapılara ulaşmaya çalışan ölçeklerin sahip olması gereken psikometrik özelliklerden biridir. Aynı yapı, ülkeler, kültürel birimler, zaman hususları veya ülkeler içindeki bölgeler gibi farklı gruplar arasında aynı şekilde ölçüldüğünde, ölçümün gruplar için değişmez olduğu sonucuna varılır (Horn and McArdle, 1992; Meredith, 1993; Vandenberg and Lance, 2000). Ölçme değişmezliği sağlandığında gizil değişken üzerinde yapılan gruplar arası karşılaştırmaların geçerliliği sağlanabilir; aynı zamanda gelecekte yapılacak olan çalışmalarla gizil değişken puanlarının oluşumu, belirleyicileri ve sonuçları da kıyaslanarak değerlendirilebilir (Schoot, Lugtig and Hox, 2012).

Horn ve McArdle 'a göre (1992) yaş, cinsiyet veya kültür gibi değişkenlerin yer aldığı her çalışmada gerekli çıkarım ve yorumların yapılabilmesi için ölçme değişmezliği gereklidir. Ölçme değişmezliğinin temel sorusu, olguların farklı koşullar altında gözlemlenmeleri sonucunda ölçülen niteliğin aynı ölçme sonuçlarını verip veremediğidir. Eğer değişmezliğin varlığına dair bir kanıt elde edilemediyse bireyler ve gruplar arasındaki farklılıklara dair bilimsel çıkarımlar yapabilmek mümkün olmamaktadır.

Ölçme değişmezliği çoklu grup doğrulayıcı faktör analizi ile dört adet hiyerarşik modelin test edilmesi ile ortaya konmaktadır. Bu modeller sırasıyla “yapısal değişmezlik (configural invariance)”, “zayıf faktöriyel değişmezlik (metric invariance= weak factorial invariance)”, “güçlü faktöriyel değişmezlik (strong factorial invariance= scalar invariance)” ve “katı faktöriyel değişmezlik (strict factorial invariance)” olarak adlandırılmaktadır. Aşamalılık içeren bu dört modelin test edilmesi gereklidir (Meredith, 1993). Zayıf teknikten güçlü tekniğe doğru gerçekleştirilen değişmezlik testleri çoklu grup doğrulayıcı faktör analizleri ve yapısal eşitlik modellemeleri kullanılarak gösterilir (Horn ve McArdle, 1992).

2.4.1 Yapısal Değişmezlik (Configural Invariance)

Ölçme değişmezliğindeki en temel seviyedir. Yapısal değişmezlik test edilirken faktör yükleri, faktör varyansları ve hata varyansları serbest bırakılır. Bu aşamada test edilen hipotez ölçme aracının faktör yapısının gruplar arasında değişmez olduğudur. Grupların aynı faktör yapısına sahip olup olmadıkları incelenir. Eğer yapısal değişmezlik kanıtı elde edilemezse, bu durum gruplarda farklı yapıların ölçüldüğü anlamına geldiğinden (Wu, Li ve Zumbo, 2007), sonraki aşamalarda hipotezlerin test edilmesinin anlamı olmayacaktır.

2.4.2 Zayıf Faktöriyel Değişmezlik (Weak Factorial Invariance)

Aşamalılık gereği yapısal değişmezliğin doğrulanması durumunda zayıf faktöriyel değişmezliğin testine geçilebilmektedir. İlk kez parametre sınırlamasına geçilen bu modelde sadece faktör yükleri serbesttir. Bu aşamada test edilen hipotez, gizil değişkene ait ölçme aracını oluşturan maddelere ait faktör yüklerinin gruplar arasında değişmez olduğudur. Zayıf faktöriyel değişmezliğin sağlanması durumunda tahmini faktör varyanslarının ve kovaryanslarının grupsal karşılaştırmaları savunulabilir. Benzer bir ifadeyle bir madde zayıf faktöriyel değişmezlik şartını sağlarsa madde üzerindeki fark puanları ülkeler arasında anlamlı bir şekilde karşılaştırılabilir ve bu farklılıklar benzer yapıdaki ülkeler arası farklılıkların da göstergesidir (Steenkamp ve Baumgartner, 1998).

Zayıf faktöriyel değişmezliğin desteklenmemesi, ortak grupların bir veya daha fazlasının veya maddelerin en azından bir alt kümesinin, gruplar arasında farklı anlamlara sahip olduğu anlamına gelmektedir (Gregorich, 2006).

2.4.3 Güçlü Faktöriyel Değişmezlik (Strong Factorial Invariance)

Ölçek değişmezliği olarak da adlandırılan güçlü faktöriyel değişmezlik modelinde, bir ölçme aracına ait maddelerin faktör puanlarının sıfır olması durumunda elde edilen regresyon sabitinin gruplar arasında değişmez olduğu hipotezi test edilir. Bu aşama örtük ortalamaların karşılaştırılabilmesi için önemlidir. Güçlü faktöriyel değişmezliğin sağlanması gözlenen puanların gizil puanlarla ilişkili olduğu anlamına gelir; yani, gizil yapıda aynı puana sahip olan bireyler hangi gruptan olursa olsun, gözlenen değişkende aynı puanı elde ederler (Milfont and Fischer, 2010).

Güçlü faktöriyel değişmezlik zayıf faktöriyel değişmezlik modelini tanımlayanların üzerinde bir dizi kısıtlama içerir (Widaman and Reise, 1997). Yalnızca faktör yüklerinin eşitliğinin test

edildiği zayıf faktöriyel değişmezlik modeline kıyasla bu modelde regresyon sabitinin de gruplar arası eşitliği gerekmektedir. Zayıf faktöriyel değişmezlik modeli, güçlü faktöriyel değişmezlik modeli için bir ön koşuldur (Tucker, Özer, Lyubomirsk ve Boehm, 2006). Güçlü faktöriyel değişmezliğin sağlanmaması durumunda, örtük ortalamaların karşılaştırması belirsizdir (Cheung and Rensvold, 2002). Buradan hareketle yapılan grup karşılaştırmalarının da gerçekliğinden söz edilemez.

2.4.4 Katı Faktöriyel Değişmezlik (Strict Factorial Invariance)

Bu aşamada test edilen hipotez, hata varyanslarının gruplar arasında değişmez olduğudur. Katı faktöriyel değişmezlik modelinde, faktör yükleri ve regresyon sabitinin yanı sıra hata varyanslarının da eşitliği gerekmektedir. Ortalama ve varyans kestirimlerinin gruplar arası karşılaştırmaları için, katı faktöriyel değişmezlik kanıtı elde edilmelidir (Gregorich, 2006). Katı faktöriyel değişmezlik kanıtının elde edilmesi madde varyanslarının sabit olduğu anlamına gelir. Madde varyansları gözlenen değişkenler arasındaki korelasyonları etkilediğinden bu korelasyonları karşılaştırılabilmesi için katı faktöriyel değişmezlik sağlanmalıdır (Tucker, Özer, Lyubomirsk ve Boehm, 2006). Ölçme değişmezliğine ilişkin aşamalı süreçte son aşamada yer alan bu model, en çok sınırlandırılmış modeldir.

2.5 İlgili Araştırmalar

Bu bölümde kronolojik sıralama ile daha önceden yapılmış ölçme değişmezliği çalışmalarına yer verilmiştir.

Uzun ve Öğretmen (2010), TIMSS 1999 verileri üzerinde yürüttükleri çalışmalarında, öğrencilerin fen bilimleri başarısını belirleyen değişkenleri belirlemiş ve bu değişkenlerin cinsiyet bazında değişmezliğini incelemişlerdir. Öz yeterlik, önem, tutum ve sınıf içi öğrenci etkinlikleri için aşamalı şekilde ölçme değişmezliği testleri yapılmış ve elde edilen uyum

indeksleri sonucunda modelde yer alan deęişkenlerin yalnızca metrik deęişmezlik modeline kadar uyum gösterdikleri görülmüştür. Cinsiyet bazında oluşturulan grupların ise katı deęişmezlik koşullarını sağlayamadıkları belirlenmiştir.

