



**TIMSS 2019 SEKİZİNCİ SINIF MATEMATİĞE YÖNELİK GÜVEN
ÖLÇEĞİNİN HIZALAMA METODUNA GÖRE ÖLÇME
DEĞİŞMEZLİĞİNİN İNCELENMESİ**

Yakup Dokur

YÜKSEK LİSANS TEZİ

EĞİTİM BİLİMLERİ ANABİLİM DALI

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

HAZİRAN, 2023

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren(....) ay sonra tezdin fotokopi çekilebilir.

YAZARIN

Adı : Yakup
Soyadı : DOKUR
Bölümü : Eğitimde Ölçme ve Değerlendirme
İmza :
Teslim tarihi :

TEZİN

Türkçe Adı : TIMSS 2019 Sekizinci Sınıf Matematiğe Yönelik Güven Ölçeğinin Hizalama Metoduna Göre Ölçme Değişmezliğinin İncelenmesi

İngilizce Adı : Investigation of Measurement Invariance of TIMSS 2019 Eighth Grade Mathematics Confidence Scale by Alignment Method

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazar Adı Soyadı: Yakup DOKUR

İmza:

JÜRİ ONAY SAYFASI

Yakup DOKUR tarafından hazırlanan “TIMSS 2019 Sekizinci Sınıf Matematiğe Yönelik Güven Ölçeğinin Hizalama Metoduna Göre Ölçme Değişmezliğinin İncelenmesi” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Mehtap ÇAKAN

(Eğitimde Ölçme ve Değerlendirme Bilim Dalı, Gazi Üniversitesi)

Başkan: Dr. Öğr. Üyesi Ceren MUTLUER

(Eğitimde Ölçme ve Değerlendirme Bilim Dalı, Bolu Abant İzzet Baysal Üniversitesi)

Üye: Prof. Dr. Şeref TAN

(Eğitimde Ölçme ve Değerlendirme Bilim Dalı, Gazi Üniversitesi)

Tez Savunma Tarihi: 08/05/2023

Bu tezin Eğitim Bilimleri Anabilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Şaban ÇETİN

Eğitim Bilimleri Enstitüsü Müdürü



Aileme

TEŞEKKÜR

Yüksek lisans çalışmam süresince yardımını hiçbir zaman esirgemeyen değerli danışmanım Prof. Dr. Mehtap ÇAKAN hocama teşekkürlerimi sunarım.

Tez savunma jürimde bulunarak değerli katkılar sağlayan Prof. Dr. Şeref TAN ve Dr. Öğr. Üyesi Ceren MUTLUER hocalarıma teşekkür ederim.

Yüksek lisans eğitimim boyunca her konuda yardımcı olan değerli hocalarıma teşekkür ederim. Eğitim hayatım boyunca bilgi ve deneyimlerini paylaşan üzerimde emeği olan saygıdeğer hocalarıma teşekkürü bir borç bilirim.

Çalışmam boyunca bana her zaman destek olan değerli arkadaşlarım Ar. Gör. Ömer KAMIŞ'a ve Nurullah ERYILMAZ'a teşekkür ederim.

Hayatım boyunca maddi ve manevi desteklerini esirgemeyen, her zaman yanımda olan annem, babam ve ablama sonsuz teşekkürlerimi sunarım.

**TIMSS 2019 SEKİZİNCİ SINIF MATEMATİĞE YÖNELİK
GÜVEN ÖLÇEĞİNİN HIZALAMA METODUNA GÖRE ÖLÇME
DEĞİŞMEZLİĞİNİN İNCELENMESİ**

(Yüksek Lisans Tezi)

Yakup Dokur
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
Haziran, 2023

ÖZ

Bu çalışmada TIMSS uygulamasında yer alan duyuşsal özelliklerden biri olan matematiğe yönelik güven ölçeği puanları araştırılmıştır. Araştırmada TIMSS 2019 sınavına katılan sekizinci sınıf öğrencilerinin cevapladığı öğrenci anketlerinden matematiğe yönelik güven ölçeği verileri ele alınmıştır. Bu doğrultuda araştırmanın amacı TIMSS 2019 sınavına katılan sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeği puanlarının belirlenen ülkelere göre ölçme değişmezliğinin incelenmesidir. İlişkisel tarama modelinde olan bu araştırma farklı kültürleri temsil etmesi açısından yedi ülke (Şili, Hong Kong, İsrail, Katar, Rusya, Türkiye ve İngiltere) üzerinde yürütülmüştür. Ülkelerin birbirleriyle karşılaştırılmasında önem arz eden ölçme değişmezliği bir geçerlik araştırmasıdır. Çalışmada ölçme değişmezliği çoklu grup doğrulayıcı faktör analizi (ÇGDFA) ve hizalama metodu kullanılarak test edilmiştir. Çoklu grup doğrulayıcı faktör analizine ait bulgular incelendiğinde matematiğe yönelik güven ölçeğinin yapısal, metrik ve kısmi ölçek değişmezliğini sağladığı görülmüştür. “Matematikte genellikle iyiyimdir”, “Matematik birçok sınıf arkadaşına göre bana daha zor gelir” ve “Zor matematik problemlerini çözmede iyiyimdir” maddelerinin kısmi ölçek değişmezliğini sağladığı bulunmuştur. Hizalama yöntemine ait bulgularda ise

model için deęişmezlięi saęlamayan gruplar faktör yükleri için %39,68; kesişimler için %46,03 bulunmuştur. Deęişmezlięi saęlamayan grupların ortalaması ise %42,86 bulunmuştur. Bu deęer önerilen sınırların üzerinde olduęundan dolayı hizalama yönteminde matematięe yönelik güven ölçeęinin ölçme deęişmezlięi istenen düzeye ulaşmamıştır. Hizalama yönteminde parametreler deęişmezlik indeksleri gibi olmadığı için ölçme deęişmezlięi saęlanmasa bile parametreler yorumlanabilir. Hizalama yönteminde matematięe yönelik güven ölçeęi için tüm kriterler birlikte deęerlendirildięinde faktör yüklerine göre “Matematik güçlü yönlerimden biri deęildir” maddesi deęişmezlik düzeyi en yüksek maddedir. Kesişimlere göre “Matematik birçok sınıf arkadaşına göre bana daha zor gelir” ve “Zor matematik problemlerini çözmede iyiyimdir” maddeleri deęişmezlik düzeyi en yüksek maddelerdir. Hem faktör yükleri hem kesişimlere göre “Öğretmenim bana matematikte iyi olduęumu söyler” maddesi deęişmezlik düzeyi en düşük maddeyi ifade etmektedir. Araştırmının sonucunda sekizinci sınıf öğrencilerinin matematięe yönelik güvenini ölçmeyi amaçlayan söz konusu ölçeęin ÇGDFA ile yedi ülke arasında yapısal ve metrik deęişmezlięi saęladığı; ölçeęin üç madde için ise kısmi ölçek deęişmezlięini saęladığı bulunmuştur. Hizalama metoduna göre ölçme deęişmezlięinin saęlanmadığı sonucuna ulaşılmıştır. Dięer bir deyişle matematięe yönelik güven ölçeęinin yedi ülke için eşdeęer olmadığı söylenebilir. Dolayısıyla bu ülkeler için matematięe yönelik güven puanlarının karşılaştırılmasında bu durum dikkate alınmalıdır.

Anahtar Kelimeler :Ölçme deęişmezlięi, TIMSS 2019, hizalama metodu, çoklu grup doğrulayıcı faktör analizi, matematięe yönelik güven.

Sayfa Adedi : xv + 83

Danışman : Prof. Dr. Mehtap ÇAKAN

**INVESTIGATION OF MEASUREMENT INVARIANCE OF TIMSS
2019 EIGHTH GRADE MATHEMATICS CONFIDENCE SCALE
BY ALIGNMENT METHOD**

(M.S. Thesis)

Yakup Dokur

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

June, 2023

ABSTRACT

In this study, confidence scale scores for mathematics, which is one of the affective features in the TIMSS application, were investigated. In the research, the data of the confidence scale in mathematics from the student questionnaires answered by the eighth grade students who participated in the TIMSS 2019 exam were discussed. In this direction, the aim of the research is to examine the measurement invariance of the eighth grade students' mathematics confidence scale scores who participated in the TIMSS 2019 exam according to the determined countries. This research, which is in the relational screening model, was conducted on seven countries (Chile, Hong Kong, Israel, Qatar, Russia, Turkey and England) in order to represent different cultures. Measurement invariance, which is important in comparing countries with each other, is a validity study. In the study, measurement invariance was tested using multiple group

confirmatory factor analysis (MGCFA) and alignment method. When the findings of the multi-group confirmatory factor analysis were examined, it was seen that the confidence scale for mathematics provided structural, metric and partial scalar invariance. It was found that the items "I am generally good at math", "Math is more difficult for me than many of my classmates" and "I am good at solving difficult math problems" provide partial scalar invariance. In the findings of the alignment method, the groups that noninvariance for the model were 39.68% for factor loadings; for intercepts, it was found 46.03%. The mean of the groups that noninvariance was found to be 42.86%. Since this value is above the recommended limits, the measurement invariance of the confidence scale for mathematics in the alignment method did not reach the desired level. In the alignment method, parameters can be interpreted even if measurement invariance is not ensured, as parameters are not like invariance indexes. When all criteria for the confidence scale in mathematics are evaluated together in the alignment method, the item "Mathematics is not one of my strengths" is the item with the highest level of invariance according to factor loadings. According to the intercepts, the items "Math is more difficult for me than many of my classmates" and "I am good at solving difficult mathematical problems" are the items with the highest level of invariance. According to both factor loadings and intercepts, the item "My teacher tells me I am good at mathematics" expresses the item with the lowest level of invariance. As a result of the research, it was found that the said scale, which aims to measure the confidence of eighth grade students in mathematics, provides structural and metric invariance between seven countries with MGCFA; It was found that the scale provided partial scalar invariance for three items. According to the alignment method, it was concluded that measurement invariance was not achieved. In other words, it can be said that the confidence scale for mathematics is not equivalent for seven countries. Therefore, this situation should be taken into account when comparing confidence scores for mathematics for these countries.

Key Words : Measurement invariance, TIMSS 2019, alignment method, multiple group confirmatory factor analysis, mathematics confidence.

Page Number : xv + 83

Supervisor : Prof. Dr. Mehtap ÇAKAN

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU	i
ETİK İLKELERE UYGUNLUK BEYANI	ii
JÜRİ ONAY SAYFASI.....	iii
TEŞEKKÜR.....	v
ÖZ	vi
ABSTRACT	viii
İÇİNDEKİLER.....	x
TABLolar LİSTESİ	xii
BÖLÜM I	1
GİRİŞ.....	1
Problem Durumu	1
Problem İfadesi	6
Araştırma Soruları.....	6
Araştırmanın Amacı	7
Araştırmanın Önemi.....	7
Sınırlılıklar	8
Sayıtlar	8
BÖLÜM II.....	9
KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR.....	9
Matematik Başarısına Etki Eden Duyuşsal Özellikler	9
Geçerlik.....	13
Faktör Analizi.....	14
Ölçme Değişmezliği	16
Kısmi Değişmezlik.....	19
Model Uyum İndeksleri	20
Hizalama Yöntemi.....	24

Hizalama Yönteminde Model Uyum Kriterleri	29
İlgili Araştırmalar	32
BÖLÜM III	42
YÖNTEM.....	42
Araştırmanın Modeli	42
Araştırmanın Çalışma Grubu	42
Verilerin Toplanması.....	43
Matematiğe Yönelik Güven Ölçeği.....	43
Verilerin Analizi	44
Doğrulamalı Faktör Analizi Varsayımları	45
Modelin Oluşturulması.....	47
Doğrulamalı Faktör Analizi Sonuçları	47
BÖLÜM IV	49
BULGULAR VE YORUM	49
Birinci Araştırma Sorusuna İlişkin Bulgular	49
İkinci Araştırma Sorusuna İlişkin Bulgular.....	53
BÖLÜM V	62
TARTIŞMA, SONUÇ VE ÖNERİLER	62
Birinci Araştırma Sorusuna İlişkin Sonuçlar.....	62
İkinci Araştırma Sorusuna İlişkin Sonuçlar	65
Öneriler	68
Araştırmacılara Yönelik Öneriler	68
Uygulayıcılara Yönelik Öneriler.....	69
KAYNAKLAR.....	70

TABLÖLAR LİSTESİ

Tablo 1. <i>Uyum İyiliği Katsayıları</i>	23
Tablo 2. <i>Çalışma Grubunu Oluşturan Öğrencilerin Ülkelere Göre Dağılımı</i>	43
Tablo 3. <i>Matematiğe Yönelik Güven Ölçeği Maddeleri</i>	44
Tablo 4. <i>Ülkelere İlişkin Doğrulamalı Faktör Analizi Uyum İndeksleri</i>	50
Tablo 5. <i>Çoklu Grup Doğrulamalı Faktör Analizi Sonucu Ölçme Değişmezliği Uyum İndeksleri</i>	50
Tablo 6. <i>Çoklu Grup Doğrulamalı Faktör Analizi Sonucu Kısmi Ölçme Değişmezliği Uyum İndeksleri</i>	52
Tablo 7. <i>Hizalama Yönteminde Matematiğe Yönelik Güven Ölçeği Maddelerinin Faktör Yüklerinin ve Kesişimlerinin Değişmezliğinin Sonuçları</i>	54
Tablo 8. <i>Hizalama Yönteminde Ölçme Parametrelerine Göre Değişmezliği Sağlamayan Ülke Sayıları</i>	55
Tablo 9. <i>Hizalama Yönteminde Kesişim ve Faktör Yüklerine İlişkin Uyum İstatistikleri</i> ...	56
Tablo 10. <i>Hizalama Yöntemi Metriğinde Maddelerin Kesişim ve Faktör Yüklerine İlişkin Parametre Tahminleri</i>	58
Tablo 11. <i>TIMSS 2019 8. Sınıf Matematiğe Yönelik Güven Ölçeğinin Gizil Ortalama Karşılaştırması ve Sıralaması</i>	60

ŞEKİLLER LİSTESİ

Şekil 1. TIMSS 2019 8. Sınıf matematiğe yönelik güven ölçeğine ait ölçme modeli47



SİMGELER VE KISALTMALAR LİSTESİ

χ^2 : Ki-kare

AGFI: Adjusted Goodness of Fit Index (Düzeltilmiş uyum iyiliği istatistiği)

CFI: Comparative Fit Index (Karşılaştırmalı Uyum İndeksi)

ÇGDFA: Çoklu Grup Doğrulayıcı Faktör Analizi

DFA: Doğrulayıcı Faktör Analizi

GFI: Goodness of Fit Index (Uyum İyiliği İstatistiği)

IEA: International Association for the Evaluation of Educational Achievement (Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu)

ML: Maximum Likelihood

MLR: Robust Maximum Likelihood

MTK: Madde-Tepki Kuramı

NFI: Normed Fit Index (Normlu Uyum İndeksi)

OECD: The Organisation for Economic Co-operation and Development (Ekonomik İşbirliği ve Kalkınma Örgütü)

PIRLS: Progress in International Reading Literacy Study (Uluslararası Okuma Becerilerinde Gelişim Projesi)

PISA: Programme for International Student Assessment (Uluslararası Öğrenci Değerlendirme Programı)

RMSEA: Root Mean Square Error of Approximation (Yaklaşık Hataların Ortalama Karekökü)

SRMR: Standardized Root Mean Square Residual (Standartlaştırılmış Artık Ortalamaların Karekökü)

TIMSS: Trends in International Mathematics and Science Study (Uluslararası Matematik ve Fen Eğilimleri Araştırması)

TLI: Tucker Lewis Index

VIF: Varyans Şişkinlik Faktörü

YEM: Yapısal Eşitlik Modellemesi

BÖLÜM I

GİRİŞ

Bu bölümde öncelikle problem durumundan bahsedilmiş ardından problem ifadesine, araştırmanın amacına, önemine ve sınırlılıklarına yer verilmiştir.