Güzeller (2011), PISA 2009 uygulamasında yer alan öğrenci anketinin sekiz tanesi OECD üyesi olmak üzere toplam on ülke arasındaki ölçme deęişmezliğini incelemiştir. Ankette yer alan bilgisayar tutum boyutunu ele alan araştırmacı, kültürler arası deęişmezliği incelemek üzere tüm öğrencilerin verilerini analizde kullanmıştır. Elde ettiği sonuçlar ölçme deęişmezliğinin sağlandığını, bir başka deyişle bilgisayar tutumunun kültürler arasında eşdeğer olduğunu, bu doğrultuda bilgisayar tutumuna dair puanların söz konusu ülkeler için sıralanabileceğini göstermiştir.

Hansson ve Gustafsson (2013), TIMSS 2003 verilerini kullanarak sosyoekonomik durum deęişkeninin İsveçli ve İsveçli olmayan öğrencilerin oluşturduğu alt gruplarda ölçme deęişmezliğini araştırmışlardır. Okunan kitap sayısı, anne babanın eğitim seviyesi ve okuldan beklenti içerikli ölçek maddelerinden seçilen 4 adet madde gözlenen deęişken olarak, standart matematik başarı puanı ise sonuç deęişkeni olarak belirlendikten sonra yapılan ölçme deęişmezliği analizleri sonucunda, gruplar arası metrik deęişmezliğin sağlandığı görülmüştür. Bu bulgu, sosyoekonomik durum deęişkeninin tüm gruplarda aynı anlama geldiği şeklinde yorumlanabilir.

Segeritz ve Pant (2013), PISA 2013 Almanya verileri ile yürüttükleri çalışmada ülkede ağırlıklı bulunan göçmen gruplarının (Türk ve eski SSCB kökenli) öğrenci öğrenme yaklaşımları aracı üzerinden ölçme deęişmezliğini araştırmışlardır. OECD öncülüğünde bir grup tarafından oluşturup geliştirilen bir araç olan öğrenci öğrenme yaklaşımları aracı 9 ölçek ve 45 maddeden oluşmaktadır. Aracın faktör yapısı incelenerek uygun olmayan maddeler elendikten sonra yapısal deęişmezlik ve metrik deęişmezliğin sağlandığı görülmüş, ancak

güçlü faktöriyel değişmezlik şartlarının sağlanamadığı görülmüştür. Ayrıca tüm ölçekler için metrik değişmezliğin sağlandığını ve bu nedenle birimlerin gruplar arasında karşılaştırılabilir olduğunu belirtmişlerdir. Araştırmadaki en ilgi çekici sonuç ise göçmen öğrencilerin akademik başarı açısından yerli öğrencilerin gerisinde olmalarına rağmen, soruları yanıtlamada daha olumlu bir eğilim göstermeleri olmuştur.

Uyar ve Doğan (2014), araştırmalarında PISA 2009 Türkiye örnekleminde öğrenci anketinde yer alan öğrenme stratejilerine ilişkin bir modeli cinsiyet, okul türü ve istatistiksel 12 bölge gruplarına göre incelemiş ve ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemeye çalışmışlardır. Yaptıkları analizler sonucunda modelin cinsiyet ve okul türleri gruplarında ölçme değişmezliğinin sağlanmadığını, istatistiksel bölge gruplarında ise tüm değişmezlik koşullarını yerine getirerek değişmezliğin sağlandığını belirlemişlerdir.

Asil ve Brown (2015), PISA 2009 okuma becerileri testinin dil, kültür ve ekonomik gelişim düzeylerinde ölçme değişmezliğini test etmişlerdir. Araştırmaya, örneklem büyüklükleri göz önünde bulundurulduğundan 55 ülke dahil edilmiştir. Avustralya'nın referans olarak seçildiği çalışmada İngilizce konuşulan yalnızca üç ülkenin Avustralya ile güçlü faktöriyel değişmezliği olduğu sonucuna ulaşılmıştır. Ayrıca elde edilen verilerden hareketle, eğitimde sosyo-ekonomik kaynakların ölçme değişmezliğinde önemli bir rol oynadığı ancak dil faktörünün ve dil eğitimi uygulamalarının etkilerinin daha az olduğu görülmüştür.

Başusta ve Gelbal (2015), PISA 2009 öğrenci anketinde yer alan fen bilgisi ve teknolojileri ile ilgili maddelere ait faktörleri belirleyerek, bu faktörlerin cinsiyet grupları arasında ölçme değişmezliğine sahip olup olmadığını inceledikleri çalışmalarında, Türkiye örneklemini ele almışlardır. Anket verilerinin analizi için yapısal eşitlik modellemesi kullanılmış ve ölçme modeli aşama aşama test edilmiştir. Analiz sonuçlarından elde edilen uyum indeksleri karşılaştırılmış ve ölçme değişmezliğinin cinsiyet değişkeni bakımından sağlandığı sonucuna

ulaşlmıştır. Buradan hareketle ölçme modelinin kız ve erkek öğrenci grupları için aynı biçimde ölçme yaptığı söylenebilir.

Kıbrıslıoğlu (2015), PISA 2012 uygulamasında matematik öğrenme alt boyutu anketleri ile oluşturulan matematik öğrenme modelinin ülke ve cinsiyet grupları arasında ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemeye çalıştığı araştırmasında Türkiye, Endonezya ve Çin-Şangay örneklemelerini ele almıştır. Açıklayıcı faktör analizi ile oluşturduğu matematik öğrenme algısı modelini doğruladıktan sonra, ölçme değişmezliği analizlerinde çoklu grup doğrulayıcı faktör analizini kullanarak modeli incelemiştir. Elde edilen araştırma sonuçlarına göre model ülkeler arasında ölçme değişmezliğinin tüm koşullarını sağlayamamış olup, cinsiyet grupları arasında ise model ölçme değişmezliğini her aşamada sağlamıştır. Buradan hareketle modele ait ortalama, varyans ve kovaryansların yalnızca cinsiyet grupları arasında karşılaştırılabilir olduğu sonucuna ulaşmışlardır.

Ayvallı (2016), PISA 2012 uygulamasındaki matematik okuryazarlığı testinin Türkiye'deki cinsiyet ve bölge grupları arasında ve OECD ülkeleri arasında ölçme değişmezliğinin sağlanıp sağlanmadığını çoklu grup doğrulayıcı faktör analizi aracılığıyla test etmiştir. Tüm gruplarda ayrı ayrı doğrulayıcı faktör analizi sağlaması yapıldıktan sonra kovaryans matris eşitliği de sağlanmış, ardından ölçme değişmezliği testlerine geçilmiştir. Yapılan analizler sonucunda Türkiye'deki cinsiyet ve bölge grupları arasında ölçme değişmezliğinin sağlandığı, OECD üyesi ülkeler arasında ise ölçme değişmezliğinin sağlanmadığı görülmüştür.

Karakoç Alatl (2016), araştırmasında PISA 2012 matematik ve fen okuryazarlık testleri ile okuma becerileri testlerinden elde edilen puanların uygulamaya katılan Avustralya, Şangay-Çin, Türkiye ve Fransa grupları arasında ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemeyi amaçlamıştır. Dil değişkenine göre incelenen çalışmada Çoklu Grup Doğrulayıcı Faktör Analizi, Genelleştirilmiş Aşamalı Doğrusal Modelleme, Simultaneous Item Bias Test

ve Madde Tepki Kuramı'na dayalı Olabilirlik Oran gibi farklı teknikler kullanılmıştır. Araştırma sonuçlarına göre, PISA 2012 matematik ve fen okuryazarlığı testleri ile okuma becerileri testlerinin belirlenen ülke örneklemi arasında dil değişkenine göre ölçme değişmezliğinin sağlanmadığını belirlemiştir. Ayrıca, temel model olan yapısal değişmezliğinin sağlandığı ancak metrik değişmezlik modelinin sağlanmadığı ortaya koyulmuştur.

Kıbrıslıoğlu Uysal ve Akın Arıkan (2018), PISA 2006 ve PISA 2015 uygulamalarında ortak biçimde uygulanmış olan fen öz yeterlilik ölçeğinin Türkiye örneğinde yıllara ve cinsiyet gruplarına göre ölçme değişmezliğini sağlayıp sağlamadığını araştırmışlardır. Yıllara ve cinsiyete bağlı ölçme değişmezliğinin varlığına karar verme aşamasında çoklu-grup doğrulayıcı faktör analizi kullanılmış ve uyum indeksleri karşılaştırılarak değişmezlik aşamaları incelenmiştir. Elde edilen bulgulara göre, fen öz yeterlilik ölçeğinin 2006 ve 2015 yıllarında cinsiyet grupları arasında ölçme değişmezliğini her aşamada doğruladığı belirlenmiştir. Her iki cinsiyet grubunun yıllar arası karşılaştırmasında da toplam grubun yıllar arası karşılaştırmasında da yapısal değişmezlik ve zayıf faktöriyel değişmezlik kanıtları elde edilmiştir.

Şekercioğlu ve Koğar (2018), PISA 2015 uygulamasında matematik ve fen okuryazarlığı ile okuduğunu anlama becerisi testlerinin yaygın kullanılan yedi adet test dili açısından ölçme değişmezliğini inceleme ve araştırmaya dahil edilen bilişsel test maddelerine ilişkin değişen madde fonksiyonlarını incelemeyi amaçlamışlardır. Toplam verinin %59.14'ünü temsil eden İngilizce, İspanyolca, Arapça, Portekizce, Çince, Almanca ve Fransızca dillerinin ele alındığı araştırmada, yapılan analizler sonucunda bilişsel testlerin üçü için de en iyi çalışan modelin güçlü faktöriyel değişmezlik olduğu belirlenmiş ve ölçme değişmezliğinin sağlanmadığı bulgusu elde edilmiştir.

BÖLÜM III

YÖNTEM

Bu bölümde araştırma deseni, araştırma verileri, verilerin düzenlenmesi ve verilerin analizi ile ilgili bilgilere yer verilmiştir.

3.1 Araştırma Deseni

Bu araştırmanın amacı TIMSS 2015 uygulamasında bilişsel test olarak yer alan dördüncü sınıf matematik başarı testinin maddelerin OECD ülkeleri arasında eşdeğer olup olmadığının belirlenmesidir. Bu bağlamda çalışma, TIMSS 2015 uygulamasında kullanılan başarı testi maddeleri arasındaki ilişkileri farklı gruplar arasında incelediği için ilişkisel tarama modeli kullanılarak yürütülmüştür. Bilimsel araştırmalarda ilişkisel tarama modeli, iki ya da daha fazla değişken arasında var olduğu kestirilen ilişkinin düzeyini ve yönünü belirleme amaçlı kullanılmaktadır (Karasar, 2005).

3.2 Araştırma Verileri

TIMSS 2015 4. Sınıf düzeyinde 45 ülkeden toplam 253.546 öğrenciye uygulanmıştır. Uygulamaya katılan ülkeler içerisinde OECD üyesi olan 24 ülke, 24 OECD ülkesinden toplam 132.226 öğrenci ve bu öğrencilerin içerisinde de 7. kitapçığın ortak olarak uygulandığı 9.461 öğrenci araştırmanın verisini oluşturmaktadır. Araştırmada kullanılan veriler TIMSS uygulamasının resmi sitesinden elde edilmiştir. (www.timssandpirls.bc.edu).

Araştırmanın dahil edilen ülkeler kapsamında öğrencilerin ülkelere göre dağılımı Tablo 3.1'deki gibidir.

Tablo 3.1. Ülkelere Göre Örneklem Dağılımı

Ülke	N	%
Almanya	285	3.01
Amerika	720	7.61
Avustralya	431	4.56
Belçika	384	4.06
Çek Cumhuriyeti	376	3.97
Finlandiya	357	3.77
Hollanda	321	3.39
İngiltere	285	3.01
İrlanda	310	3.28
İspanya	538	5.69
İsveç	284	3.00
İtalya	307	3.24
Japonya	314	3.32
Kanada	893	9.44
Kore	331	3.50
Macaristan	354	3.74
Norveç	286	3.02
Polonya	346	3.66
Portekiz	333	3.52
Slovakya	414	4.38
Slovenya	329	3.48
Şili	343	3.63
Türkiye	461	4.87

Tablo 3.1'in devamı: Ülkelere Göre Örneklem Dağılımı

Ülke	N	%
Yeni Zelanda	459	4.85

Tablo 3.1 incelendiğinde katılımcı 24 OECD üyesi ülkeden araştırmaya dahil edilen öğrenci dağılımlarına göre en fazla katılımcıya sahip ülke %9.44 (893 öğrenci) ile Kanada, en az katılımcıya sahip ülke ise %3 (284 öğrenci) ile İsveç'tir. Diğer ülkelerin katılımcı oranları ise %5.69 ve %3.01 (538 ve 285 öğrenci) arasında değişmektedir.

TIMSS 2015 uygulamasının değerlendirme çerçevesinde; matematik ve fen bilimleri alanlarında başarı testleri, öğrenci başarısını etkileyen eğitimsel ve sosyal ortamlar hakkında bilgi toplayan anketlerden oluşmaktadır. Başarı testleri ile öğrencilerin matematik ve fen bilimleri alanlarındaki bilgi becerilerinin ölçülmesi amaçlanmaktadır.

TIMSS uygulamasında çoktan seçmeli ve açık uçlu sorular kullanılmaktadır. Çoktan seçmeli sorular dört seçenekli olmak üzere, her birinin tek bir doğru seçeneği bulunmaktadır. Verilen yanlış cevap sayısının doğru cevaplar üzerinde bir etkisi bulunmamaktadır. Açık uçlu sorularda ise öğrenci; açıklama yaparak, birtakım verilere dayalı çıkarımlarda bulunarak veya şekiller çizerek kendi cevaplarını oluşturmaktadırlar. Bu soru tipinde puanlama her bir soru için özel olarak belirlenmiş puanlama anahtarına göre yapılmaktadır. TIMSS 2015'te 4.sınıf düzeyinde öğrenciler, her bölüm için 36 dakika süre tanınan başarı testlerine katıldıktan sonra 30 dakika süre verilen anketleri tamamlamaktadır (TIMSS, 2015).

TIMSS uygulamasında 4. ve 8. sınıf düzeylerinde matematik ve fen bilimleri sorularının öğrenme alanı ve bilişsel alan olmak üzere iki temel boyutu bulunmaktadır. Tablo 3.2'de TIMSS 2015 4. sınıf düzeyinde yer alan soruların bilişsel alanlara göre dağılımı verilmiştir.

Tablo 3.2. TIMSS 2015 4.Sınıf Bilişsel Alanlara Göre Dağılımı

Bilişsel Alan	Matematik	Fen Bilimleri
Bilgi	%40	%40
Uygulama	%40	%40
Akıl Yürütme	%20	%20

Tablo 3.2’deki verilere bakıldığında TIMSS 2015 uygulamasında kullanılan matematik ve fen bilimleri sorularının aynı oranda bilişsel alanlara ayrıldığı ifade edilebilmektedir. Her iki alanda da soruların %40 bilgi ve uygulama, %20 akıl yürütme bilişsel alanlarına ayrıldığı görülmektedir.

Tablo 3.3’te TIMSS 2015 4. sınıf düzeyinde yer alan Matematik sorularının öğrenme alanlarına göre dağılımları verilmiştir.

Tablo 3.3. TIMSS 2015 4.Sınıf Düzeyi Matematik Sorularının Öğrenme Alanlarına Göre Dağılımı

Öğrenme Alanı	Alt Öğrenme Alanı	Yüzde
	Doğal Sayılar	
Sayılar	Kesirler ve Ondalık Sayılar	%50
	Doğal Sayılarla İşlemler	
	Örüntüler ve İlişkiler	
Geometrik Şekiller ve Ölçme	Noktalar, Doğrular ve Açılar	%35
	İki ve Üç Boyutlu Şekiller	
	Veri Okuma ve Yorumlama	
Veri Gösterme	Verileri Düzenleme ve Sunma	%15

Tablo 3.3'teki veriler incelendiğinde, öğrenme alanları içerisinde ağırlığın %50 ile en fazla Sayılar öğrenme alanına verildiği, Geometrik Şekiller ve Ölçme öğrenme alanına %35 ve Veri Gösterme öğrenme alanına %15 oranında ağırlık verildiği görülmektedir.

3.3 Verilerin Düzenlenmesi

Bu araştırmada, TIMSS 2015 uygulamasında kullanılan 14 adet test kitapçığından tüm ülkelere uygulanmış olan 7 numaralı kitapçık analiz için seçilmiştir. Matematik başarısına ait 10'u açık uçlu, 15'i çoktan seçmeli madde olmak üzere toplam 25 adet soru analize dahil edilmiş ve çoktan seçmeli maddelerde her doğru seçenek "1", yanlış seçeneklerden her biri ise "0" olarak yeniden kodlanmıştır. Açık uçlu maddelerde ise "doğru cevap" olarak belirlenmiş yanıtlar "1"; "yanlış cevap", "kısmi doğru cevap", "erişilememiş" veya "atlanmış veya geçersiz" olarak belirlenmiş yanıtlar ise "0" olarak yeniden kodlanmıştır. Araştırmada kullanılan maddeler Tablo 3.4'te verilmiştir.