Problem Durumu

Eğitim ulusların her alanda ülkelerini en iyiler kategorisine taşımayabilmeleri, bilimsel ve teknolojik ilerleme kaydedebilmeleri için vazgeçilmez bir unsurdur. Her ülke kendine özgü eğitim politikası uygulayarak bu bağlamda kendini geliştirmektedir. Bu gelişim ve ilerlemeleri takip edebilmek için ulusal ve uluslararası düzeyde çeşitli araştırmalar yürütülmektedir. Uygulanan bu araştırmalar ülkelerin birbirlerine göre başarısını, konumunu ve hangi düzeyde bulunduğu hakkında bilgi vermektedir. Ulusal düzeyde ABİDE (Akademik Becerilerin İzlenmesi ve Değerlendirilmesi), uluslararası düzeyde ise PISA (Programme for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study) ve PIRLS (Progress in International Reading Literacy) gibi geniş ölçekli araştırmalar yapılmaktadır.

Geniş ölçekli araştırmalar, eğitim-öğretim sürecinin çeşitli aşamalarında var olan aksaklıkların belirlenip giderilmesinde etkili bir biçimde kullanılabilir. Bunun için, test

sonuçlarının farklı boyutlardan bakılarak analiz edilmesi, sonuçların doğru yorumlanması ve etkili kullanılması önem taşımaktadır. Test sonuçlarının doğruluk derecesini artırmak için elde bulunan verilerden yüksek oranda yararlanılması, verilerin analizinde farklı tekniklerin ve yaklaşımların kullanılması gerekmektedir (Çakan, 2003). Başta politika yapıcılar olmak üzere ülkeler, belirli periyotlarla düzenlenen bu uygulamalara katılmakta ve eğitim politikalarını bu çerçevede şekillendirmektedir.

Uluslararası düzeyde uygulanan TIMSS, PISA ve PIRLS gibi araştırmalar, ülkelerin öğrenci başarılarının değerlendirilmesi ve ülkelerin eğitim sistemleri arasındaki farklılıkları ortaya koyması bakımından araştırmaya katılan ülkelere önemli bilgiler sağlamaktadır.

Bunlardan PISA, The Organisation for Economic Co-operation and Development (OECD) tarafından üçer yıllık periyotlar halinde 15 yaş grubu öğrencilere uygulanır. PISA'nın amacı öğrencilerin modern toplumlara tam katılım için gerekli olan bilgi ve becerileri ne ölçüde edindiklerini değerlendirmektedir. Değerlendirme sadece öğrencilerin bilgiyi yeniden üretip üretemeyeceğini tespit etmekle kalmaz; aynı zamanda öğrencilerin öğrendiklerinden ne kadar iyi sonuçlar çıkarabileceklerini ve bu bilgileri hem okul içinde hem de dışındaki ortamlarda uygulayabileceklerini incelemektedir. PISA uygulamasında öğrencilere; matematik okuryazarlığı, fen bilimleri ve okuma becerileri alanlarında başarı testleri uygulanarak, öğrenme ortamları ve aileleriyle ilgili bilgiler toplamak için de çeşitli anketler uygulanmaktadır. Türkiye de OECD üyesi olduğu için PISA uygulamalarına 2003 yılından itibaren katılmaktadır (OECD, 2019).

Diğer yandan PIRLS, International Association for the Evaluation of Educational Achievement (IEA) tarafından her beş yılda bir 10 yaş grubu dördüncü sınıf öğrencilerine uygulanır. PIRLS'de uygulanan test ve anketler aracılığıyla öğrencilerin okuma alışkanlıkları, dil ve okuma becerileri tespit edilmektedir. PIRLS, öğretim ve öğrenmenin iyileştirilmesine yönelik politikalar ile ilgili en iyi bilgileri sağlayarak öğrencilerin başarılı

ve kendi kendine yeten okuyucular olmalarına yardımcı olmaktadır. (Mullis, Martin, Foy & Hooper, 2017). Türkiye, PIRLS uygulamalarına 2001 ve 2021 yıllarında katılmıştır.

TIMSS ise IEA tarafından dört yıllık periyotlar halinde gerçekleştirilerek dördüncü ve sekizinci sınıf düzeyindeki öğrencilerin matematik ve fen bilimlerinde kazandıkları bilgi ve becerilerin değerlendirilmesini amaçlamaktadır. TIMSS uygulamasında öğrencilerin başarı puanlarıyla birlikte uygulamaya katılan öğrencilere, bu öğrencilerin velilerine, öğretmenlerine, ve okul idarecilerine uygulanan bağlam anketleriyle ilgili de veri toplanmaktadır. Böylece, ülkelerin kendi eğitim sistemlerini değerlendirmesine ve uluslararası düzeyde karşılaştırmalı çalışmalar yapılmasına fırsat sunmaktadır (Milli Eğitim Bakanlığı [MEB], 2020).

Türkiye, TIMSS araştırmasına uzun süredir katılan ülkeler arasında yer almaktadır. 1995 yılından bu yana düzenlenen TIMSS uygulamasına Türkiye 1995 ve 2003 yıllarında her iki sınıf düzeyinde de katılmamış, 1999 ve 2007 yıllarında sadece sekizinci sınıf düzeyinde katılmış, 2011 ve 2015 yıllarında dördüncü ve sekizinci düzeylerinde son olarak da 2019 yılında beşinci ve sekizinci sınıf düzeylerinde katılmıştır. Türkiye TIMSS 2019 uygulamasına kadar fen ve matematik alanlarında ortalama başarısını zamanla artırmış ancak değerlendirmede referans kabul edilen ölçek orta noktasının (500 puan) altında ya da ölçek orta noktası düzeyinde performanslar sergilemiştir (MEB, 2020).

Uluslararası düzeyde uygulanan bu araştırmada ülkelerin başarı ya da başarısızlığının altında yatan nedenlerin araştırılması için de çeşitli bağlam anketleri uygulanmaktadır.

Hem dördüncü hem de sekizinci sınıf düzeyinde TIMSS'e katılan öğrencilere, matematik ve fen öğrenmeye yönelik deneyimleri ve tutumlarını belirlemek amacıyla çeşitli anketler uygulanmaktadır. Öğretmenler ve okul müdürleri, okul ve sınıf kaynakları hakkında veri sağlamak için anketler doldurmakta ve dördüncü sınıf öğrencilerinin ebeveynleri, öğrencilerin ev bağamları hakkında yine bir anket doldurmaktadır. Ulusal bağamları ve politikaları daha iyi anlamak için her bir katılımcı ülkeden temsilciler bir müfredat

anketi doldurarak TIMSS uygulamasına çeşitli katkılar sağlamaktadır. (Martin, von Davier & Mullis, 2020).

TIMSS 2019 bağlam anketlerinin amacı, geçmişle bağlantı kurarak geleceğe bir köprü inşa etmektir. Öğrenci başarısı ve matematikteki eğilimlerin bir araştırması olarak, TIMSS'in eğitim bağlamını karakterize etmedeki ilk önceliği, değerlendirmeden değerlendirmeye başarı değişikliklerini yorumlamaya yardımcı olabilecek etkenler hakkında veri toplamaktır. Bu tür bağlamsal veriler, bir değerlendirmeden diğerine değişiklikleri ölçmek için kullanılan bağlam anketi ölçekleri olarak ifade edilmektedir. Öğrenci başarısındaki değişiklikleri eğitim politikalarındaki veya uygulamadaki değişikliklerle ilişkilendirmek, politika veya uygulamanın öğrencinin öğrenmesi için faydalı olduğuna dair güçlü bir kanıt kaynağı olabilmektedir (Mullis & Martin, 2017).

TIMSS'e katılan öğrencilere uygulanan bağlamsal anketler çeşitli psikolojik yapılardan oluşmaktadır. Tutumlar, inançlar ve değerler gibi psikolojik yapılar, doğrudan ölçülemeyen gizil değişkenleri oluşturmaktadır. Psikolojik ölçümler gizil yapının göstergeleri olarak işlev görür. Gruplar arasında gizil bir yapıyı anlamlı bir şekilde karşılaştırmak için, gözlemlenen her gösterge tüm gruplarda gizil değişkenle aynı şekilde ilişkili olmalıdır (Milfont & Fischer, 2010).

Duyuşsal özellikler, her öğrenme sürecinde olduğu gibi matematik başarısında da önemli bir role sahiptir. Matematiğe yönelik kaygı, tutum, ilgi, öz yeterlik inancı ve benlik algısı gibi duyuşsal özellikler matematik başarısında oldukça önemlidir (Gülleroğlu, 2017). TIMSS uygulamasına katılan ülkelerde hem dördüncü hem de sekizinci sınıf öğrencilerinin matematiğe yönelik tutumları ile matematik başarıları arasındaki ilişkiler olumlu bulunmuştur (Mullis, Martin, Foy, Kelly & Fishbein, 2020). Oral ve McGivney (2013), TIMSS 2011 uygulaması Türkiye örneklemini ele aldıkları çalışmalarında matematiğe yönelik güven değişkeninin matematik başarısı ile doğru orantılı olduğu sonucuna ulaşmışlardır. Benzer şekilde Ertürk ve Erdiç Akan (2018), TIMSS 2015 uygulaması

Türkiye örnekleminde matematiğe yönelik güven değişkeninin hem dördüncü hem de sekizinci sınıf düzeyinde matematik başarısı ile ilişkili en önemli değişken olduğu sonucuna ulaşmışlardır. Matematiğe yönelik duyuşsal özelliklerden biri olan matematikte kendine güven değişkeni öğrencinin matematiği öğrenmede kendine inanması ve matematik dersinde iyi olduğunu düşünmesini ifade etmektedir (Demir & Kılıç, 2010). Stankov ve Lee (2014) güven değişkeninin akademik başarı ile en yüksek korelasyona sahip olduğunu ifade etmiştir. Bu çalışmada matematik başarısının altında yatan en önemli etmenlerden biri olan matematiğe yönelik güven ele alınmıştır.

TIMSS uygulamasında “Matematikte kendine güven anketi” aracılığıyla konuya özel benlik kavramı ölçülmektedir (Martin, von Davier & Mullis, 2020). Öğrenciler farklı konularda yetenekleri hakkında farklı görüşlere sahip olma eğilimindedirler. Aynı zamanda öz değerlendirmeleri, geçmiş deneyimleri ve akranlarına kıyasla kendilerini nasıl gördükleri bu anket yardımıyla tespit edilebilmektedir (Marsh & Craven, 2006).

TIMSS uygulamasında yer alan bağlamsal anketlerden toplanan bilgiler küresel eğitimde hesap verilebilirliği ilerletip öğrenci çıktılarına ilişkin önemli bilgiler sağlasa da, ülkeler arası karşılaştırmaların geçerliliği konusunda sorgulamalara yol açmıştır (Oliveri & von Davier 2011). Bağlamsal anketlerin altında yatan, çok tartışılan konulardan biri de ölçme değişmezliği veya ölçek puanı eşdeğerliği varsayımdır (Glassow, Rolfe & Hansen, 2021).

Psikolojik araştırmalar genellikle grupları psikolojik değişkenler üzerinden karşılaştırmayı esas alır. Özellikle kültürler arası araştırmalarda, bir kültür grubunda yeterli psikometrik özellikler gösterdiği tespit edilen bir araç, başka bir kültür grubuna çevrilir ve uygulanır. Grupları karşılaştırırken araştırmacılar genellikle aracın (anketler, yetenek ölçekleri) tüm gruplarda aynı psikolojik yapıyı ölçtüğünü varsaymaktadır. Bu varsayım genellikle doğru değildir ve test edilmesi gerekir. Bu nedenle ölçme değişmezliğinin test edilmesi son yıllarda önemli bir konu haline gelmiştir (Milfont & Fischer, 2010).

Ölçme değişmezliği ile ilgili sorunlar, araştırmacıların, gizil faktör ortalamalarını ve gizil değişkenler arasındaki ilişkileri ülkeler veya zaman bazında karşılaştırmayı gerektiren önemli soruları yanıtlamalarını sıklıkla engellemektedir. Uluslararası karşılaştırmalarda ölçme değişmezliği konusundaki endişenin nedeni, psikolojik özelliklerin veya yapıların kültürler arasında ölçülmesiyle ilgili zorluktan kaynaklanmaktadır. Kültürel faktörler, yanıtlayıcıların ölçme aracında yer alan soruları nasıl yorumladığını ve yanıtladığını etkileyebilmektedir. (Glassow vd., 2021).

Ülkeleri birbirine göre kıyaslamak için ölçme değişmezliği temel teşkil ettiğinden yapılan araştırmalarda ölçme değişmezliğinin sağlanması da büyük önem taşımaktadır. Bu bağlamda çalışmada TIMSS 2019 uygulamasında yer alan sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeği puan verileri kullanılarak ölçeğin ölçme değişmezliği çoklu grup doğrulayıcı faktör analizi ve yeni bir yöntem olan hizalama metoduyla incelenmiştir. Bu çalışma sayesinde matematiğe yönelik güven ölçeği puan verilerine göre yapılan karşılaştırmaların anlamlı olabilmesi, elde edilen sonuçların ve bu sonuçlara dayalı olarak yapılan yorumların geçerliği de araştırılmıştır. Çalışmanın bu yönüyle literatüre katkı sağlaması beklenmektedir.

Problem İfadesi

TIMSS 2019 uygulamasına katılan sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeği puan verileri belirlenen ülkelere göre ölçme değişmezliğini sağlamakta mıdır?

Araştırma Soruları

Çalışmanın araştırma soruları aşağıda verilmiştir;

- 1) TIMSS 2019 uygulaması sonucunda elde edilen veri seti için sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeği puanları Şili, Hong Kong, İsrail,

Katar, Rusya, Türkiye ve İngiltere arasında çoklu grup doğrulayıcı faktör analiziyle bakıldığında yapısal, metrik, skalar ve katı değişmezliği sağlamakta mıdır?

- 2) TIMSS 2019 uygulaması sonucunda elde edilen veri seti için sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeği puanları, hizalama metodu kullanıldığında Şili, Hong Kong, İsrail, Katar, Rusya, Türkiye ve İngiltere arasında ölçeğe değişmezliğini sağlamakta mıdır?