Tablo 3.4. *Analizde Kullanılan Matematik Başarı Testi Maddeleri*

Madde No	Maddeler	Madde Formatı
M041298	Which Rectangle is 1/4 Shaded	Çoktan Seçmeli
M041007	Which Could be Pat's Number	Çoktan Seçmeli
M041280	Number Closest to 1362/ 32	Çoktan Seçmeli
M041059	Fraction of Circles That are Black	Açık Uçlu
M041046	Which is the Largest Fraction	Çoktan Seçmeli
M041048	Height of Wall Made of 4 Rows	Çoktan Seçmeli
M041169	Shape and Its Reflection	Çoktan Seçmeli
M041333	How Many Cubes are in the Box	Çoktan Seçmeli
M041262	What is The Area of the Triangle	Çoktan Seçmeli

Tablo 3.4'ün devamı: Analizde Kullanılan Matematik Başarı Testi Maddeleri

Madde No	Maddeler	Madde Formatı
M041267	Draw Line With Given Specification	Açık Uçlu
M041271	Selecting Correct Bar Graph	Çoktan Seçmeli
M041276A	Who Won the Triathlon	Açık Uçlu
M041276B	What Sue Needs to Improve on	Açık Uçlu
M061026	8 Thousands + 4 Hundreds + 5 Ones	Çoktan Seçmeli
M061273	$27 \times 43 =$	Çoktan Seçmeli
M061034	Number of People That Got Off the Train	Açık Uçlu
M061040	Shaded Fraction of a Square	Çoktan Seçmeli
M061228	Art Teacher Cuts Paper for Her Class	Açık Uçlu
M061166	Find the Value of Win a Subtraction Sentence	Açık Uçlu
M061171	Total Number of Trading Cards Mona and Ben Bought	Çoktan Seçmeli
M061080	Mark an X on Parallel Sides of Trapezoid	Açık Uçlu
M061222	Distance Between Shawn and Rick on Number Line	Çoktan Seçmeli
M061076	Sandra's Model for a Decorated Cube	Açık Uçlu
M061084	Complete Pie Chart of Friends' Favorite Flowers	Açık Uçlu

3.4. Verilerin Analizi

Verilerin analizine geçilmeden önce veri seti analize uygun hale getirilme amacıyla düzenlenmiş, maddelere dair yeniden kodlama işlemleri yapılmıştır. Öncelikle temel sayıtları test etmek amacıyla merkezi eğilim ölçüleri, daha sonra iç tutarlılık kapsamında KR-20 güvenilirlik katsayısı, daha sonra ise veri setinin normalliğinin incelenmesi amacıyla çarpıklık ve basıklık katsayıları hesaplanmıştır. Katılımcı ülkelerin matematik başarı test

puanları dağılımı üzerinden hesaplanan merkezi eğilim ölçüleri (ortalama, mod, medyan), standart sapma, ranj, basıklık ve çarpıklık katsayıları ve KR-20 iç tutarlılık katsayısı bulguları Tablo 3.5'te verilmiştir.

Tablo 3.5. *Ülkelerin Matematik Başarı Testi Puanları için Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları*

Ülke	N	\bar{X}	Mod	Medyan	S_s^*	Ranj	K_y^*	B_s^*	KR-20
Almanya	285	11.86	13	12	4.56	23	.155	-.620	.79
Amerika	720	14.24	17	14	5.24	24	-.085	-.701	.84
Avustralya	431	12.31	15	12	5.24	24	.019	-.733	.84
Belçika	384	14.50	16	15	3.93	21	.013	-.194	.70
Çek Cumhuriyeti	376	12.54	11	13	4.86	24	-.024	-.512	.81
Finlandiya	357	13.81	17	14	4.45	24	-.358	-.244	.78
Hollanda	321	13.56	14	14	3.96	20	-.189	-.324	.70
İngiltere	285	14.04	13	14	5.22	23	-.161	-.680	.84
İspanya	538	12.17	15	12	4.44	23	-.060	-.631	.77
İsveç	284	12.36	13	12	5.06	23	.110	-.800	.83
İtalya	307	10.93	10	11	4.32	23	.136	-.243	.75
Japonya	314	18.24	20	19	4.63	22	-.691	-.004	.82
Kanada	893	12.01	11	12	4.78	24	.181	-.513	.79
Kore	331	19.13	20	20	3.98	19	-.924	.610	.80
Macaristan	354	14.01	19	14	5.85	24	-.031	-1.055	.87
Norveç	286	11.15	8	11	4.47	23	.162	-.196	.77
Polonya	346	13.81	15	14	4.93	25	-.091	-.548	.82
Portekiz	333	13.57	11	14	4.86	21	-.112	-.706	.81

Tablo 3.5'in devamı: Ülkelerin Matematik Başarı Testi Puanları için Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları

Ülke	N	\bar{X}	Mod	Medyan	S_s^*	Ranj	K_y^*	B_s^*	KR-20
Slovakya	414	10.66	8	10	4.95	24	.180	-.538	.82
Slovenya	329	12.48	9	12	4.87	24	.069	-.635	.81
Şili	343	10.15	13	10	4.76	24	.442	-.113	.82
Türkiye	461	11.38	12	11	5.36	24	.159	-.694	.85
Yeni Zelanda	459	12.18	12	12	5.29	24	-.017	-.787	.84

* S_s : Standart sapma, * K_y : Çarpıklık katsayısı, * B_s : Basıklık katsayısı

Tablo 3.5'te yer alan verilere göre, ülkelere ait matematik başarı testi puanlarının merkezi eğilim ölçülerinin (aritmetik ortalama, mod, medyan) birbirine oldukça yakın oldukları görülmüştür. Normallik değerleri incelendiğinde çarpıklık katsayılarının tüm ülkelerde -1 ile +1 arasında yer aldığı, basıklık katsayılarının Macaristan dışındaki tüm ülkelerde -1 ile +1 arasında yer aldığı görülmüştür. Çarpıklık ve basıklık katsayılarının -1 ile +1 arasında yer alması testten alınan puanların normal dağılıma sahip olduğunu göstermektedir (Mertler ve Vannatta, 2005). Buradan hareketle, ülkelere ait verilerin çoğunluğunun normale yakın dağılım gösterdiği söylenebilir. KR-20 güvenilirlik katsayıları incelendiğinde, ülkelere ait güvenilirlik katsayıları değerlerinin .70 ile .87 (Belçika, Hollanda ile Macaristan) aralığında olduğu görülmüştür. Veri grubuna ait güvenilirlik katsayılarının kabul edilebilir düzeylerde olması için .70 ile .80 aralığında yer alması beklenmektedir (Nunnally ve Bernstein, 1994).

Her bir ülkeye ait tek faktörlü modelin doğrulanıp doğrulanmadığını belirlemek amacıyla doğrulayıcı faktör analizi yapılmış ve uyum indeksleri incelenmiştir. Daha sonra kovaryans matrislerinin eşitliği testi ve çoklu grup doğrulayıcı faktör analizi yapılmıştır. Analizlere ait sonuçlara bulgular bölümünde yer verilmiştir.

Gerçekleştirilen analizlerde en yüksek olabilirlik tahmini yöntemi (maximum likelihood estimation) kullanılmıştır. Bu yöntem, parametrelerin kestiriminde en yüksek olasılığı veren modelin bulunmasına dayanır ve bu tür veriler için en sık uygulanan istatistiksel çıkarım yöntemidir (White, 1982).

Çoklu grup doğrulayıcı faktör analizine geçilmeden önce veri setinin karşılaması gereken varsayımlardan biri olan çoklu bağlantı varsayımı incelenmiştir. Çoklu bağlantı, bağımsız değişkenlerden birinin diğerinin yerine kullanılabilecek kadar üst seviyede ilişkili olmaları durumudur ve maddeler arası korelasyon katsayılarının .90'ın üzerine çıkması durumu çoklu bağlantı sorununa işaret etmektedir (Çokluk ve diğerleri, 2010). Bu bağlamda maddeler arası korelasyon katsayıları her bir ülke için ayrı ayrı hesaplanmıştır. Matematik maddelerinin korelasyon katsayılarının .004 ve .435 aralığında olduğu belirlenmiştir. Buna göre, bağımsız değişkenler arasında çoklu bağlantı problemi olmadığı yorumu yapılabilir.