Araştırmanın Amacı

Bu araştırma TIMSS 2019 uygulamasına katılan sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeğinin belirlenen ülkelere göre ölçeğe değişmezliğinin incelenmesini amaçlamaktadır.

Bu araştırma sayesinde TIMSS 2019 uygulamasına sekizinci sınıf düzeyinde belirlenen ülkelerden katılan öğrencilerin matematiğe yönelik güven puanlarına göre sıralamalarının daha geçerli bir şekilde yapılması araştırmanın diğer bir hedefini oluşturmaktadır. Diğer bir ifadeyle araştırmanın amacı, matematiğe yönelik güven puanlarının belirlenen ülkelere göre hangi düzeyde olduğunun geçerli bir şekilde tespiti amaçlanmaktadır.

Araştırmanın Önemi

Ülkemizde ölçeğe değişmezliği ile ilgili çalışmaların sayısı son yıllarda artış gösterse de yeterli düzeye ulaşamamıştır. Ölçeğe değişmezliği çalışmaları bir geçerlik araştırması olduğu için önem arz etmektedir. Özellikle PISA, TIMSS, PIRLS gibi geniş ölçekli uluslararası uygulamaların geçerlik kanıtı olarak sunulması, eğitime yön vermesi açısından büyük önem taşımaktadır (Başusta & Gelbal, 2015). Literatür incelendiğinde TIMSS ve PISA gibi uluslararası araştırmalar ile ilgili yapılan çalışmalarda genellikle performans testleri ve anket verileri kullanılarak kültür, cinsiyet gibi farklı alt gruplara göre ölçeğe değişmezliği çoklu grup doğrulayıcı faktör analizi ile araştırılmıştır. Bu çalışmada TIMSS 2019 verileri ele

alınarak sınava katılan sekizinci sınıf öğrencilerinin cevapladığı öğrenci anketlerinden matematiğe yönelik güven ölçeği kullanılmıştır. Yapılan araştırmalara ek olarak hizalama metodu adı verilen yeni bir yöntemle de belirlenen ülkelere göre ölçme değişmezliği incelenmiştir. Hizalama yöntemi ölçme değişmezliğinin derecesi hakkında bilgi vermekte, hangi ölçme parametrelerinin yaklaşık olarak değişmez olduğu ve hangilerinin olmadığı da bu yöntemle belirlenebilmektedir. Yani hizalama yöntemi sayesinde ölçme aracındaki her bir değişkenin ölçme değişmezliğine ne kadar katkıda bulunduğu ile ilgili bilgiler elde edilebilmekte, ölçme aracının en değişmez değişkenlerini belirlemek için de kullanılabilmektedir (Asparouhov & Muthén, 2014). Bu yönüyle çalışmanın alana katkı sağlaması beklenmektedir.

Sınırlılıklar

Bu araştırma;

1. TIMSS 2019 sekizinci sınıf öğrenci anketlerinden matematiğe yönelik güven ölçeği
2. TIMSS 2019 uygulamasına sekizinci sınıf düzeyinde katılan yedi ülke

ile sınırlıdır.

Sayıtlar

Bu çalışmada TIMSS 2019 uygulamasına katılan sekizinci sınıf öğrencilerinin araştırmada yer alan maddelere içtenlikle cevap verdikleri varsayılmaktadır.

BÖLÜM II

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde ilk olarak matematik başarısına etki eden duyuşsal özellikler, geçerlik, faktör analizi, ölçme değişmezliği ve hizalama yöntemi kavramları ele alınarak açıklanmıştır. Daha sonra çalışmayla ilişkili olarak ilgili araştırmalara yer verilmiştir.

Matematik Başarısına Etki Eden Duyuşsal Özellikler

Öğretme ve öğrenme karmaşık bir süreçtir ve birçok faktörden doğrudan veya dolaylı olarak etkilenir. Öğrenmenin nasıl gerçekleştiğini tam olarak açıklayan tek bir model bulunmasa da, öğretme ve öğrenme sürecinde öğrenci özellikleri ve okul faktörlerinin önemli unsurlar olduğu açıktır (Akyüz, 2014).

Öğrenme güdüsünün kaynağı olarak görülen ve öğrenme düzeyini etkileyen öğrenci niteliklerinden biri duyuşsal giriş özellikleridir. Duyuşsal giriş özellikleri, öğrencinin öğrenilecek öğrenme birimine yönelik göstereceği ilgi, tutum ve özgüveninden oluşan bir bileşkedir. İlgilere, tutumlara ve kendi görüş tarzlarına göre bireylerin duyuşsal bakımdan neleri öğrenmeye hazırlıklı oldukları tespit edilirse, öğrencilerin daha kolay ve daha hızlı öğrenmeleri sağlanabilir. Öğrenmede anahtar bir rol oynayan duyuşsal özellikler akademik olarak bireylerin başarılı olabilmeleri için gerekli eğilimlerden biri olarak karşımıza

çıkılmaktadır. Duyuşsal özellikler bireylerin düşünme, anlama, harekete geçme durumlarında büyük önem teşkil etmekte ve bireyler arasındaki öğrenme farklılıkların kaynağının yaklaşık dörtte birini oluşturmaktadır (Bloom, 1976/2016).

Duyuşsal giriş öğelerinden biri olan akademik özgüven öğrencinin öğrenme özgeçmişine bağlı olarak herhangi bir öğrenme birimini öğrenip öğrenemeyeceğine ilişkin kendini algılama biçimidir. Öz yeterlik ise bireyin karşılaşılabileceği güçlüklerde, olayın üstesinden gelip gelemeyeceğine yönelik etkinlikler düzenlemesi ve bu etkinlikleri başarılı bir şekilde yapabilme kapasitesine ilişkin kendi öz yargısıdır (Senemoğlu, 2013). Öz-yeterlik, bireyin belirli bir görevi tamamlamak için gereken eylemleri başarılı bir şekilde üstlenme yeteneklerine ilişkin algılarıyla ilgilidir (Ferla, Valcke & Cai, 2009). Benlik kavramı, kişinin belirli bir yetkinlik alanıyla ilişkili olarak benlik algısını ifade eder. (Stankov, Morony & Lee, 2014). Benlik kavramının ölçüleri, kişilik ölçülerine benzerdir; yani kişinin kendi alışılmış eğilimleri hakkındaki inançlarına atıfta bulunur. Benlik kavramı, belirli bir akademik alana özgü olması bakımından daha spesifikdir (Morony, Kleitman, Lee & Stankov, 2013). Kaygı, öz-yeterlik ve benlik kavramının zıttı olarak görülebilir ve gerçekten de kaygı, bu yapılarla olumsuz bir ilişki kurma eğilimindedir. Diğer bir deyişle kaygı kavramsal olarak kendinden şüphe duymaya benzer. Kaygı, kişinin bir görevi düşündüğü veya yerine getirdiği zaman fizyo-duygusal tepkilerini ifade eder (Stankov, Morony & Lee, 2014).

Bazı kaynaklarda özyeterlik ve güven sıklıkla birbirinin yerine kullanılsa da, aslında benzer ancak özdeş olmayan kavramları temsil ederler (Çiftçi & Yıldız, 2019). Kavramsal olarak güven ve öz-yeterlik birbirine yakın kavramlardır. Bilişsel olmayan ölçümler arasında, güven ve öz-yeterliğin en yüksek korelasyona sahip olduğu bulunmuştur. İki kavram arasındaki en önemli fark, güveni değerlendirirken katılımcının bir problemin çözümünü bulması gerekmezken, öz-yeterliği değerlendirirken çözümün kaydedilmesi gerektiğidir. Öz-yeterlik, kişinin sonuç üretme yeteneğine olan inancını ifade eder. (Stankov, Morony &

Lee, 2014). Kısaca iki kavram arasındaki temel fark, maddenin ölçülüp ölçülmediğidir. Bu nedenle özyeterlik ve güven farklı yapıları ifade etmektedir (Stankov, 2013). Böylece her iki kavram da incelendiğinde öz-yeterliğin belirli bir duruma özgü bir inanç olduğu, güvenin ise genel bir kişilik özelliği olduğu söylenebilir (Çiftçi & Yıldız, 2019).

Benlik kavramı ve öz-yeterlik yapıları üzerine yapılan pek çok çalışma birbirinden bağımsız olarak ilerlemiştir. Sonrasında hem kavramsal analizler hem de bu iki yapıyı karşılaştıran deneysel kanıtlar, bu değişkenlerin farklı yapıları temsil ettiğini göstermektedir (Bong & Skaalvik, 2003; Ferla, Valcke & Cai, 2009; Lee, 2009). Ferla, Valcke & Cai (2009) araştırmalarında öğrencilerin akademik benlik kavramının, akademik öz-yeterlik inançlarını güçlü bir şekilde etkilediğini tespit etmişlerdir. Akademik benlik kavramı, duygusal-motivasyonel değişkenler için daha iyi bir yordayıcı iken, akademik öz-yeterliğin akademik başarı için daha iyi bir yordayıcı olduğu sonucuna ulaşmışlardır.

Stankov, Morony ve Lee (2014) matematik başarısını yordamayı amaçladıkları çalışmalarında benlik kavramı, öz-yeterlik ve matematik kaygısı ölçeklerine güveni de ekleyerek öz inanç yapısı oluşturmuşlardır. Araştırma bulguları kendine inanç ve güvenin aynı yapıyı tanımladığı ve dört kendine inanç değişkeninin tümüyle çoklu regresyon analizi yapıldığında, güvenin diğer kendine inanç değişkenlerine göre yordayıcılığının daha fazla olduğunu bulmuşlardır. Bu nedenle, tek başına güvenin, matematik performansındaki diğer öz inanç ölçütlerinin toplamından daha fazla varyansı açıklayabildiği sonucuna ulaşılmıştır.

Eğitimin amacı, öğrencilerin akademik başarısını geliştirmektir. Bunun için en önem verilmesi gereken konulardan biri de öğrencilerin öz inançlarını geliştirmeye yönelik çalışmalara odaklanmaktır. (Stankov & Lee 2014). Baker ve LeTendre (2005), öğrencilerin matematik başarısının ülkelerin gelecekteki ekonomik gücü ile yakından ilgili olduğunu vurgulamaktadır. Böylece matematik başarısının geliştirilmesinin önemini ifade etmişlerdir. Buradan hareketle matematik başarısının geliştirilmesinde duyuşsal özelliklerin önemli bir etkiye sahip olduğu söylenebilir.

Bireylerin matematiğe yönelik olarak geliştirdikleri tutumları, inançları ve duyguları, genellikle matematiğe olan ilgilerinde, tepkilerinde ve bireysel yaşamlarında matematiği kullanmalarında önemli bir rol oynamaktadır. Örneğin, matematikte kendine daha fazla güvenen öğrencilerin, karşılaştıkları çeşitli bağlamlarda matematiği kullanma olasılıkları diğerlerinden daha yüksektir. Matematiğe karşı olumlu duygular besleyen öğrenciler, matematiğe karşı kaygı duyan öğrencilere göre matematiği daha iyi öğrenebilecek konumdadırlar. Bu nedenle matematik eğitiminin amaçlarından biri de, öğrencilerin matematiği başarılı bir şekilde kullanmaları ve kişisel-sosyal faydaları için daha fazla matematik öğrenmelerini sağlayan tutumlar, inançlar ve duygular geliştirmeleridir (OECD, 2013).

Öğrencilerin matematik öğrenme ve öğretme sürecinde tutum ve inançlarla ilgili eğilimleri alan bilgisi kadar büyük bir öneme sahiptir (Wilkins & Ma, 2003). Literatür incelendiğinde yapılan araştırmalar öğrencilerin matematik becerilerinde ve öğrenme sürecinde matematik başarısının bilişsel olmayan en iyi yordayıcısının güven olduğunu (Akyüz, 2014; Arıkan, van de Vijver & Yağmur, 2016; Çiftçi & Yıldız, 2019; Doğan & Barış, 2010; Koç, 2019; Morony, Kleitman, Lee & Stankov, 2013; Sarı, Arıkan & Yıldızlı, 2017; Stankov & Lee 2014; Stankov, 2013; Stankov, Lee, Luo & Hogan, 2012; Stankov, Morony & Lee, 2014) göstermektedir. Bir öğrencinin kendine güveni, önemli ölçüde gelecekte kendine güvenin gelişimini, aynı zamanda başarı yönelimini ve başarısının gelişimini öngörmektedir (Hannula, Maijala & Pehkonen, 2004).

Galbraith ve Haines (1998) matematiğe yönelik güveni şu biçimde ifade etmişlerdir:

Matematik güveni yüksek öğrenciler, çabalarının karşılığını aldıklarına inanırlar, zor konuları öğrenmek için endişelenmezler, iyi sonuçlar almayı beklerler ve bir ders olarak matematik konusunda kendilerini iyi hissederler. Kendine güveni düşük öğrenciler yeni materyal öğrenmek konusunda gergindirler, tüm matematiğin zor olacağını düşünürler, doğal olarak matematikte zayıf olduklarını düşünürler ve matematik hakkında diğer tüm derslerden daha fazla endişe duyarlar. (s.278).

Geçerlik

Geçerlik, bir testin ölçmek istediği özelliği başka özelliklerle karıştırmadan tam ve doğru olarak ölçebilme derecesi olarak ifade edilmektedir (Tekin, 2010). Bilindiği üzere farklı geçerlik türleri mevcuttur ve ihtiyaca göre bu geçerlik türlerinden biri veya daha fazlası test edilmektedir. Cronbach ve Meehl (1955) kapsam geçerliği, ölçüt bağıntılı geçerlik ve yapı geçerliği olmak üzere üç farklı geçerlik türünden bahsetmiştir.

Cronbach tarafından geçerliğin kanıta dayalı olduğu fikri ortaya atılmış; geçerliği, kanıt toplama süreci olarak ifade etmiştir (Crocker & Algina, 1986). Kane (1992)'e göre geçerlik test puanlarının yorumu ile ilişkilidir. Yorum, puanlardan kararlara götüren bir argümanı içermektedir. Argümandaki çıkarımlar ve varsayımlar için kanıtlar yorumu desteklemelidir.

Geçerlik bu süreçte kanıt kaynakları etrafında şekillenmeye başlamış ve gelişim göstermeye devam etmiştir. 1999 standartları ve bunun paralelinde hazırlanan 2014 standartlarında bu kanıt kaynakları açıkça belirtilmiş, yeni standartlara geçişin önemi böylece ortaya çıkmıştır. Geçerlik, kanıtların ve teorilerin testlerin önerilen kullanımları için test puanlarını destekleme derecesini ifade etmektedir. Bu nedenle geçerlik testlerin geliştirilmesi ve testlerin değerlendirilmesinde en temel husustur. Geçerleme süreci önerilen puan yorumlarına sağlam bir bilimsel temel sağlamak için ilgili kanıtların toplanmasını içermektedir (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014).

Messick (1995), geçerliğin merkezinde yapı geçerliğinin bulunduğunu diğer geçerlik türlerinin, yapı geçerliğinin alt boyutları olduğunu ifade etmektedir. Tüm boyutlar ile ilgili kanıtların toplandığı yapı geçerliğinin bütünleşik bir geçerlik olduğunu belirtmektedir. Kline (2000) da yapı geçerliliğinin diğer yaklaşımları içerdiğini, dolayısıyla tüm geçerlik türlerinin yapı geçerliğinin değerlendirilmesi ile ilgili olduğunu belirtmektedir.

Yapı geçerliğine dayalı çalışmalarda ise tanımlanması gereken ilk kavram psikolojik yapılardır. Cronbach ve Meehl (1955) yapıyı, bireylerin var olduğu kabul edilen özellikleri olarak tanımlamışlardır.

Yapı geçerliği ölçme aracını oluşturan parçaların birbiriyle ilişkilerinden yola çıkarak maddelerin belirli bir yapıyı sergileyip sergilemediklerini belirleme yani araçla ölçülmek istenen kuramsal yapıyı ortaya koyabilme derecesidir (Tekin, 2010). Bireyin tutum, performans, yetenek gibi psikolojik özelliklerini ölçmek amacıyla çok sayıda ölçülebilir, gözlenebilir sorular oluşturularak hazırlanan soruların belirtilen özellikleri ne derece doğru ölçtüğü, yapı geçerliğiyle ilgilidir. Yapı geçerliğine dair kanıt elde etmek için en sık kullanılan tekniklerden biri de faktör analizidir (Büyüköztürk, 2012).

Faktör Analizi

Faktör analizi, birbiriyle ilişkili çok sayıda değişkeni bir araya getirerek kavramsal olarak anlamlı daha az sayıda yeni değişken bulmayı ve keşfetmeyi amaçlayan çok değişkenli bir istatistiktir (Büyüköztürk, 2012).

Crocker ve Algina (1986) faktör analizinin üç temel amacının olduğunu belirtmiştir:

- (i) Bir modele ait bir dizi testin tüm çiftleri arasındaki korelasyonlar için hesaplanması gereken ortak faktörlerin sayısına karar vermek.
- (ii) Test iç korelasyonları için hesaplanan ortak faktörlerin niteliğine karar vermek.
- (iii) Ortak faktör varyansı ile ilişkili bir gözlenen değişkene ait varyans oranını belirlemek.

Faktör analizi temel olarak açıklayıcı ve doğrulayıcı faktör analizi olmak üzere iki boyutta ele alınmaktadır. Açıklayıcı faktör analizi, yeni geliştirilmekte olan veya başka bir dile çevrilmiş olan ölçeğin faktör yapısını ortaya koymayı amaçlar. Doğrulayıcı faktör analizi ise

önceden geliştirilmiş bir ölçeğin yeni bir araştırmada kullanıldığında orijinal faktör yapısına uygunluğunu tespit etmeye yaramaktadır (Suhr, 2006).

Açımlayıcı faktör analizi değişkenler arası ilişkilerden yola çıkarak faktör bulmaya yönelik bir işlemdir. Doğrulayıcı faktör analizinde ise değişkenler arasındaki ilişkiye yönelik daha önceki hipotez ya da kuramın test edilmesi söz konusudur (Büyüköztürk, 2012). Doğrulayıcı faktör analizi, modelin tüm özelliklerinin önceden bilinmesini gerektirdiğinden açımlayıcı faktör analizinden farklıdır, dolayısıyla faktör modelinin değerlendirilmesi için güçlü bir deneysel temel gereklidir. Bu nedenle, çoğunlukla açımlayıcı faktör analizi uygulaması yapı geçerliliği doğrulaması sürecinde önceki aşamalarda kullanılır. Doğrulayıcı faktör analizi önceden geliştirilmiş bir ölçeğin ilgili kuramsal yapıyı ölçüp ölçmediğini doğrulamak amacıyla kullanılmaktadır (Brown, 2015).

Brown (2015)'a göre doğrulayıcı faktör analizi aşağıda belirtilen bir dizi sonucu vermektedir:

- (a) Faktör yükleri (yani, her bir madde ile gözlenemeyen ya da gözlemlenmek istenen değişken arasındaki ilişki),
- (b) Hatalar (yani, gözlemlenemeyen faktörlerle açıklanamayan her bir maddenin değişkenliği),
- (c) Kesişimler (eşikler) (her madde için bir yanıt kategorisinden diğerine geçme olasılığı; örneğin, likert tipi bir ölçek için “bazen” kategorisinden “her zaman” kategorisine geçme olasılığı gibi).

Çoklu grup doğrulayıcı faktör analizi ise doğrulayıcı faktör analizinin özel bir uygulama alanıdır. Bu analiz yöntemi, aynı anda birden fazla grup için yapısal parametrelerin eşitliğinin test edilmesini sağlar. Bu bağlamda, gruplar arasındaki eşitliğin faktör yapısı açısından değerlendirilmesi ölçme değişmezliği olarak da ifade edilmektedir (Brown, 2015).

Çoklu grup doğrulayıcı faktör analizi ölçme değişmezliğini test etmek için yaygın olarak

kullanılan bir yöntemdir. (Vandenberg & Lance, 2000; Kline, 2011; Milfont & Fischer, 2010). Çoklu grup doğrulayıcı faktör analizi, klasik doğrulayıcı faktör analizinin bir devamıdır ve çoklu grup karşılaştırmasına dayanır. Verileri gruplara ayırır ve her biri için modeli belirler (Kline, 2011).

Ölçme Değişmezliği

Psikolojik ve gelişimsel araştırmalarda hızla zorunlu hale gelmekte olan ölçme değişmezliğinin kavramsal önemi literatüre günümüzde daha fazla girmiştir (Putnick & Bornstein, 2016). Yirmi birinci yüzyılın başında metodolojistler özellikle yapısal eşitlik modelleme çerçevesi içinde ölçme değişmezliğinin önemine vurgu yaparak dikkatleri üzerinde toplamışlardır (Steenkamp & Baumgartner, 1998; Vandenberg, 2002; Putnick & Bornstein, 2016). Psikolojik bir yapı farklı kültürel gruplarda ölçüldüğünde, puanların anlamlarının aynı olacağı ve puanların gruplar arasında karşılaştırılabileceği varsayılmaz (van de Vijver & Leung, 1997). Bu nedenle, tasarlanan araçlar ve psikolojik yapılar, incelenen kültürel gruplar arasında denklikleri açısından istatistiksel olarak test edilmelidir (Byrne, 2015). Kültürlerarası araştırmaların genellikle farklı kültürel gruplar arasındaki psikolojik yapıların karşılaştırılmasına odaklandığı dikkate alındığında psikolojik ölçmelerin denkliği de önem kazanmıştır (Milfont & Fischer, 2010).

Byrne ve Watkins (2003), ölçme değişmezliğini, ölçme aracında yer alan maddelerin farklı gruplar arasında aynı şekilde algılanıp yorumlanması olarak tanımlamışlardır. Ölçme değişmezliği, gizil bir yapıyı ortaya koymak için geliştirilmiş olan ölçeğin psikometrik bir özelliğidir. Aynı yapı ülkeler, kültürel birimler, zaman unsurları veya ülkeler içindeki bölgeler gibi farklı gruplar arasında aynı şekilde ölçüldüğünde ölçümün değişmez olduğu kabul edilir (Vandenberg & Lance, 2000; Millsap, 2012).

Ölçme değişmezliğini test etmek için Jöreskog (1971) tarafından geliştirilen tipik olarak faktör yapılarının karşılaştırılabilirliğinin değerlendirilmesi stratejisi izlenir. Stratejisinde, iç içe modeller, parametre kısıtlamalarının birer birer eklenmesini gerektiren azalan sayıda parametre (veya artan serbestlik derecesi) ile hiyerarşik bir düzende ilerler. Giderek daha kısıtlayıcı olan bu modeller verilerin modele uyup uymadığı konusunda test edilir (Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000).

Ölçme değişmezliği, madde tepki kuramına (MTK) dayalı olarak veya yapısal eşitlik modellemesi (YEM) çerçevesinde test edilebilir (Putnick & Bornstein, 2016). Ölçme değişmezliğini yapısal eşitlik modellemesine dayalı olarak test eden yaklaşımlardan biri olan çoklu grup doğrulayıcı faktör analizi hiçbir kısıtlamanın olmadığı modelden en kısıtlı modele doğru gruplar arası karşılaştırma yaparak parametrelerin değişmezliğini inceleyen aşamalı bir süreçtir (Horn & McArdle, 1992). Çoklu grup doğrulayıcı faktör analizi, farklı karşılaştırılabilirlik düzeylerinin, yani yapısal değişmezlik, metrik değişmezlik ve skalar değişmezlik gibi keşfedilmesi gereken ölçme değişmezliğini test etmek için yaygın olarak kullanılır (Kline, 2011; Vandenberg & Lance, 2000).

Meredith'e (1993) göre ölçme değişmezliğinin test edilmesi için dört aşama bulunmaktadır. Bu aşamalar sırasıyla şu şekildedir:

1. Yapısal değişmezlik (Biçimsel değişmezlik)
2. Metrik değişmezlik (Zayıf faktöriyel değişmezlik)
3. Skalar değişmezlik (Güçlü Faktöriyel değişmezlik-Ölçek değişmezliği)
4. Katı değişmezlik (Hata varyansı değişmezliği)

Yapısal (Konfigürasyon) değişmezlik: Bu model, ölçme değişmezliğini oluşturmanın ilk adımudur ve temel model yapısının gruplar arasında değişmez olması durumunda tahmin edilir. Bu durum farklı gruplardan katılımcıların yapıları aynı şekilde kavramsallaştırdığını

gösterir (Milfont & Fischer, 2010). Yalnızca sabit ve serbestçe tahmin edilen parametrelerin aynı modelinin gruplar arasında ne ölçüde geçerli olduğunu ve dolayısıyla hiçbir eşitlik kısıtlaması uygulanmadığını test etmede kullanılır (Byrne, 2015). Bu model, faktöriyel yapı gruplar arasında aynı olacak şekilde kısıtlanarak test edilir (Milfont & Fischer, 2010). Yapısal model, sonraki tüm denklik testlerinin karşılaştırıldığı temel model olarak hizmet etmesi bakımından önemlidir (Byrne, 2015).

Metrik değişmezlik: Bu model, farklı grupların maddelere aynı şekilde cevap verip vermediğini test eder. Metrik değişmezlik sağlanırsa, elde edilen derecelendirmeler gruplar arasında karşılaştırılabilir ve gözlenen madde farklılıkları, temeldeki gizil yapının grup farklılıklarını gösterecektir. Araştırmalar, değişmezlik testine devam etmeden önce en azından kısmi metrik değişmezliğin oluşturulması gerektiğini ileri sürmektedir (Vandenberg & Lance, 2000). Bu model, tüm faktör yükleri gruplar arasında aynı olacak şekilde sınırlanarak test edilir (Milfont & Fischer, 2010). Bu aşamada, farklı grupların maddeleri aynı şekilde cevapladığı böylece her bir gruptan gelen puanları karşılaştırmanın da anlamlı olduğu sayıltıları bulunmaktadır (Steenkamp & Baumgartner, 1998).

Skalar değişmezlik: Gizil araçları karşılaştırmak için skalar veya kesişim değişmezliği gereklidir. Skalar değişmezliğin oluşturulması, gözlenen puanların gizil puanlarla ilişkili olduğunu gösterir; yani, gizil yapıda aynı puana sahip olan bireyler, grup üyeliğinden bağımsız olarak gözlenen değişkende aynı puanı alırlar. Bu model, maddelerin kesişimlerinin gruplar arasında aynı olacak şekilde sınırlandırılmasıyla test edilir. Bu model gruplar arasında puanları karşılaştırmak için gerekli olan son modeldir. Tüm ek testler isteğe bağlıdır ve belirli bağlamlarda teorik olarak anlamlı olabilir (Milfont & Fischer, 2010). Bu değişmezlik düzeyi, farklı gruplardan gelen katılımcıların yanıt ölçeğini benzer şekilde yorumladıkları fikrini desteklemektedir. (Millsap & Olivera-Aguilar 2012). Alt gruplar arasında gizil değişken ortalamalarını ve varyanslarını karşılaştırmak için skalar değişmezlik gereklidir (Millsap, 2012). Gizil yapıların ortalamaları karşılaştırıldığında ölçek

değişmezliği de sağlanmış olur (Meredith, 1993). Gruplar arasında anlamlı karşılaştırmalar yapmak için bir yapı, skalar düzeyde değişmezliği veya daha yüksek olanı karşılamalıdır.

Katı değişmezlik (Hata varyansı değişmezliği): Bu model ölçme hatası miktarının ülkeler arasında değişmez olduğunu göstermektedir (Steenkamp & Baumgartner, 1998). Gruplar arasında her bir öge için aynı düzeyde ölçme hatasının mevcut olup olmadığını test etmek için, tüm hata varyansları gruplar arasında eşit olacak şekilde sınırlandırılır (Milfont & Fischer, 2010).

Kısmi Değişmezlik

Kısmi ölçme değişmezliği, bazı gözlenen değişkenler ile bu değişkenlere ait gizil değişkenler arasındaki ilişkiler gruplar arasında değişmezliği sağlarken diğer bazı göstergelerin bu değişmezliği sağlamaması biçiminde tanımlanmaktadır (Byrne, Shavelson & Muthen, 1989).

Kısmi ölçme değişmezliği testleri, hangi faktör yüklerinin veya kesişimlerin gruplar arasında farklılık gösterdiğini belirlemeyi amaçlamaktadır. Gruplar arasında bunlardan yalnızca biri farklıysa, gizil değişken üzerindeki herhangi bir farklılığın bu yük veya kesişimdeki bir farktan veya gerçek gizil değişken grubu farkından kaynaklanabilmektedir. Gruplar arasında eşit olarak kısıtlanmış en az iki faktör yükü/kesişim olduğu sürece, modeldeki gizil faktör ortalamaları arasındaki farklar hakkında geçerli çıkarımlar yapılabilmektedir (Byrne, Shavelson & Muthén, 1989).

Kısmi ölçme değişmezliğinde model uyumu için ilk olarak en büyük modifikasyon indeksine sahip madde serbest bırakılarak söz konusu model daha sonra tekrar analiz edilir ve değerlendirilir. Belirtilen model yine verilere uymuyorsa, kalan en büyük modifikasyon indeksine sahip madde bir sonraki modelde serbest bırakılır. Bu süreç, büyük modifikasyon indeksine sahip değer kalmayana veya model verilere uyana kadar tekrarlanır (Yoon & Kim, 2014).

Ölçme değişmezliği çalışmalarında model veri uyumunun kontrolü için birtakım uyum indeksleri bulunmaktadır.

Model Uyum İndeksleri

Literatürde model-veri uyumunun değerlendirilmesine yönelik çeşitli görüşler bulunmaktadır. Brown'a (2015) göre modellerin değerlendirilmesinde ortak kabul kuralları bulunmamakla birlikte genel olarak farklı türde uyum indekslerinin değerlendirilmesi gerektiği belirtilmektedir. Tasarlanan modele ilişkin uyumun değerlendirilmesinde bazı uyum indekslerinden faydalanılır. Bu uyum indeksleri mutlak uyum indeksleri ve karşılaştırmalı (artımlı, fazlalık) uyum indeksleri olmak üzere ikiye ayrılmaktadır.