Araştırmada, ölçme değişmezliği değerlendirmesinde model karşılaştırmaları esnasında X^2 , X^2/sd , CFI ve SRMR değerleri esas alınmıştır. Büyük örneklemeler üzerinden yürütülen analizlerde veri setinin normalden belli bir ölçüde sapma gösterebileceğinin göz önünde bulundurulması, veri setinin kategorik puanlamaya dayalı olması ve modellere ait X^2 değerleri arasında manidar bir fark olup olmadığını belirlemede kullanılacak T_s değeri hesaplamada Satorra-Bentler X^2 ($S-B_{X^2}$) değerine gereksinim duyulması sebepleriyle araştırmada asimptotik kovaryans matrisi ayrıca hesaplanmıştır. Asimptotik kovaryans matrisinin de analizde kullanılmış olması sebebiyle, $S-B_{X^2}$ değeri analizde model uyumunu değerlendirmede ve modelleri karşılaştırmada kullanılmıştır.

Modellerin hiyerarşik yapısı bağlamında karşılaştırmalar yuvalanmış (nested) modeller arasında yapılmıştır. Buna göre yapısal değişmezlik modeli ile zayıf faktöriyel değişmezlik modeli (Model 1 ve Model 2), zayıf faktöriyel değişmezlik modeli ile güçlü faktöriyel

değişmezlik modeli (Model 2 ve Model 3), güçlü faktöriyel değişmezlik modeli ile katı faktöriyel değişmezlik modeli (Model 3 ve Model 4) arasında karşılaştırmalar yapılmıştır. Karşılaştırmalar yapılırken ilk olarak, iki model arasındaki X^2 değerleri farkına (ΔX^2) ve serbestlik dereceleri farkına (Δsd) bakılmıştır. ΔX^2 'nin istatistiksel olarak manidar olup olmadığına, Ki-kare dağılım tablosundaki $p < .05$ anlamlılık düzeyinde serbestlik dereceleri farkına karşılık gelen kritik değere bakılarak karar verilmiştir. Ancak bu araştırmada $S-B_{x^2}$ değeri kullanıldığı için farkın manidarlığını değerlendirmeden önce T_s değerleri hesaplanmıştır. Hesaplanan T_s değerinin kritik ki-kare değerinden büyük olması durumunda karşılaştırılan modeller arasındaki farkın manidar olduğu ve ölçme değişmezliğinin sağlanmadığı, T_s değerinin kritik ki-kare değerinden küçük olması durumunda ise modeller arası farkın manidar olmadığı yorumu yapılabilmektedir. Bunun yanında araştırmada, ΔCFI ve $\Delta SRMR$ değerleri de karşılaştırma yaparken ele alınmıştır.

Model karşılaştırmaları yapılırken örneklem büyüklüğüne bağlı bazı kesme değerleri referans alınmıştır. Söz konusu kesme değerleri, örneklem değerinin küçük olduğu ($n < 300$) durumda gruplar arası zayıf faktöriyel değişmezlik testi için $\Delta CFI \leq -.005$ ve $\Delta SRMR \geq .025$, güçlü ve katı faktöriyel değişmezliğin karşılaştırılmasında $\Delta CFI \geq -.005$ ve $\Delta SRMR \geq .005$ olarak ele alınmıştır. Örneklem büyüklüğünün yeterli olduğu ($n > 300$) durumda gruplar arası zayıf faktöriyel değişmezlik testi için $\Delta CFI \geq -.010$ ve $\Delta SRMR \geq .030$, güçlü ve katı faktöriyel değişmezliğe ait karşılaştırmalarda ise $\Delta CFI \geq -.010$, $\Delta SRMR \geq .010$ olarak ele alınmıştır (Chen, 2007).

BÖLÜM IV

BULGULAR

Bu bölümde araştırma kapsamında elde edilen, doğrulayıcı faktör analizi bulgularına, kovaryans matrislerinin eşitliği testi bulgularına ve çoklu-grup doğrulayıcı faktör analizi bulgularına yer verilmiştir.

4.1. Doğrulayıcı Faktör Analizi Bulguları

Doğrulayıcı faktör analizi sonuçları doğrultusunda elde edilen bulgular Tablo 4.1’de verilmiştir.

Tablo 4.1. Ülkelerin Matematik Başarı Testi Puanları Doğrulayıcı Faktör Analizi Bulguları

Ülke	S-B _x ² (sd)	x ² /sd	CFI	NNFI	SRMR	RMSEA
Almanya	266.95(275)	0.97	1.00	1.00	.046	.000
Amerika	442.62(275)	1.61	.98	.98	.036	.029
Avustralya	321.84(275)	1.17	.99	.99	.039	.020
Belçika	281.24(275)	1.02	.99	.99	.043	.008
Çek Cumhuriyeti	369.67(275)	1.34	.96	.96	.046	.030
Finlandiya	284.54(275)	1.03	1.00	.99	.042	.010
Hollanda	384.82(275)	1.4	.89	.88	.055	.035
İngiltere	379.10(275)	1.38	.96	.96	.053	.037
İrlanda	283.34(275)	1.03	1.00	1.00	.045	.010
İspanya	345.81(275)	1.26	.97	.97	.039	.022
İsveç	334.63(275)	1.22	.97	.97	.049	.028
İtalya	343.92(275)	1.25	.95	.94	.052	.029
Japonya	304.76(275)	1.11	.99	.99	.050	.019

Tablo 4.1'in devamı: Ülkelerin Matematik Başarı Testi Puanları Doğrulayıcı Faktör Analizi Bulguları

Ülke	S-B χ^2 (sd)	χ^2 /sd	CFI	NNFI	SRMR	RMSEA
Kanada	478.40(275)	1.74	.96	.96	.035	.029
Kore	314.86(275)	1.14	.98	.98	.050	.021
Macaristan	331.81(275)	1.21	.99	.99	.042	.024
Norveç	291.54(275)	1.06	.99	.99	.048	.015
Polonya	381.13(275)	1.39	.96	.96	.048	.033
Portekiz	390.05(275)	1.42	.95	.94	.050	.035
Slovakya	442.60(275)	1.61	.95	.95	.048	.038
Slovenya	328.14(275)	1.19	.98	.98	.047	.024
Şili	312.96(275)	1.14	.99	.98	.044	.020
Türkiye	330.73(275)	1.2	.99	.99	.038	.021
Yeni Zelanda	448.28(275)	1.63	.97	.96	.045	.037

Tablo 4.1'deki bulgulara göre, tüm katılımcı ülkelerin matematik başarı testi puanlarına ait χ^2 /sd oranının 3'ün altında olduğu gözlenmiştir. CFI değerlerine bakıldığında, Hollanda dışındaki tüm ülkelerin değerlerinin .90'ın üzerinde olduğu; yalnızca Hollanda'ya ait CFI değerinin .89 olduğu görülmüştür. Benzer biçimde, Hollanda dışında analize dahil edilen ülkelerin tamamına ait NNFI değerinin .90 ve üstünde olduğu, yalnızca Hollanda'ya ait NNFI değerinin .88 olduğu görülmüştür. SRMR değerleri incelendiğinde, 21 ülkeye ait SRMR değerinin .05'in altında olduğu; Hollanda, İngiltere ve İtalya'ya ait SRMR değerlerinin (.55, .53 ve .52) .50'nin üzerine bir miktar çıktığı görülmektedir. RMSEA değerlerinin ise tüm ülkelerde .05'in altında yer aldığı görüldüğünden; genel olarak, 24 ülkeye ait her bir modelin doğrulayıcı faktör analizi sonuçlarına göre doğrulandığı ifade edilebilir.

4.2. Kovaryans Matrislerinin Eşitliği Testi Bulguları

Analiz kapsamında, çoklu grup doğrulayıcı faktör analizine geçilmeden önce OECD üyesi ülkelerinin matematik başarı testi puanlarına ilişkin kovaryans matrislerinin eşitliği test edilmiş ve sonuçları Tablo 4.2’de verilmiştir.