Mutlak uyum indeksleri, ön modelin örneklem verisi ile ne kadar iyi uyduğunu belirler ve önerilen modelin ne derece uyuma sahip olduğunu gösterir (McDonald & Ho, 2002). Bu ölçümler, önerilen teorinin verilere ne kadar iyi uyduğunun en temel göstergesini sağlar. Karşılaştırmalı uyum indekslerinden farklı olarak, hesaplamaları bir temel modelle karşılaştırmaya dayanmaz, bunun yerine modelin hiçbir modelle kıyaslanmadan ne kadar iyi uyum sağladığının bir ölçüsüdür (Jöreskog & Sörbom, 1993). Bu kategoriye Ki-kare (χ^2) testi, RMSEA, GFI, AGFI ve SRMR dahildir.

Karşılaştırmalı uyum indeksleri ki-kare değerini doğrudan kullanmak yerine bu değeri referans bir model ile kıyaslar ve sıfır hipotezini reddetmek şeklinde işler. Bu modeller için tüm değişkenlerin ilişkisiz olduğu boş hipotezi test eder (McDonald & Ho, 2002). Bu kategorideki istatistikler: CFI, TLI ve NFI değerleridir.

Ki-kare (χ^2): Ki-kare değeri, genel model uyumunu değerlendirmek için kullanılan geleneksel bir ölçümdür ve örneklem ile uygun kovaryans matrisleri arasındaki tutarsızlığın büyüklüğünü değerlendirir (Hu & Bentler, 1999). Bu test çok değişkenli normalliği varsayar ve normallikten ciddi sapmalar model uygun biçimde belirtilmiş olsa bile modelin

reddedilmesine neden olabilir (McIntosh, 2006). Ki-Kare istatistiksel bir anlamlılık testi olduğu için örneklem büyüklüğüne duyarlı olduğundan dolayı büyük örneklem kullanıldığında ki-kare istatistiği neredeyse her zaman modeli reddeder (Jöreskog & Sörbom, 1993).

Yaklaşık hataların ortalama karekökü (RMSEA): RMSEA bilinmeyen ancak optimum olarak seçilmiş parametre tahminleriyle modelin evrenin kovaryans matrisi ile ne kadar iyi uyacağını bildiren istatistiktir (Byrne, 2011). RMSEA indeksi için 0 ile 0,05 arasındaki değerler modelin mükemmel uyum gösterdiği anlamına gelir (Jöreskog & Sörbom, 1993; Brown, 2015). Doksanların başına kadar, 0,05-0,10 aralığındaki bir RMSEA, uygun bir uyum göstergesi olarak kabul edilirken 0,10'un üzerindeki değerler, zayıf bir uyumu belirtmiştir. Daha sonra 0,08 ile 0,10 arasındaki bir RMSEA'nın ortalama bir uyum sağladığı ve 0,08'in altında daha iyi bir uyum gösterdiği düşünülmüştür (MacCallum, Browne & Sugawara, 1996).

Uyum iyiliği istatistiği (GFI) : GFI, tahmini evren kovaryansı tarafından açıklanan varyans oranını hesaplar (Tabachnick & Fidell, 2013). GFI, değeri 0 ile 1 arasında değişmektedir. Ayrıca, parametre sayısı arttıkça GFI'nin arttığı ve ayrıca büyük örneklemelerde artma eğiliminde olduğu da gözlemlenmiştir (Bollen, 1990; Miles & Shevlin, 1998). Geleneksel olarak, GFI için 0,90'lık birçok amaçlı kesme noktası önerilmiştir, ancak simülasyon çalışmaları, faktör yükleri ve örneklem büyüklükleri düşük olduğunda 0,95'lik daha yüksek bir kesme noktasının daha uygun olduğunu göstermiştir (Miles & Shevlin, 1998).

Ayarlanmış uyum iyiliği istatistiği (AGFI): GFI ile ilgili olarak, daha doygun modellerin uyumu azaltarak, GFI'yi serbestlik derecelerine göre ayarlayan AGFI'dir (Tabachnick & Fidell, 2013). AGFI, örneklem büyüklüğü ile artma eğilimindedir. GFI'de olduğu gibi, AGFI için de değerler 0 ile 1 arasında değişir ve genellikle 0,90 veya daha büyük değerlerin iyi uyum modelleri gösterdiği kabul edilir. Genellikle kovaryans yapı analizlerinde rapor edilirler.

Standartlaştırılmış artık ortalamaların karekökü (SRMR): SRMR, girdi matrisinde gözlemlenen korelasyonlar ile model tarafından tahmin edilen korelasyonlar arasındaki ortalama tutarsızlık olarak ifade edilebilir. Buna göre, bir artık korelasyon matrisinden türetilir (Brown, 2015). Veri seti içerisindeki örnekleme bulunan kovaryans matrisi ve varsayımda bulunulan kovaryans modelinin artık değerlerinin arasındaki farkın karekökü alınarak SRMR değeri bulunur (Kline, 2011). SRMR, 0 ile 1 arasında bir değer alabilir ve 0 mükemmel uyumu gösterir. Yani, SRMR değeri ne kadar küçükse model o kadar iyi uyum sağlar (Brown, 2015). SRMR için değerler, 0,08'e yakın ve altındaki değerler kabul edilebilir uyum sağlar (Hu & Bentler, 1999). Ancak modelde ve büyük örneklem boyutlarına dayalı modellerde çok sayıda parametre olduğunda SRMR'nin daha düşük olacağı unutulmamalıdır.

Normlu uyum indeksi (NFI): NFI, modelin χ^2 değerini boş modelin χ^2 değeriyle karşılaştırarak modeli değerlendirir. Boş/bağımsızlık modeli, ölçülen tüm değişkenlerin ilişkisiz olduğunu belirttiği için en kötü durum senaryosudur. Bu istatistik için değerler 0 ile 1 arasındadır ve Bentler ve Bonnet (1980) 0,90'dan büyük değerleri önererek modelin iyi bir uyum sağladığını belirtmiştir. Daha yeni öneriler, kesme kriterlerinin $NFI \geq 0,95$ olması gerektiğini belirtmektedir (Hu ve Bentler, 1999). Bu indeksin önemli bir dezavantajı, örneklem büyüklüğüne duyarlı olması, 200'den küçük örneklem için uygunluğu olduğundan düşük tahmin etmesidir (Bentler, 1990). Bu nedenle yalnız kullanılması tavsiye edilmez (Kline, 2011).

Tucker-Lewis indeksi (TLI) : Örneklem büyüklüğünün etkisini ortadan kaldırmak amacıyla NNFI (normsuz uyum indeksi) diğer adıyla TLI istatistiği ortaya atılmıştır. TLI istatistiği model karmaşıklıkça düşme eğilimi göstermektedir Bununla birlikte, küçük örneklemelerin kullanıldığı durumlarda, TLI değeri, iyi uyuma işaret eden diğer istatistiklere rağmen zayıf uyumu gösterebilir (Bentler, 1990; Kline, 2011; Tabachnick & Fidell, 2013). TLI eşik değeri olarak literatürde çok farklı görüşler yer almıştır. TLI > 0,80 gibi eşik

değerler yanı sıra $TLI > 0,95$ gibi yüksek eşik değerlere de rastlanmaktadır (Hu & Bentler, 1999, Byrne, 2011).

Karşılaştırmalı uyum indeksi (CFI) : CFI, örneklem büyüklüğü küçük olduğunda bile iyi performans gösteren örneklem büyüklüğünü hesaba katan NFI'nin revize edilmiş bir şeklidir (Tabachnick & Fidell, 2013). NFI gibi, bu istatistik de tüm gizil değişkenlerin ilişkisiz olduğunu varsayar (boş/bağımsızlık modeli) ve örneklem kovaryans matrisini bu boş modelle karşılaştırır. NFI'de olduğu gibi, bu istatistik için değerler 0 ile 1 arasında olup 1'e yakın değerler iyi uyumu gösterir. CFI 0,90'lık bir kesme kriteri başlangıçta geliştirilmiştir, ancak son çalışmalar yanlış tanımlanmış modellerin kabul edilmediğinden emin olmak için 0,90'dan büyük bir değere ihtiyaç olduğunu göstermiştir (Hu & Bentler, 1999). Bu yüzden, CFI 0,95 değeri şu anda iyi uyumun göstergesi olarak kabul edilmektedir (Hu & Bentler, 1999). Tablo 1'de model-veri uyumunun değerlendirilmesinde kullanılan istatistikler ve değer aralıkları verilmiştir.

Tablo 1
Uyum İyiliği Katsayıları

Uyum İyiliği Katsayıları	Kabul Edilebilir Uyum Bölgesi	İyi Uyum Bölgesi
χ^2	$p > 0,05$	$p > 0,05$
χ^2/sd	$2 \leq \chi^2/sd \leq 5$	$0 \leq \chi^2/sd \leq 2$
RMSEA	$0,05 \leq RMSEA \leq 0,10$	$0 \leq RMSEA \leq 0,05$
SRMR	$0,05 \leq SRMR \leq 0,08$	$0 \leq SRMR \leq 0,05$
CFI	$0,90 \leq CFI \leq 0,95$	$0,95 \leq CFI \leq 1,00$
TLI	$0,90 \leq TLI \leq 0,95$	$0,95 \leq TLI \leq 1,00$

Ki-Kare istatistiksel bir anlamlılık testi olması ve örneklem büyüklüğüne duyarlı olması bakımından büyük örneklemeler kullanıldığında neredeyse her zaman modeli reddeder (Jöreskog & Sörbom, 1993). TIMSS ve PISA gibi uluslararası geniş ölçekli uygulamalarda ki-kare istatistiğinin model uyumsuzluğu nedeniyle genel model uyumuna karar vermek için

kullanımı uygun bulunmamış ve bu çalışmada ki-kare istatistiği dikkate alınmamıştır (Rutkowski & Svetina, 2014). Chen (2007) metrik değişmezliğin sağlanması için ΔCFI değerinin 0,010 veya daha düşük, $\Delta RMSEA$ değerinin 0,015 veya daha düşük, $\Delta SRMR$ değerinin ise 0,030 veya daha düşük bulunduğu kabul edilebilir olduğunu belirtmiştir. Skalar değişmezlik ve katı değişmezliğin sağlanması için ise ΔCFI değerindeki 0,010'luk değişime ek olarak $\Delta RMSEA$ ve $\Delta SRMR$ değerlerinde 0,015'lik değişimlerin kabul edilebileceğini ifade etmiştir. Rutkowski ve Svetina (2014) ise TIMSS ve PISA gibi büyük örneklemelerde metrik değişmezliğin sağlanabilmesi için CFI'daki değişimin (ΔCFI) 0,020'den küçük ve RMSEA'daki değişimin ($\Delta RMSEA$) 0,030'dan küçük olması; ölçek (skalar) değişmezliğin sağlanabilmesi için CFI ve RMSEA'daki değişimin - 0,010 ile 0,010 arasında olması gerektiğini ifade etmiştir. Bu çalışmada ölçme değişmezliği kriteri olarak Chen (2007) tarafından önerilen ΔCFI ve $\Delta RMSEA$ kesme değerleri dikkate alınmıştır.

Hizalama Yöntemi

Çoklu grup doğrulayıcı faktör analizi genellikle ölçme değişmezliği analizinde çok sayıda grubu karşılaştırmak için kullanılır ancak çoğunlukla iki grubun karşılaştırılmasında bazı dezavantajları vardır. Herhangi bir ölçümde gruplar arasında ikili karşılaştırmalar parametre ve grup sayısı arttıkça katlanarak artar böylece ölçme değişmezliğini yanlış tespit etme şansı da yükselir (Rutkowski & Svetina, 2014). Birçok gruba çoklu grup doğrulayıcı faktör analizi yaklaşımı, birçok nedenden dolayı pratik olamayacak kadar hantaldır (Muthén & Asparouhov, 2013). Likert tipi ölçeklerin çoklu grup doğrulayıcı faktör analizi bağlamında sürekli olarak modellenmesi, model uyumsuzluğunun gruplar arasındaki ölçme farklılıklarından mı yoksa verilerin kategorik doğasının göz ardı edilmesinden mi kaynaklandığı belirsiz olduğundan, model uyum indekslerinin değerlendirilmesini zorlaştırabilir (Lubke ve Muthén, 2009).

Verilerdeki grup sayısı ve örneklem büyük olduğunda, çoklu-grup doğrulayıcı faktör analizinin ölçme değişmezliğini test etmede dezavantajları ve sınırlılıkları vardır. Bu yöntem ile pek çok olası değişmezlik ihlali meydana gelebilir ve kabul edilebilir model uyumlarının elde edilmesi zor olabilir. Geleneksel çoklu-grup doğrulayıcı faktör analizinde skalar değişmezliğe ulaşmak, özellikle grup boyutu büyük olduğunda genellikle zordur (He vd., 2019; Marsh vd., 2018).

Asparouhov ve Muthén (2014), modifikasyon indeksleri tarafından yönlendirilen çoklu manuel model ayarlamaları gerektiren çoklu grup doğrulayıcı faktör analizine alternatif olarak hizalama yöntemi olarak adlandırılan ve birçok grubun analizine uygun yeni bir yöntem önermiştir. Hizalama yönteminin temel varsayımı, yaklaşık bir ölçme değişmezliği modeli olduğudur.

Hizalama yöntemi, kesin ölçme değişmezliği gerektirmeden yapılandırma modelinden başlayarak ve faktör ortalamalarının ve varyanslarının gruplar arasında değişmesine izin vererek mümkün olduğunca fazla değişmezlik bulma fikrine dayanır. Hizalama yöntemi, kesin ölçme değişmezliğini varsaymaz. Hizalama yönteminin güçlü yanı, yapılandırma modeline dayalı olması ve çok sayıda grup için en uygun modelleri tahmin edebilmesidir. (Asparouhov & Muthén, 2014; Muthén & Asparouhov, 2013, 2014, 2018).

Yöntemin en temel özelliklerinden biri, ülke karşılaştırmaları gibi birçok grup için modelleri uygun bir şekilde tahmin etme yeteneğidir. Çoklu grup doğrulayıcı faktör analizi, skalar modelin zayıf model uyumu ve çok fazla büyük modifikasyon indeksi nedeniyle çok sayıda grup için pratik değildir. Aksine, hizalama yöntemi, yapılandırma modeline dayanır ve esas olarak ölçme değişmezliği analizini otomatikleştirir ve büyük ölçüde basitleştirir. Ayrıca yöntem, her gruptaki her model parametresi için ayrıntılı bir parametre değişmezliği hesabı sağlar (Asparouhov & Muthén, 2014).

Hizalama yöntemini uygulamak için dört kilit nokta vardır. İlk olarak, hizalama yöntemi benzersizliği optimize etmez, çünkü birincil amaç geçerli grup karşılaştırmaları için yansız gizil faktör ortalamalarını tahmin etmektir. İkincisi, konfigürasyon değişmezliği, hizalama optimizasyonunun bir varsayımdır, çünkü prosedürde sadece faktör yükleri ve kesişimler optimize edilir ve bu süreç için yeterli bir konfigürasyon modeli M_0 gereklidir. Üçüncüsü, optimizasyon prosedürü açımlayıcı faktör analizlerinde döndürme yöntemlerine benzer şekilde çalıştığı için, birkaç büyük değişmez olmayan parametrenin varlığı ve yaklaşık olarak değişmeyen birçok parametre, hizalama optimizasyonunun başka bir varsayımdır. Dördüncüsü, hizalama optimizasyon modeli, araştırmacı girdisi gerektiren iki şekilde tanımlanabilir. Referans grubunun faktör ortalaması ve varyansı, sırasıyla 0 ve 1'e sabitlenebilir (sabit hizalama optimizasyon seçeneği) veya faktör ortalaması serbestçe tahmin edilebilir (serbest hizalama optimizasyon seçeneği).

Hizalama yöntemi, değişmez olmayan parametrelerin miktarını azaltmanın mümkün olduğunu varsaymaktadır. Ölçme değişmezliği, hizalama yöntemiyle iki adımda test edilir. İlk adımda, faktör yükleri ve kesişimlerin gruplar arasında serbest olduğu yapılandırma modelinde faktör ortalamalarını 0'a ve tüm gruplardaki faktör varyanslarını 1'e sabitleyerek faktör metriğini ayarlar. Temel model " M_0 " olarak adlandırılan bu yapılandırma modeli, gruplar arası parametre kısıtlamaları içermediği için çoklu grup faktör analizi modelleri arasında en uygun modeldir. İkinci adımda hizalama optimizasyonu yapılır. Bu aşamada, faktör ortalamaları ve varyansları serbest bırakılır. Faktör ortalamaları ve varyansları, ölçme değişmezliğini sağlamayan grup sayısını en aza indiren bir basitlik fonksiyonu ile hesaplanır. Basitlik fonksiyonu (F), her bir ölçme kesişimi ve faktör yük değerlerinin bileşenler olduğu her grup çifti için kayıp fonksiyonundan oluşur (Asparouhov & Muthén, 2014; Muthén & Asparouhov, 2010, 2018).

F fonksiyonu, her bir grup çifti her kesişim ve faktör yükü parametresi bileşeni için, kayıp fonksiyonu (f) aracılığıyla ölçeklenen parametreler arasındaki farkın toplam kayıp

fonksiyonuna eklendiğini ifade eder. Böylece basitlik fonksiyonu, orta ölçüde değişmez olmayan birçok parametre yerine büyük ölçüde değişmez olmayan az sayıda parametre ve yaklaşık olarak değişmez birçok parametre ile optimize edilir. Bu durum, açıklayıcı faktör analizindeki rotasyon fonksiyonlarının orta büyüklükteki yükleri değil, büyük veya küçük yükleri hedeflemesine benzer (Asparouhov & Muthén, 2014; Muthén & Asparouhov, 2104, 2018).

Basitlik fonksiyonunun en aza indirilmesi, birden fazla yerel optimum değerine bağlı olarak hesaplanır. Genellikle yerel optimumlar, genel optimum uyum fonksiyonu değerine yakın olan uyum fonksiyonu değerlerini verir (Asparouhov & Muthén, 2014).

Önerilen son hizalanmış model, M_0 modeliyle aynı uyuma sahiptir; yani, hizalanmış model değişmezlik miktarını en aza indirmeye çalışmasına rağmen, model uyumundan ödün vermez. M_0 ile son hizalanmış model arasındaki ilişki, açıklayıcı faktör analizindeki döndürülmemiş model (sabit sayıda faktöre sahip tüm doğrulayıcı faktör analizi modelleri arasında en iyi uyuma sahip olan) ile yükleme matrisini uyumdan ödün vermeden basitleştiren döndürülmüş model arasındaki ilişkiye paraleldir. Model, döndürülmemiş modelle aynı uyuma sahiptir (Asparouhov & Muthén, 2014).

Hizalama yönteminde sabit ve serbest optimizasyon olmak üzere iki farklı hizalama optimizasyonu kullanılabilir. Sabit optimizasyonda birinci grubun faktör ortalaması ve faktör varyansı sırasıyla 0 ve 1 olarak sabitlenir; serbest optimizasyonda, birinci grubun faktör ortalaması ve varyansı üzerinde herhangi bir kısıtlama yoktur ve bunlar serbestçe tahmin edilir. Maksimum olabilirlik (ML) ve Bayes hizalama tahmini olmak üzere iki tür tahmin edici kullanılabilir. ML asimptotik teoriye dayanırken, Bayes yöntemini kullanmak için deneysel araştırma sonuçlarına dayalı olarak modeldeki parametrelerin dağılımlarının belirtilmesi gerekir (Asparouhov & Muthén, 2014).

Marsh vd. (2018) çalışmasında serbest optimizasyonu kullanarak model tanımlamasının zayıf olduğunu bildirmiş bu nedenle, sabit yaklaşımı kullanarak sonraki analizde referans grubu için faktör ortalaması 0'a en yakın olan grubu seçmek amacıyla elde edilen analiz çıktısını kullanmıştır.

Hizalama analizi yapılırken her ölçme parametresi için en büyük değişmez grup bulunur, burada değişmez grup kümesindeki her grup için o gruptaki ölçme parametresi, değişmez kümedeki tüm gruplarda o parametre için ortalama değerden istatistiksel olarak anlamlı değildir. Değişmez kümede olmayan her grup için parametre, o parametre ortalamasından istatistiksel olarak önemli ölçüde farklıdır. Algoritma çoklu ikili karşılaştırmaya dayanmaktadır; çoklu testte 0,05'ten daha küçük p değerleri kullanılır (Asparouhov & Muthén, 2014).

Algoritmadaki ilk adım, değişmez grupların bir başlangıç kümesini belirlemektir. Her bir grup çifti için ikili karşılaştırmada elde edilen p değeri 0,01'den büyükse iki grup bağlanır. Daha sonra, o parametre için en büyük bağlı küme belirlenir, bu grupların başlangıç kümesidir. Başlangıç kümesi açıklanan prosedür kullanılarak değiştirilecektir. İlk olarak, mevcut değişmez küme kullanılarak ortalama parametre hesaplanır. Daha sonra, her grubun parametre değerini mevcut ortalama ile karşılaştırmak için her grup için bir anlamlılık testi yapılır. p değeri 0,001'in üzerindeyse grup değişmez kümeye eklenir; bu değer altında ise grup değişmez kümeden çıkarılır. Daha sonra, değişmez küme stabilize olana ve değişmez kümeden hiçbir grup eklenmeyene veya çıkarılmayana kadar bu işlemi tekrarlanır (Asparouhov & Muthén, 2014).

Hizalama yönteminin TIMSS, PISA gibi çok kültürlü verilerin kullanıldığı çalışmalarda diğer ölçme değişmezliği yöntemlerine göre başlıca üç avantajı vardır:

- 1) İlk olarak, maksimum grup sayısı hem madde hem de ölçek düzeyinde ölçme değişmezliğinin sağlandığı tek bir analizle belirlenebilir.

- 2) İkinci olarak, madde tabanlı analizde, faktör yükü ve her birinin madde kesişimi tek adımlı bir analizle ülkeleri ikili olarak karşılaştırılabilir. Ayrıca her bir maddenin ölçme değişmezliğine ne kadar katkıda bulunduğu belirlenir.
- 3) Üçüncü olarak, faktör ortalamalarının hangi ülkeler arasında anlamlı bir farklılık gösterip göstermediği belirlenmekte ve böylece ülkeler arası karşılaştırmalar ölçek düzeyinde tespit edilmektedir (Sırgancı, Uyumaz & Yandı, 2020).

Sonuç olarak hizalama yönteminin özellikleri incelendiğinde;

- 1) İlk olarak, prosedür otomatik olduğu için istatistiksel olarak yönlendirilen çok sayıda hipotez testi yapmak için değişiklik indekslerini inceleme durumu söz konusu olmaz.
- 2) İkinci olarak, hizalama optimizasyonundan sonra model veri uyumu, yapılandırma modeli ile aynıdır.
- 3) Üçüncüsü, az sayıda maddeden ve karmaşık bir örneklem çerçevesinde toplanan verilerden oluşan ölçme araçlarına uygulanabilir.
- 4) Ayrıca, yöntem, madde düzeyindeki verilerin yaygın olmayan bir özelliği olan verilerin normal olarak dağılmasını gerektirmez (Asparouhov & Muthén, 2014).

Hizalama Yönteminde Model Uyum Kriterleri

Hizalama yönteminin performansı, varsayım kontrolleri ve madde düzeyinde analizler yoluyla öncelikle önemli ölçüde değişmez olmayan maddelerin sayısı ve bunların dereceleri ile değerlendirilir.

Muthén ve Asparouhov (2014), hizalama yönteminde iyi bir performans için Monte Carlo simülasyonlarından hareketle madde düzeyinde anlamlılık testlerine dayanarak, maddelerin

%25'inden fazlasının değişmez olmaması gerektiğine dair bir genel kural önermişlerdir. Bu öneri, Flake ve McCoach'un (2018) simülasyonlarında elde ettiği sonuçla yani maddelerin %29'undan azının değişmez olmaması gibi iyi bir performans göstermesiyle desteklenmiştir.

Hizalama optimizasyonu, değişmez olmayan maddeler ve değişmez olmayan gruplar hakkında bilgi sağlar. Değişmezlik için, hizalama uyum indeksleri olan uyum fonksiyonu katkı değeri ve R^2 değeri incelenmelidir. Her maddenin basitlik fonksiyonuna katkısı maddenin değişmezlik düzeyini yansıtır. Katkı ne kadar küçükse, maddeler o kadar değişmezdir. Bu değerlerin toplamı optimize edilmiş toplam basitlik fonksiyonu değerine eşittir. Her bir ölçme parametresi için hesaplanan R^2 değeri, gruplar arası faktör ortalaması ve varyansındaki değişimle açıklanan yapılandırma modelindeki gruplar arasındaki parametre değişimini ifade etmektedir. R^2 tüm madde parametre tahminlerindeki değişkenliğin ne kadarının grupların faktör ortalamaları ve varyansları ile açıklanabileceğini ölçen değişmezlik etki büyüklüğü ölçüsüdür. 1'e yakın bir değer tam değişmezliği gösterir, çünkü madde parametrelerindeki R^2 değişkenliği tamamen grup ortalama farklılıklarıyla açıklanırken, 0'a yakın bir değer grup ortalama farklılıklarının madde parametresindeki değişkenliğin hiçbirini açıklamadığını gösterir (Asparouhov & Muthén, 2014). Diğer bir deyişle bir madde için faktör yükü demografik değişkendeki her alt grup için ayrı ayrı kestirilmekte ve bu kestirimler arasında bir varyans oluşmaktadır. İşte R^2 bu varyansın ne kadarının grup ortalama ve varyansından geldiğini ifade etmektedir.

Hizalama yönteminde değişmezlik veya performans derecesini değerlendirmek için kesin kurallar tam anlamıyla oluşturulmamıştır ve bunun biraz daha araştırılması gerekmektedir. Bu nedenle, değişmezlikten kaynaklanan olası sapmaların anlamlı olup olmadığını ölçmek için her bir madde düzeyinde test için etki büyüklükleri aracılığıyla madde farklılıklarının büyüklüğünü incelemenin gerekli olduğu ifade edilmektedir.

Hizalama yönteminin önemli bir avantajı, gruplar arasında madde parametre dağılımlarının normalliği gibi belirli bir dağılım varsayımının gerekli olmamasıdır. Örneğin, grupların bir

alt kümesi büyük değişmezlik gösterebilirken, geri kalan gruplar çok az değişmezlik gösterebilir. Hangi grupların değişmezliğe katkıda bulunduğuna ilişkin bilgiler de hizalama yöntemiyle elde edilebilir (Muthén & Asparouhov, 2013, 2014).

Faktör ortalamalarının ne kadar iyi tahmin edildiğini kontrol etmek için bir Monte Carlo simülasyon çalışması yürütmek, hizalama analizinin faydalı bir şekilde genişletilmesidir. Çalışmalar büyük ölçüde ölçme değişmezliği gösterebilir, yani birçok ölçme parametresi birçok grupta büyük değişmezlik gösterir. O zaman sorun, faktör araçlarının gruplar arasında güvenilir bir karşılaştırmayı sağlamak için yeterince iyi tahmin edilmemesi olabilir (Muthén & Asparouhov, 2014). Monte Carlo simülasyonları, grup faktör ortalamaları ile tahmin edilen hizalama faktör ortalamaları arasındaki korelasyona dayalıdır, gruplar üzerinden hesaplanır ve replikasyonlar üzerinden ortalaması alınır. Hizalama analizi sonuçlarının güvenilirliğini doğrulamak için, faktör ortalamaları sıralamaları arasında en az 0,98'lik bir korelasyon gereklidir (Muthén & Asparouhov, 2013, 2014).

İlgili Araştırmalar

Bu kısımda ölçme değişmezliği ve hizalama metodu ile ilgili alanyazında yapılan araştırmalara yer verilmiştir.

Ölçüoğlu (2015), çalışmasında TIMSS 2011 Türkiye örnekleminde sekizinci sınıf öğrencilerinin ait verileri kullanılarak, matematik başarısını etki eden bazı değişkenleri modellemiş ve modelin coğrafi bölgelere göre ölçme değişmezliğini araştırmıştır. Çalışmada matematik başarısını belirlemede pozitif yönlü duyuşsal özellikler en önemli etkenken, ev ortamı değişkeninin orta düzeyde etkili olduğu bulunmuştur. Okul ortamı değişkeni ile matematik başarısı arasında bir ilişki bulunmadığını belirtmiştir. Modelde yer alan değişkenlerin coğrafi bölgelere göre ölçme değişmezliğini inceleyerek ölçme değişmezliğinin sağlanamadığı sonucuna ulaşmıştır.

Başusta ve Gelbal (2015), PISA 2009 Türkiye örneklemine ait öğrenci anketinde yer alan fen bilgisi ve teknolojileri ile ilgili maddelerdeki faktörleri tespit ederek bu faktörlerin cinsiyete göre ölçme değişmezliğini yapısal eşitlik modellemesine dayalı olarak araştırmışlardır. Sonuç olarak cinsiyete göre ölçme değişmezliğinin sağlandığını, ölçme aracından elde edilen puanlara göre kız ve erkek öğrencilerin belirtilen ölçme modeli çerçevesinde karşılaştırılabileceğini tespit etmişlerdir.

Kıbrıslıoğlu (2015), çalışmasında PISA 2012 uygulamasına katılan Türkiye, Endonezya ve Çin-Şangay örneklemlerine ait matematik öğrenme alt boyutu anketleri ile oluşturulan matematik öğrenme modelinin belirlenen ülkelere ve cinsiyet gruplarına göre ölçme değişmezliğini çoklu grup doğrulayıcı faktör analizi kullanarak incelemiştir. Araştırma sonucunda modelde yer alan ülkeler arasında ölçme değişmezliğinin sağlanmadığı, cinsiyet grupları arasında ise ölçme değişmezliğinin sağlandığı tespit edilmiştir.