Tablo 4.2. OECD Üyesi Ülkelere Ait Matematik Başarı Testi Puanlarının Kovaryans Matrislerinin Eşitliği Testi Analiz Bulguları

Grup	SB-X ² (sd)	p	X ² /sd	RMSEA	GFI	CFI	SRMR
OECD üyesi ülkeler	13170.60 (7475)	.00	1.76	.044	.90	.92	.081

Tablo 4.2’deki analiz sonuçları incelendiğinde; SB-X²/sd oranının 3’ün altında olduğu, RMSEA değerinin .05’in altında olduğu, GFI değerinin .90’a eşit ve CFI değerinin .90’ın üzerinde olduğu, SRMR değerinin ise .05’in üzerine çıktığı görülmektedir. Elde edilen bulgular genel olarak değerlendirildiğinde, ülkeler arasındaki kovaryans matrisleri arasında uyum olduğu söylenebilir.

4.3. Çoklu-Grup Doğrulayıcı Faktör Analizi Bulguları

TIMSS 2015 4.sınıf matematik test puanlarının OECD üyesi ülkelere göre ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemek amacıyla yapılan çoklu grup doğrulayıcı faktör analizi sonuçları Tablo 4.3’te verilmiştir.

Tablo 4.3. OECD Üyesi Ülkelerin TIMSS 2015 4.Sınıf Matematik Başarı Testi Puanlarına ait Çoklu-Grup Doğrulayıcı Faktör Analizi Bulguları

	SB- χ^2 (sd)*	ΔX^2 (Δ sd)	X^2 /sd	ΔX^2 (Δ sd)	CFI	Δ CFI	SRMR	Δ SRMR
MODEL1 ^A	15034.70 (7750)	-	1.9399	-	.90	-	.081	-
MODEL2 ^B	12495.26 (7175)	2539.44 (575)	1.7415	0.1984	.93	-.03	.051	.030
MODEL3 ^C	8381.66 (6600)	4113.6 (575)	1.2699	0.4716	.98	-.05	.045	.006
MODEL4 ^D	10845.62 (7175)	-2463.96 (-575)	1.5116	-0.2417	.95	.03	.079	-.034

*p<.05

AYapısal Değişmezlik (Faktör yükleri, faktör korelasyonları ve hata varyansları sabit)

BZayıf Faktöriyel Değişmezlik (Faktör yükleri serbest, faktör korelasyonları ve hata varyansları sabit)

CGüçlü Faktöriyel Değişmezlik (Faktör yükleri ve hata varyansları serbest, faktör korelasyonları sabit)

DKatı Faktöriyel Değişmezlik (Hata varyansları serbest, faktör yükleri ve faktör korelasyonları sabit)

Öncelikle karşılaştırılacak ülkelerin faktör yapılarının benzer olup olmadığının incelendiği yapısal değişmezlik modeli test edilmiştir. Yapılan analize göre; S-B χ^2 ve serbestlik derecesi oranının (X^2 /sd) 2'nin altında olduğu, CFI değerinin .90 olduğu ve SRMR değerinin .08'in bir miktar üzerine çıktığı belirlenmiştir. Model 1'e dair uyum indekslerinin genel olarak kabul edilebilir düzeyde olduğu ve modelin doğrulandığı söylenebilir. Bu durumda yapısal değişmezlik modelinin sağlandığı yorumu yapılabilir.

Yapısal değişmezlik (Model 1) ve zayıf faktöriyel değişmezlik (Model 2) modelleri karşılaştırıldığında, S-B ΔX^2 ve Δ sd oranının iyileştiği, CFI değerinin manidar biçimde farklılaştığı (Δ CFI <-.01) ve benzer şekilde SRMR değerindeki değişimin manidar olduğu (Δ SRMR > .025) belirlenmiştir. S-B ΔX^2 ve Δ sd oranındaki iyileşmenin manidar olup olmadığının belirlenmesi için T_s değeri hesaplanmış ve 2972.648 olarak bulunmuştur. T_s değerinin X^2 dağılımı tablosundaki kritik değer üzerinde olduğu belirlenmiştir, $X^2_{\text{fark}}(575) = 631.893$, p<.05. Bulgular genel olarak ele alındığında, tüm uyum indekslerinde farkın manidar olduğu görüldüğünden yapısal değişmezlik ve zayıf faktöriyel değişmezlik modelleri arasında anlamlı bir fark olduğu söylenebilir.

Zayıf faktöriyel değişmezlik (Model 2) ve güçlü faktöriyel değişmezlik (Model 3) modelleri karşılaştırıldığında, $S-B\Delta_X^2$ ve Δ_{sd} oranına ait değerin iyileştiği görülmüştür. İyileşmenin manidar olup olmadığının belirlenmesi amacıyla T_s değeri hesaplanmış ve 7194.932 olarak bulunmuştur. T_s değerinin X^2 dağılımı tablosundaki kritik değerin üzerinde olduğu belirlenmiştir, $X^2_{fark}(575) = 631.893$, $p < .05$. Bu nedenle, zayıf faktöriyel değişmezlik modeli ile güçlü faktöriyel değişmezlik modelleri arasında manidar bir farkın olduğu söylenebilir. Ayrıca Model 2 ve Model 3 bağlamında CFI değerinin ($\Delta CFI < -.01$) manidar biçimde farklılaştığı ancak SRMR değerinde ($\Delta SRMR < .01$) manidar bir farklılaşma olmadığı görülmüştür.

Son olarak, güçlü faktöriyel değişmezlik (Model 3) ve katı faktöriyel değişmezlik (Model 4) modelleri karşılaştırıldığında $S-B\Delta_X^2$ ve Δ_{sd} oranının kötüleştiği görülmüş ve T_s değeri hesaplanmıştır. 2739.085 olarak bulunan T_s değerinin X^2 dağılımı tablosundaki kritik değerin üzerinde olduğu tespit edilmiştir, $X^2_{fark}(575) = 631.893$, $p < .05$. Ancak, CFI ($\Delta CFI > -.01$) ve SRMR ($\Delta SRMR > .01$) değerlerindeki değişimlerin manidar olmadığı söylenebilir. Bulgular genel olarak ele alındığında, üç uyum indeksinden ikisinde manidar bir fark görülmediğinden güçlü faktöriyel değişmezlik modeli ile katı faktöriyel değişmezlik modeli arasındaki farkın manidar olmadığını söylemek mümkündür.

Yukarıdaki bulgulara dayanarak, dört model içerisinde en iyi çalışan modelin güçlü faktöriyel değişmezlik modeli olduğu görülmektedir. Buradan hareketle analiz, matematik başarı testinin yapısının OECD ülkeleri için eşit olmadığını ve ölçme değişmezliğinin sağlanmadığını ortaya koymaktadır.

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde araştırmanın bulgularına ilişkin sonuçlara ve bu sonuçlara dayalı olarak önerilere yer verilmiştir.

5.1. Sonuç ve Tartışma

Araştırmada, TIMSS 2015 matematik başarı testinden elde edilen puanların OECD üyesi ülkeler için ölçme değişmezliğini sağlayıp sağlamadığı incelenmiştir. Öncelikle, veri setinin sağlanması gereken temel sayıtlara bakılmış, merkezi eğilim ölçüleri (aritmetik ortalama, mod, medyan), standart sapma, ranj ve güvenilirlik katsayıları hesaplanmıştır. Elde edilen bulgularla birlikte, dağılımın normale yakın olduğu yorumu yapılmıştır. Güvenilirlik katsayılarının ise her ülke için .70'in üstünde çıktığı, buradan hareketle test maddelerine ait iç tutarlılığın kabul edilebilir düzeyde olduğu belirlenmiştir.

Araştırma yapılan ülke gruplarına ait ölçme modellerinin doğrulanıp doğrulanmadığını incelemek için doğrulayıcı faktör analizi yapılmıştır. Analiz sonuçlarına göre iyilik uyum indekslerinin yeterli düzeyde olduğu ve ülkelere ait modellerin doğrulandığı belirlenmiştir. Ölçme değişmezliği analizlerine geçilmeden önce, ülke gruplarına ait kovaryans matrislerinin eşitliği test edilmiştir. Test sonucunda elde edilen bulgular ülkelere ait kovaryans matrisleri arasında iyi düzeyde uyum olduğunu göstermiştir.