Ayvalı (2016), araştırmasında PISA 2012 uygulamasında yer alan matematik okuryazarlığı testinin OECD ülkeleri arasında ve Türkiye'deki cinsiyet ve bölge grupları arasında ölçme değişmezliğini çoklu grup doğrulayıcı faktör analizi ile incelemiştir. Sonuçlar Türkiye'deki

cinsiyet ve bölge gruplarına göre ölçme değişmezliğinin sağlandığını, OECD üyesi ülkelere göre ise ölçme değişmezliğinin sağlanamadığını göstermiştir.

Karakoç Alatl (2016), çalışmasında PISA 2012 araştırmasına katılan Avustralya, Fransa, Şangay-Çin ve Türkiye örneklemelerinin, matematik ve fen okuryazarlık testleri ile okuma becerileri alt testlerinin dil değişkenine göre ölçme değişmezliğini çoklu grup doğrulayıcı faktör analizi tekniğini kullanarak belirlemeyi amaçlamıştır. Ayrıca, çalışmada değişen madde fonksiyonu ve yanlılık da incelenmiştir. Sonuç olarak, PISA 2012 matematik ve fen okuryazarlığı ile okuma becerileri testlerinin belirlenen örneklemeler arasında dil değişkenine göre ölçme değişmezliğini sağlamadığını tespit etmiştir.

Gülleroğlu (2017), PISA 2012 uygulaması Türkiye örnekleminde öğrencilerin matematik okuryazarlığı testine ait duyuşsal özelliklerinin (matematiğe yönelik ilgi, matematik kaygısı, matematiğe yönelik benlik algısı ve matematik öz yeterliği) cinsiyete göre ölçme değişmezliğini çoklu grup doğrulayıcı faktör analizi ile incelemeyi amaçlamıştır. Matematik kaygısı ve matematiğe yönelik ilgi değişkenleri için kız ve erkek öğrencilerin maddelere benzer şekilde yanıt verdiği ve bu gruplardan elde edilen puanların karşılaştırılabilir olduğu; matematik öz yeterliği değişkenine ait ölçme değişmezliği sağlanmadığı için cinsiyete göre karşılaştırma yapmanın uygun olmayacağı sonucuna ulaşılmıştır.

Ertürk ve Erdinç Akan (2018), TIMSS 2015 Türkiye örnekleminde dördüncü sınıf öğrencilerinin matematik başarısında etkili olan matematiği sevmeye, matematiğe olan ilgi ve matematiğe ilişkin özgüven gizil değişkenleri için cinsiyete göre ölçme değişmezliğini çok gruplu doğrulayıcı faktör analizi yöntemini kullanılarak incelemişlerdir. Araştırmada cinsiyete göre ölçme değişmezliğini sadece matematiği sevmeye gizil değişkeninin sağladığı, bu değişken üzerinden yapılacak karşılaştırmaların anlamlı olacağı sonucuna ulaşılmıştır.

İmrol (2019) araştırmasında, PISA 2012 uygulaması Türkiye örneklemine ait matematiğe yönelik motivasyon ve öz inanç yapılarının okul türü, istatistiki bölge birimi ve sosyo-ekonomik düzey gruplarına göre ölçme değişmezliğini incelemiştir. Sonuç olarak

motivasyon yapısının tüm gruplara göre ölçme değışmezliğini sağladığı; öz-inanç yapısının ise sadece istatistiki bölge birimlerine göre ölçme değışmezliğini sağladığı tespit edilmiştir. Öz-inanç yapısının okul türleri ve sosyo-ekonomik düzey ölçüt değışkenlerine göre ölçme değışmezliğini sağlamadığı dolayısıyla bu ölçeğin, okul türleri ve sosyo-ekonomik düzeye ilişkin yapılacak karşılaştırmalarda kullanılmaması önerilmiştir.

Asar (2019), araştırmasında PISA 2015 matematik okuryazarlığı testinin uygulandığı farklı dilleri konuşan 12 ülke arasında ölçme değışmezliğinin sağlanıp sağlanmadığını incelemiştir. Ölçme değışmezliğinin aşamalarını sırası ile uygulamış, modelin yapısal ve metrik değışmezlik koşullarını sağladığını fakat ölçek değışmezliği aşamasını sağlamadığını tespit etmiştir. Sonuç olarak, belirtilen ülkeler arasında ölçme değışmezliğinin sağlanmadığı bulunmuştur.

Güngör (2019) çalışmasında PISA 2015 uygulaması Türkiye örnekleminde öğrenci anketlerinden fen öğreniminde araçsal güdülenme ve fen özyeterliği alt ölçekleri yardımıyla fen motivasyonu ve özyeterliği modeli oluşturmuş, modelin cinsiyetler ve istatistiki bölgeler açısından ölçme değışmezliğini çoklu grup doğrulayıcı faktör analizi ile incelemiştir. Çalışmada, modelin cinsiyete göre sadece şekil değışmezliğini, bölgelere göre ise zayıf değışmezlik aşamasını sağladığını bulunmuştur. fen motivasyonu ve özyeterliği modeli ölçme değışmezliğini sağlayamamıştır. Böylece, cinsiyete göre öğrencilerin ve farklı bölgelerde yaşayan öğrencilerin puanları arasında karşılaştırmalar yapılamayacağı sonucuna ulaşılmıştır.

Tiryaki (2019), araştırmasında PISA 2015 uygulamasına katılan Türkiye ve ABD örneklemini ele alarak fen bilimlerine yönelik öğrenci tutum anketlerinin altı alt testine çoklu grup doğrulayıcı faktör analizi uygulayarak ölçme değışmezliğini, madde tepki kuramına olabilirlik oranı yöntemiyle ise değışen madde fonksiyonuna dayalı olarak incelemiştir. Sonuç olarak, fen öğrenmekten zevk alma ve fen öğreniminde araçsal güdülenme alt anketlerinin ölçme değışmezliğini sağladığı bulunmuştur. Geniş fen konularına ilgi, fene

yönelik öz yeterlik, fen bilimlerine yönelik epistemik inançlar ve bilim aktiviteleri alt anketlerinde ise sadece metrik değişmezliğin sağlandığı tespit edilmiştir.

He, Barrera-Pedemonte ve Buchholz (2019), çalışmalarında 2015 PISA ve TIMSS 2015 (8. sınıf) uygulamasının her ikisine de katılan 29 ülkedeki araçsal motivasyon, bilimden zevk alma ve okula ait olma duygusu ile ilgili öğrenci anket verilerini kullanmışlardır. Bu anketlerin ölçme değişmezlik düzeylerini incelemeyi, farklı ölçekleme yöntemlerinin yakınsamasını ve anketler arasında öğrenci başarısı ile bu yapıların yakınsamasını kontrol etmeyi amaçlamışlardır. PISA'da okula ait olma duygusu ölçeği dışında her iki uygulamadaki üç ölçeğin de metrik değişmezliğe ulaştığını bulmuşlardır. Çalışmalarında karşılaştırmalı çıkarımlar yapmadan önce ölçme değişmezliğini kontrol etmenin önemini vurgulamışlardır.

Öncü (2019), çalışmasında TIMSS 2015 sekizinci sınıf matematik başarı testini kullanarak OECD ülkelerine göre ölçme değişmezliğinin sağlanıp sağlanmadığını araştırmıştır. Ülkelerin sekizinci sınıf matematik başarı testi puanları için test istatistiklerini değerlendirdiğinde tüm ülkelere ait ölçme modellerinin tüm gruplar için ayrı ayrı doğrulandığına karar vermiştir. Ardından tüm gruplar için kovaryans matrislerinin eşitliği testini yaparak ölçme değişmezliğini çoklu grup doğrulayıcı faktör analizi ile test etmiştir. Araştırma sonucunda OECD üyesi ülkeler için ölçme değişmezliğinin sağlanmadığını tespit etmiştir.

Polat (2019), araştırmasında TIMSS 2015 uygulamasına katılan Türkiye, Singapur ve Suudi Arabistan'ın yer aldığı sekizinci sınıf öğrenci anketinin matematik ve fen dersini sevme, matematik ve fen aktivitelerine katılım, matematik ve fen dersinde kendine güven, matematik ve fen dersine verilen değer alt anketleri için ayrı ayrı matematik ve fen duyuşsal özellik modeli oluşturmuş, bu modellerin kültürlere, cinsiyete ve bölgelere göre çoklu grup doğrulayıcı faktör analizi ile ölçme değişmezliğini incelemiştir. Sonuç olarak matematik ve fen duyuşsal özellik modeli, kültürlere ve bölgelere göre ölçek değişmezliğini, cinsiyete göre

katı değişmezliği sağlamıştır. Cinsiyete göre ölçme değişmezliği sağlandığı için cinsiyete göre yapılan karşılaştırmaların anlamlı olduğu sonucuna ulaşmıştır.

Tavlıca (2019), araştırmasında TIMSS 2015 dördüncü sınıf matematik başarı testinin OECD ülkelerine göre ölçme değişmezliğini incelemiştir. Çoklu-grup doğrulayıcı faktör analizi sonucu elde edilen bulgulara göre, modeller karşılaştırıldığında en iyi çalışan modelin güçlü faktöriyel değişmezlik modeli olduğu belirlenmiştir. Sonuç olarak TIMSS 2015 dördüncü sınıf matematik başarı testi puanlarının OECD üyesi ülkeler arasında ölçme değişmezliğini sağlamadığı, bu nedenle söz konusu ülkeler arasında karşılaştırma yapılamayacağı sonucuna ulaşmıştır.

Karasu (2020), çalışmasında TIMSS 2015 Türkiye örnekleminde dördüncü sınıf öğrencilerinin velilerine uygulanan ev anketi verileri ile oluşturulan erken matematik aktiviteleri modelinin bölgelere ve cinsiyete göre ölçme değişmezliğini ölçek ve madde düzeyinde incelemiştir. Erken matematik aktiviteleri ile ilgili model oluşturularak örtük sınıf analizi ile uygun sınıf sayısı belirlenerek modelin çoklu-grup örtük sınıf analizi yöntemi ile bölgelere ve cinsiyete göre ölçme değişmezliği incelenmiştir. İki ayrı analizde de en uygun model olarak üç örtük sınıflı homojen model seçilmiştir. Homojen modelin kabulünden sonra, modelin bölgelere ve cinsiyete göre ölçek ve madde düzeyinde değişmez olduğu sonucuna varılmıştır.

Pedrero ve Manzi (2020), 58 ülkenin yer aldığı çalışmalarında, TIMSS 2011 verilerini kullanarak dördüncü sınıf düzeyindeki öğrencilerin matematik ve fen bilimlerindeki akademik başarıları ile ilişkili öğrenci anketlerinden içsel motivasyon, katılım ve öz inançlar ölçeklerinin ölçme değişmezliğini araştırmışlardır. Matematik ve fen bilimlerindeki içsel motivasyon ölçeklerinin daha yüksek derecede kültürler arası denklik sergilediği, öz inançlar ölçeğinin ise hiçbir derecede değişmezlik göstermediğini bulmuşlardır. Böylece, motivasyon ölçeği ülkeler arasında geçerli karşılaştırmalar yapmak için değişmeyen tek ölçektir. Sonuç olarak, motivasyon ve katılım yapılarının kültürler arası olarak eşdeğer olduğunu; kendine

inanç ölçeğinin ölçme değişmezliğini sağlamadığını tespit etmişlerdir. Madde yanlılığı ile ilgili olarak, sadece motivasyon ölçeğinin metrik değişmezlik gösterdiğini bulmuşlardır.

Eryılmaz, Rivera Gutiérrez ve Sandoval Hernández (2020) çalışmalarında, PISA 2015 uygulamasına katılan ülkeler arasında sosyo-ekonomik statünün ölçme değişmezliğini test etmişlerdir. Ayrıca, PISA 2015'e katılan öğrencilerin sosyoekonomik geçmişlerini ölçmek için tasarlanan anketin, özellikle bölge/kıtaya göre gruplandırıldığında (Latin Amerika, Asya, Avrupa), ülkeler arasında aynı anlamı temsil edip etmediğini göstermeyi amaçlamışlardır. Sonuç olarak, bu yapının ölçme değişmezliğini sağlamadığı, yani tüm ülkelerde kültürler arası karşılaştırılabilirliği desteklemediği belirlenmiştir. Ancak ülkeler daha homojen gruplara ayrıldıktan sonra kültürler arası kısmi karşılaştırılabilirliğin desteklendiği tespit edilmiştir.

Eryılmaz (2021), çalışmasında PISA 2015 uygulamasında öğretmenlere ilk kez uygulanan iş tatminleri ölçeğinin mevcut koşullardaki iş tatmini alt ölçeğini incelemiştir. Araştırmada ölçekte yer alan bazı maddelerin ölçme değişmezliği 18 ülke örnekleminde incelemek için çoklu grup doğrulayıcı faktör analizi uygulanmıştır ve değişmezlik adımları sırası ile incelenmiştir. Sonuç olarak, ölçeğin tüm ülkelere göre yapısal, metrik ve skalar değişmezlik şartlarını sağladığı, dolayısıyla ölçeğin belirtilen ülkelere göre ölçme değişmezliğini sağladığı tespit edilmiştir.

Liou ve Lin (2021) çalışmalarında, TIMSS 2019 uygulamasına katılan Tayvan, Avustralya ve Amerika Birleşik Devletleri'ndeki sekizinci sınıf öğrencilerinin fen motivasyonel inançlarının ülkeler ve cinsiyetlere göre ölçme değişmezliğini çoklu grup doğrulayıcı faktör analizi ve gizil faktör ortalama karşılaştırmaları uygulayarak incelemişlerdir. Sonuç olarak fen motivasyonel inançlarının ölçme değişmezliği üç ülkede sadece metrik değişmezliğe ulaşmış ve bu da gizil bir ortalama karşılaştırmasını mantıksız hale getirmiştir. Her ülkede cinsiyetlere göre ölçme değişmezliği sağlanarak, cinsiyetler arasında karşılaştırılma yapılabileceği sonucuna ulaşılmıştır.

Eryılmaz ve Sandoval Hernández (2021) çalışmalarında OECD'nin Uluslararası Öğretme ve Öğrenme Anketi tarafından sağlanan ve müdürlere uygulanan “temel öğretim liderliği ölçeğinin” 48 ülke için kültürler arası model verilerinin tekdüzeliğini ve ölçme değişmezliğini incelemişlerdir. Daha sonra bu ülkeleri sosyo-demografik özelliklerine göre homojen alt gruplara ayırarak ölçme değişmezliğini yeniden test etmişlerdir. Sonuç olarak, 48 ülke için müdürlerin okul liderliğini ölçen ölçeğin tüm ülkelerde sadece metrik değişmezliği sağladığını bulmuşlardır. Bununla birlikte, alt gruplar için test edildiğinde kültürler arası karşılaştırılabilirlikte gelişmeler bulunduğunu gözlemlemişlerdir.