TIMSS 2015 uygulaması dördüncü sınıf matematik başarı testinden elde edilen puanların ülkelere göre ölçme değişmezliği testi çoklu-grup doğrulayıcı faktör analizi kullanılarak yapılmıştır. Elde edilen bulgulara göre, Model 1'e dair uyum indekslerinin genel olarak kabul edilebilir düzeyde olduğu ve modelin doğrulandığı görülmüştür. Bu durumda yapısal değişmezlik modelinin sağlandığı tespit edilmiştir. Analizde model karşılaştırmaları

yuvalanmış (nested) modeller arasında yapılmıştır. Buna göre, yapısal değişmezlik modeli (Model 1) ile zayıf faktöriyel değişmezlik modeli (Model 2), zayıf faktöriyel değişmezlik modeli (Model 2) ile güçlü faktöriyel değişmezlik modeli (Model 3) karşılaştırılmış, uyumun her iki karşılaştırmada da Model 1'e göre manidar biçimde iyileştiği görülmüştür. Güçlü faktöriyel değişmezlik (Model 3) ve katı faktöriyel değişmezlik (Model 4) modelleri ile yapılan karşılaştırma sonucunda ise uyumun Model 3'e göre kötüleştiği görülmüştür. Buradan hareketle, en iyi çalışan modelin güçlü faktöriyel değişmezlik (Model 3) modeli olduğu ve OECD üyesi ülkeler için matematik başarı testine ait puanların ölçme değişmezliğini sağlamadığı sonucuna ulaşılmıştır.

Ülke gruplarına ilişkin elde edilen bulgulara göre çalışmanın, Ayvalli'nin (2016) PISA 2012 uygulaması matematik okuryazarlığı testinin OECD üyesi ülkelere göre ölçme değişmezliğini araştırdığı çalışmasıyla benzerlik gösterdiği görülmüştür. Her iki çalışmada da ülkelere göre ölçme değişmezliği güçlü faktöriyel değişmezlik düzeyinde sağlanmıştır. Kıbrıslıoğlu'nun (2015) PISA 2012 matematik öğrenme modelinin ülkelere göre ölçme değişmezliğini araştırdığı çalışmayla aynı doğrultuda bulgular elde edilmiştir. Her iki çalışmada da ölçme değişmezliğinin sağlanamadığı, Kıbrıslıoğlu'nun (2015) çalışmasında ölçme modelinin yalnızca yapısal değişmezlik düzeyinde doğrulandığı görülmüştür. Benzer biçimde, Karakoç Atlı'nın (2016) Avustralya, Şangay-Çin, Türkiye ve Fransa grupları arasında PISA 2012 matematik ve fen okuryazarlık testleri ile okuma becerileri testlerinden elde edilen puanların dil değişkenine göre ölçme değişmezliğinin sağlanamadığı sonucuna ulaştığı çalışmasıyla paralel sonuçlar elde edilmiştir.

Elde edilen bulgular ülke boyutunda değerlendirildiğinde, Güzeller'in (2011) araştırmasıyla benzer bulgular beklenmiştir. Ancak beklenenin aksi biçimde, PISA 2009 öğrenci anketi üzerinden yürütülen ölçme değişmezliği testinin ülke grupları arasında eşitliği sonucuna ulaşıldığı görülmüştür.

Türkiye örnekleminin ele alındığı çalışmalarda benzer sonuçlar gözlemlenmiştir. Bu çalışmalarda istatistiksel bölge, cinsiyet veya yıl değişkenlerine göre PISA ve TIMSS uygulamalarında kullanılan ölçeklerden toplanan puanların değişmezlikleri araştırılmış ve birçoğunda ölçme değişmezliğinin sağlandığı bulgusu elde edilmiştir (Uzun ve Öğretmen, 2010; Uyar ve Doğan, 2014; Başusta ve Gelbal, 2015; Ayvalli, 2016; Kıbrıslıoğlu Uysal ve Akın Arıkan, 2018). Ülkeler arası yürütülen çalışmalarda da benzer sonuçlar gözlemlenmiş, bu çalışmalarda farklı ülke grupları arasında PISA ve TIMSS uygulamalarında kullanılan ölçeklerden elde edilen puanların ölçme değişmezliği sağladığına dair kanıtlar çoğunlukla elde edilememiştir (Kıbrıslıoğlu, 2015; Ayvalli, 2016; Karakoç Alatl, 2016; Şekercioğlu ve Koğar, 2018). Buradan hareketle, uluslararası düzeyde uygulanan testlerin ülkelerin katılımcıları arasında eşit algılandığı ve yapılan ölçme işleminin eş olduğu varsayımından uzaklaşmaktadır. Türkiye örneklemini üzerinden yürütülen çalışmaların birçoğunda söz konusu varsayımın doğrulanması, sonuçlardaki bu ikili ayrışmanın en önemli etkeninin kültür olduğunu göstermektedir. Dolayısıyla birden fazla ülke ile gerçekleştirilecek değerlendirme çalışmalarının farklı dil ve kültürlere uyarlanması aşamasında titizlik gösterilmesi gerekliliği ortaya çıkmaktadır.

Sonuçlar genel olarak ele alındığında, TIMSS 2015 uygulaması dördüncü sınıf matematik başarı testinin ülkelere göre ölçme değişmezliğinin sağlanamaması, katılımcı ülkeler için yapılan sıralamaları ve buna bağlı getirilen yorumları tartışılabilir konuma getirmiştir. Ölçme aracının uygulamaya dahil edilen tüm bireyler için aynı anlama gelmediği, sonuçların karşılaştırılabilir olmadığı sonucuna ulaşılmıştır.

5.2. Öneriler

Araştırmadan elde edilen sonuçlar doğrultusunda, araştırma ve uygulamaya yönelik şu önerilere yer verilebilir:

- 1) Araştırma sonuçları, OECD üyesi ülkeler arasında ölçme değişmezliğinin sağlanmadığını göstermiştir. Bu nedenle araştırma, TIMSS uygulamasından elde edilen ülke puanlarının karşılaştırılabilir olması için asıl uygulama öncesinde ölçme değişmezliği testlerinin yapılmasının önemini ortaya koymuştur.
- 2) Araştırmanın analizleri, matematik başarı testi maddelerinin yer aldığı 7 numaralı kitapçık üzerinden yürütülmüştür. Diğer kitapçıklardan elde edilen veriler ile de ölçme değişmezliği incelemesi yapılabilir.
- 3) Araştırma, TIMSS 2015 matematik test puanlarına ait veriler üzerinden yürütülmüştür. Fen bilimleri başarı testi puanlarına ait veriler üzerinden ölçme değişmezliği analizi yapılarak TIMSS 2015 uygulamasının sonuçlarına dair genellemelere ulaşabilmek mümkün olacaktır.
- 4) Araştırmada yalnızca OECD üyesi ülkeler analize dahil edilmiştir. Tüm ülkelere ait verilerle yürütülecek bir araştırma TIMSS 2015 uygulamasına dair daha kapsamlı bilgiler verecektir.
- 5) Araştırma, ülkelerin başarı testi puanları üzerinden yürütülmüştür. Gelecek çalışmalarda cinsiyet, sosyo-ekonomik durum, kültür ve dil gibi farklı değişkenlerin dahil edildiği ölçme değişmezliği incelemesi yapılabilir.

KAYNAKÇA

Anderson, L.W. & Krathwohl, D.R. (2001). *Öğrenme öğretim ve değerlendirme ile ilgili bir sınıflama* (D.A.Özçelik, Çev.) Ankara: Pegem Akademi.

Asil, M., & Brown, G. T. L. (2015). Comparing OECD PISA reading in English to other languages: Identifying potential sources of non-invariance. *International Journal of Testing*, 16(1), 71-93.

Ayvallı, M. (2016). *PISA 2012 matematik okuryazarlığı testinin ölçme değişmezliğinin incelenmesi*. (Yayınlanmamış Yüksek Lisans Tezi). Akdeniz Üniversitesi, Antalya.

Başusta, N.B., ve Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(4), 80-90.

Bayır, E., Çakıcı, Y. ve Atalay, Ö. (2016). Fen bilimleri öğretmenlerinin bilimin doğasına ilişkin görüşleri: Bilişsel harita örneği. *Kastamonu Eğitim Dergisi*, 24(3), 1419-1436.

Bowen, N.K. & Guo, S. (2011). *Structural equation modeling*. Oxford: Oxford University Press.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Newyork: Guilford Publications.

Büyüköztürk, Ş. (2002). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem Akademi.

Büyüköztürk, Ş., Çakmak, E., Akgün, Ö.E., Karadeniz, Ş. ve Demirel, F. (2008). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.

Byrne, B.M. (1998). *Structural equation modeling with LISREL, PRELIS and SIMPLIS: basic concepts, applications and programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling, 14*(3), 464-504.

Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S.H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: a cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology, 5*(982).

Çakan, M. (2003). Geniş ölçekli başarı testlerinin eğitimdeki yeri ve önemi. *Eğitim ve Bilim, 28*(128), 19-26.

Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2010). *Sosyal Bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları*. Ankara: Pegem Akademi.

Ergül, H. (1999). Uzaktan öğretimde kalite verimlilik ve üretkenlik. *Anadolu Üniversitesi Açık Öğretim Fakültesi Kurgu Dergisi, 16*, 283-296.

Gierl, M.J. (2000). Construct equivalence on translated achievement tests. *Canadian Journal of Education, 25*(4), 280-296.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical care, 44*(11 Suppl 3), S78-S94.

Güzeller, C. (2011). A study of cross-cultural equivalence of computer attitude in PISA 2009 student questionnaire. *Education and Science, 36*(162), 320-327.

Hansson, A. & Gustafsson, J. E. (2013). Measurement invariance of socioeconomic status across migrational background. *Scandinavian Journal of Educational Research*, 57(2), 148-166.

Huck, S.W. (2007). *Reading statistics and research*. Boston: Pearson Education.

Horn, J. L. & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3), 117-144.

International Association for Evaluation of Educational Assessment. (2015). *TIMSS 2015: The trends in international mathematics and science study*. 25 Haziran 2018 tarihinde https://www.iea.nl/fileadmin/user_upload/Studies/TIMSS_2015/TIMSS_2015.pdf adresinden erişildi.

Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8: structural equation modeling with the simplis command language*. USA: Scientific Software International, Inc.

Karasar, N. (2005). *Bilimsel Araştırma Yöntemi*, Ankara: Nobel Yayıncılık.

Karakoç Alatlı, B. (2016). *Uluslararası Öğrenci Değerlendirme Programı (PISA-2012) okuryazarlık testlerinin ölçme değişmezliğinin incelenmesi*. (Yayımlanmamış Yüksek Lisans Tezi). Ankara Üniversitesi, Ankara.

Kıbrıslıoğlu, N. (2015). *PISA 2012 Matematik Öğrenme Modelinin Kültürlere ve Cinsiyete Göre Ölçme Değişmezliğinin İncelenmesi: Türkiye-Çin (Şangay)-Endonezya Örneği* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.

Kıbrıslıoğlu Uysal, N. & Akın Arıkan, Ç. (2018). Measurement invariance of science self-efficacy scale in PISA. *International Journal of Assessment Tools in Education*. 5(2), 325-338.

Korkmaz, H. (2004). *Fen ve teknoloji eğitiminde alternatif değerlendirme yaklaşımları*. Ankara: Yeryüzü Yayınevi.

LoBiondo-Wood, G. & Haber, J. (2014). *Nursing Research: Methods and critical appraisal for evidence-based practice*. Missouri: Elsevier Mosby.

MEB. (2014). *TIMSS 2011 ulusal matematik ve fen raporu: 8. sınıflar*. Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü.

Mark, B. A. & Wan, T.T.H. (2005). Testing measurement equivalence in a patient satisfaction instrument. *Western Journal of Nursing Research*, 27(6), 772-787.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.

Mertler, C. A. & Vannatta, R. A. (2005). *Advanced and multivariate statistical methods: Practical application and interpretation* (Third Edition). Los Angeles: Pyrczak.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.

Milfont, T. L. & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3 (1), 111-121.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory*. (Third Edition). New York: McGraw-Hill, Inc.

Putnick, D.L. & Bornstein, M.H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review* 41, 71-90.

Rutkowski, L., von Davier, M., & Rutkowski, D. (2013). *Handbook of international large-scale assessment: background, Technical issues, and methods of data Analysis*. CRC Press, Boca Raton.

Satorra, A. & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514.

Schoot, R., Lugtig, P. & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9 (4), 486–492.

Segeritz, M. & Pant, H.A. (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*, 73(4), 601-630.

Steenkamp, B., E., M. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1),78-107.

Suhr, D.D. (2006). *Exploratory or confirmatory analysis*. NC: SAS Institute.

Şekercioğlu, G. (2018). Measurement invariance: Concept and implementation. *International Online Journal of Education and Teaching (IOJET)*, 5(3), 609-634.

Şekercioğlu, G. ve Koğar, H. (2018). The examination of measurement invariance and differential item functioning of PISA 2015 cognitive tests in terms of the commonly used languages. *Novitas-Royal (Research on Youth and Language)*. 12(2), 152- 172.

Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik*. Ankara: Seçkin Yayıncılık.

Tucker, K. L., Özer, D. J., Lyubomirsk, S. & Boehm, J. K. (2006). Testing for measurement invariance in the satisfaction with life scale: A comparison of Russians and North Americans. *Social Indicators Research*, 78, 341–360.

Uluslararası Matematik ve Fen Eğilimleri Araştırması TIMSS (2015). *TIMSS 2015 ulusal matematik ve fen ön raporu*. 13 Eylül 2018 tarihinde http://timss.meb.gov.tr/wp-content/uploads/TIMSS_2015_Ulusal_Rapor.pdf adresinden erişildi.

Uyar, Ş. ve Doğan, N. (2014). PISA 2009 Türkiye örnekleminde öğrenme stratejileri modelinin farklı gruplarda ölçme değişmezliğinin incelenmesi. *Uluslararası Türk Eğitim Bilimleri Dergisi*, 2, 30-43.

Uzun, B. ve Öğretmen, T. (2010). Fen başarısı ile ilgili bazı değişkenlerin TIMSS-R Türkiye örnekleminde cinsiyete göre ölçme değişmezliğinin değerlendirilmesi. *Eğitim ve Bilim*, 35(155), 26-35.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1), 1-25.

Widaman, K. F. & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The 59 science of prevention: Methodological advances from alcohol and substance abuse research*, 17, 281-324.

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

Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the MI literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.

Yaşlıođlu, M.M. (2017). Sosyal bilimlerde faktör analizi ve geçerlilik: keşfedici ve doğrulayıcı faktör analizlerinin kullanılması. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 46, 74-85.



EKLER

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Ayşenur TAVLICA

Doğum Tarihi : 08.03.1995

Eğitim Durumu

Lisans Öğrenimi : Akdeniz Üniversitesi, Eğitim Fakültesi, İlköğretim Matematik Öğretmenliği, 2012-2016.

Yüksek Lisans Öğrenimi : Akdeniz Üniversitesi, Eğitim Bilimleri Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Lisans, 2016- Devam ediyor.

Bildiği Yabancı Diller : İngilizce

Bilimsel Faaliyetleri : Bıçak, B.; Gök, G.; Göçer, Ö. ve Tavlıca, A.. (2017). Yayımlanmış Makalelerin Bilimsel Araştırma Süreçlerine Uygunluk Düzeylerinin Değerlendirilmesi, Uluslararası Avrasya Eğitim Araştırmaları Kongresi'nde (EJER) sunuldu (11 - 14 Mayıs 2017).

İş Denevimi

Çalıştığı Kurumlar : Milli Eğitim Bakanlığı, İlköğretim Matematik Öğretmeni, 2017- Devam ediyor.

İletişim

E-posta adresi : aysenurtavlica@gmail.com

BİLDİRİM

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

Tezimin/Raporumun tamamı her yerden erişime açılabilir.

Tezim/Raporum sadece Akdeniz Üniversitesi yerleşkelerinden erişime açılabilir.

Tezimin/Raporumun 1 yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

.../.../.....

Ayşenur TAVLICA

İNTİHAL RAPORU


AYSENUR TAVLICA TEZ TESLİM SON BENZERLİK RAPORU

ORIJINALLIK RAPORU

% 16	% 11	% 5	% 13
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	kamuexpress.com İnternet Kaynağı	% 2
2	Submitted to Ankara University Öğrenci Ödevi	% 1
3	Submitted to Hacettepe University Öğrenci Ödevi	% 1
4	www.nucem.sk İnternet Kaynağı	% 1
5	acikarsiv.ankara.edu.tr İnternet Kaynağı	% 1
6	Submitted to Kocaeli Üniversitesi Öğrenci Ödevi	% 1
7	umitgunes.net İnternet Kaynağı	% 1
8	timss.meb.gov.tr İnternet Kaynağı	% 1
9	Submitted to TechKnowledge Turkey Öğrenci Ödevi	% 1


Dr. Öğr. Üyesi Gülçin ŞEKERÇİBELİ