Uyar (2021), çalışmasında TIMSS 2015 uygulaması Türkiye örnekleminde sekizinci sınıf öğrencilerine uygulanan “matematik tutum anketinin” açımlayıcı yapısal eşitlik ve doğrulayıcı faktör analizi modellerine uygunluğu ve cinsiyete göre ölçme değişmezliğini incelenmiştir. TIMSS 2015 uygulamasında tutum maddeleri öğrencilerin matematiğe ilgisi, öğrencilerin matematik derslerini öğrenmeye yönelik görüşleri, öğrencilerin matematikte özgüvenleri ve öğrencilerin matematiğe değer vermeleri olarak dört başlık altında ele alınmıştır. Maddelerin faktör yapısı incelendiğinde 5 faktörlü ve 35 maddeden oluşan tutum anketinin açımlayıcı yapısal eşitlik modeli'ne (ESEM) uygun olduğu kabul edilmiştir. Matematik tutum anketinin cinsiyete göre değişmezliği incelendiğinde açımlayıcı yapısal eşitlik modeli'ne göre altı aşamada incelenen anketin tüm değişmezlik koşullarını sağladığı tespit edilmiştir.

Byrne ve van de Vijver (2017) araştırmalarında, 30 kültürde aile işleyişine ilişkin bir çalışmada 27 ülkeden 5.482 üniversite öğrencisinin yer aldığı Aile Değerleri Ölçeğine verilen yanıtların ölçme değişmezliğini doğrulayıcı faktör analizine dayalı olarak ve hizalama yöntemini kullanarak araştırmışlardır. Bulgular incelendiğinde hizalama yaklaşımı değişmez olmayan sonuçlar için kabul edilebilir aralıkta bulunmuş, böylece gizil ortalama tahminlerinin güvenilirliğini ve bunların gruplar arasında karşılaştırılmasını doğrulamıştır. Hizalama yaklaşımının, büyük ölçekli karşılaştırmalarda geleneksel doğrulayıcı faktör

analizi prosedürlerinin önemli sınırlamalarının üstesinden gelebilecek bir yöntem olduğu sonucuna ulaşılmıştır.

Lomazzi (2018) 59 ülkenin yer aldığı çalışmasında Dünya Değerler Araştırması'nın (WVS) cinsiyet rolü tutum ölçeğinde ölçme değişmezliğini çoklu grup doğrulayıcı faktör analizi ve hizalama optimizasyonu uygulayarak incelemiştir. Sonuç olarak çoklu grup doğrulayıcı faktör analizi ile 27 ülke için kısmi skalar değişmezlik sağlanmış, hizalama yöntemini kullanarak ise 35 ülke için kabul edilebilir bir değişmezlik derecesi elde edilmiştir.

Sırgancı, Uyumaz ve Yandı (2020), çalışmalarında PISA 2015 fen öğrenmeye yönelik ölçme aracının alt ölçeklerinden biri olan araçsal motivasyon ölçeği verilerini kullanarak hizalama yöntemi ile ölçme değişmezliğini incelemiştir. 57 ülkenin yer aldığı çalışmalarında hizalama yöntemini kullanarak skalar değişmezlik düzeyine ulaşmışlardır. Böylece hizalama yöntemi ile ölçme değişmezliği sonuçları, ölçek puanlarının karşılaştırılabilir olduğunu göstermiştir.

Glassow, Rolfe ve Hansen (2021), çalışmalarında TIMSS 2015 uygulamasında 46 eğitim sisteminde matematik öğretmenlerinin iş doyumu, okulun akademik başarıya verdiği önem, okul durumu ve kaynakları, güvenli ve düzenli okul ve öğretmenin öz-yeterliği alt anketlerinin ölçme değişmezliğini incelemek için çoklu grup doğrulayıcı faktör analizini ve hizalama yöntemini kullanmışlardır. Çoklu grup doğrulayıcı faktör analizi sonuçları okulun akademik başarıya verdiği önem, okul durumu ve kaynakları, öğretmenin öz-yeterliği alt anketleri olmak üzere sadece üç yapının metrik değişmezliğe ulaştığını göstermesine rağmen, hizalama yöntemi uygulandığında sonuçlar beş yapının tümünün kabul edilebilir ölçme değişmezliğini sağladığını göstermiştir.

Immekus (2021), çalışmasında PISA 2015 uygulamasına katılan 35 OECD ülkesinde 15 yaşındaki öğrencilerin yer aldığı her cinsiyet grubu için toplam 70 grup (yani 35 ülke×2 cinsiyet grubu) için okul ile ilgili kaygı ölçeğinin ölçme değişmezliğini incelemiştir. Değişmezlik testi için çoklu grup doğrulayıcı faktör analizi ve hizalama optimizasyon

prosedürlerini kullanmıştır. Hizalama optimizasyonunda, faktör ortalamalarının doğruluğunu araştırmak için de bir Monte Carlo çalışması kullanılmıştır.

Zakariya (2021) çalışmasında 2018 yılında gerçekleştirilen Uluslararası Öğretme ve Öğrenme Araştırması'na (TALIS) katılan 48 ülke ve ekonomide ortalama öğretmen öz-yeterliliğini karşılaştırmak ve sıralamak için bir hizalama optimizasyonu yöntemini uygulamıştır. Araştırma bulgularına göre en yüksek ortalama özyeterliliğe sahip öğretmenlerin Kolombiya, Portekiz, Birleşik Arap Emirlikleri, Macaristan ve Güney Afrika ülkelerinde yer aldığı tespit edilmiştir. Sonuç olarak araştırmada öğretmen öz-yeterliliği anketi ülkelere göre skalar değişmezliği sağlamamıştır.

Odell, Gierl ve Cutumisu (2021) çalışmalarında, PISA 2015 bilgi ve iletişim teknolojisi (BİT) aşinalık anketine katılan 47 ülke için matematik, fen ve BİT ölçeklerinin ölçme değişmezliğini hizalama yöntemini kullanarak araştırmışlardır. Araştırma sonucunda matematik ve fen ölçek puanlarının oldukça değişmez olduğu ve ülkeleri karşılaştırmak için kullanılabileceğini gösterirken, BİT ölçeğinin ölçme değişmezliğini sağlamadığı için tüm katılımcı ülkelerde BİT araçlarını karşılaştırmak için kullanılamayacağı belirlenmiştir.

Qadach, Erdogan, Schwabsky, Schechter ve Tschannen-Moran (2021) araştırmalarında İsrail, Türkiye ve ABD'den Arap, Yahudi, Türk ve Amerikan olmak üzere dört kültürü temsil eden öğretmenlere kolektif öğretmen yeterliği inanç ölçeği uygulanarak belirli parametrelerin değişmez olmadığı grupları belirlemek ve gizil faktör ortalamalarını karşılaştırmak için çoklu grup doğrulayıcı faktör analizi ve hizalama optimizasyonu kullanarak yapılandırma değişmezliğini incelemişlerdir. Araştırma bulguları gizil faktör ortalamalarının ABD için en yüksek ortalama ve Türkiye için en düşük ortalama olduğunu ortaya koymuştur. Bulgular temel model yapısının, dört kültürel ve dilsel grupta aynı olduğunu yani yapılandırma değişmezliğini sağladığını göstermektedir. Hizalama yöntemini kullanan değişmezlik testlerine dayalı olarak kolektif öğretmen yeterliği inanç ölçekleri, dört kültürde tüm faktörlerde değişmez olduğunu (3, 4, 9 ve 12. madde hariç) göstermiştir.

Kolektif öğretmen yeterliği inanç ölçekleri, incelenen dört kültürel ve dilsel grup arasında tam olarak değişmez olmasa da bu anketin dört gruptaki öğretmenler için hemen hemen aynı anlama sahip olduğu sonucuna ulaşmışlardır.

Rolfe (2021) araştırmasında İsveç'e ait 15 yıllık PISA verilerini kullanarak uygulanabilir ve uygun bir sosyoekonomik durum ölçme modeli oluşturmak ve bu modelin özelliklerini ölçme değişmezliği ve hizalama analizi ile test etmek için doğrulayıcı faktör analizini kullanmıştır. Doğrulayıcı faktör analizi ile tüm PISA döngülerinde ölçme değişmezliği testi, metrik düzeyde değişmezlik göstermiştir. Ölçme modeline hizalama optimizasyonu yöntemi uygulandığında, modelin kabul edilebilir ölçme değişmezliğini sağladığı sonucuna ulaşılmıştır.

Ding, Yang Hansen ve Klapp (2022) çalışmalarında PISA 2003 ve 2012 uygulamasına katılan 40 ülkede matematiğe ilişkin benlik kavramı ve öz-yeterliğin ölçme değişmezliğini hizalama metodu kullanarak araştırmışlardır. Hizalama yöntemi uygulandığında, hem matematiğe ilişkin benlik kavramında hem de öz-yeterlikte büyük oranda ölçme değişmezliğinin var olduğunu sonucuna ulaşmışlardır.

Eryilmaz ve Sandoval Hernandez (2023) araştırmalarında 2018 Uluslararası Öğretme ve Öğrenme Araştırması'na (TALIS)'e katılan 40 ülkeden okul müdürleri tarafından algılanan dağıtımçı liderlik ölçeğinin gizil ortalamalarını karşılaştırmak için hizalama optimizasyonu yaklaşımını kullanmışlardır. Araştırma sonucunda Güney Kore, Kolombiya, Şanghay (Çin), ve Litvanyadaki okul müdürlerinin bakış açısından, okul kararlarında en yüksek dağıtımçı liderliğe sahip olduğu; Hollanda, Belçika, Arjantin ve Japonya'nın ise en düşük seviyelere sahip olduğu bulunmuştur.

BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın modeline, çalışma grubuna ve veri toplama araçlarına yer verilmiştir.

Araştırmanın Modeli

Bu araştırma TIMSS 2019 uygulamasına katılan sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeğinin hizalama metoduna göre ölçme değişmezliğinin incelenmesini amaçladığından ilişkisel tarama modelinde bir çalışmadır. İlişkisel tarama modeli iki ve daha çok sayıdaki değişken arasında birlikte değişim varlığını veya derecesini belirlemeyi amaçlar (Karasar, 2012, s. 81). Ölçme değişmezliği analizleri özünde bireylerin ölçek maddelerine verdikleri tepkilerle sahip oldukları demografik özellikler arasındaki olası ilişkileri ortaya çıkarmayı amaçlamaktadır. Bu nedenle ilişkisel tarama araştırması olarak ele alınmasının uygun olduğu düşünülmüştür.

Araştırmanın Çalışma Grubu

IEA kuruluşu tarafından TIMSS 2019 örneklem seçimi için iki aşamalı tabakalı örnekleme yöntemi ile ilk olarak okulun bulunduğu bölge, okul türü, okulda bulunan öğrenci sayısı,

gibi okul ile ilgili çeşitli değişkenlere göre okullar belirlenip daha sonra bu okullardan rastgele şubeler seçilmiştir (Mullis, vd., 2020). Araştırma farklı kıtalarda bulunan farklı dil ve kültür gruplarını temsil edecek biçimde seçilen yedi ülke üzerinden yürütülmüştür. Bu ülkeler Şili, Hong Kong, İsrail, Katar, Rusya, Türkiye ve İngiltere'dir. Bu araştırmanın çalışma grubunu TIMSS 2019 uygulamasına sekizinci sınıf düzeyinde yedi ülkeden katılan 25.834 öğrenci oluşturmaktadır. TIMSS 2019 uygulamasına sekizinci sınıf düzeyinde katılan ülkeler ve öğrenci sayılarına ilişkin dağılım Tablo 2'de verilmiştir.

Tablo 2

Çalışma Grubunu Oluşturan Öğrencilerin Ülkelere Göre Dağılımı

Kod	Ülkeler	N	%
152	Şili	4.048	15,67
344	Hong Kong	3.237	12,53
376	İsrail	3.672	14,21
634	Katar	3.817	14,78
643	Rusya Federasyonu	3.889	15,05
792	Türkiye	4.011	15,53
926	İngiltere	3.160	12,23
	Toplam	25.834	100

Verilerin Toplanması

Bu çalışmada kullanılan veriler, belirlenen ülkelerdeki sekizinci sınıf öğrencilerinin TIMSS 2019 uygulamasındaki öğrenci anketlerinde yer alan matematiğe yönelik güven ölçeğine verdikleri cevaplardan oluşmaktadır. Veriler TIMSS araştırmasının uluslararası internet sitesinden (<https://timssandpirls.bc.edu/timss2019/international-database/>) indirilmiştir.

Matematiğe Yönelik Güven Ölçeği

Bu çalışmada TIMSS 2019 araştırmasında sekizinci sınıf öğrencilerine uygulanan matematiğe yönelik güven ölçeği kullanılmıştır. Ölçekte, öğrencilerin matematiğe olan güvenlerini ölçmek için 9 madde bulunmaktadır (5 olumsuz, 4 olumlu madde). Anket

maddeleri 4'lü likert şeklinde olup “1” kesinlikle katılmıyorum, “2” biraz katılmıyorum, “3” biraz katılıyorum ve “4” kesinlikle katılıyorum şeklinde cevap seçeneklerine sahiptir. Tablo 3’te ölçek maddeleri sunulmuştur.

Tablo 3

Matematiğe Yönelik Güven Ölçeği Maddeleri

Kodlar	Maddeler
BSBM19A	Matematikte genellikle iyiyimdir.
BSBM19B*	Matematik birçok sınıf arkadaşına göre bana daha zor gelir.
BSBM19C*	Matematik güçlü yönlerimden biri değildir.
BSBM19D	Matematikte her şeyi çabuk öğrenirim.
BSBM19E*	Matematik beni sinirlendirir.
BSBM19F	Zor matematik problemlerini çözmede iyiyimdir.
BSBM19G	Öğretmenim bana matematikte iyi olduğumu söyler.
BSBM19H*	Matematik benim için diğer derslerden daha zordur.
BSBM19I*	Matematik kafamı karıştırır.

*Ters kodlanmış maddeler

Veri setine ait güvenirlik kestirimi için likert tipi ölçeklerde kullanılan Cronbach Alfa (α) güvenirlik katsayısı incelenmiş, maddelerin iç tutarlılığının yüksek olduğu görülmüştür. Ölçeğe ait Cronbach α değeri 0,882 bulunmuştur. Cronbach α katsayıları Şili, Hong Kong, İsrail, Katar, Rusya Federasyonu, Türkiye ve İngiltere için sırasıyla 0,878; 0,904; 0,876; 0,824; 0,914; 0,889 ve 0,879 olarak hesaplanmıştır.

Verilerin Analizi

Araştırmada verilerin analizinden önce ilk olarak ölçekte yer alan olumsuz maddeler ters kodlanmıştır. Veri setinde kayıp veriler incelenmiş, tek değişkenli uç değerler, çok değişkenli uç değerler, tek değişkenli normallik, çok değişkenli normallik ve çoklu bağlantılık araştırılmıştır. Daha sonra Brown (2015) ile Van de Schoot, Lugtig ve Hox’un

(2012) önerdiği gibi ölçme değişmezliği çalışmalarına her bir grup için ayrı ayrı DFA yapılarak başlanmıştır.

TIMSS 2019 öğrenci anketinde yer alan matematiğe yönelik güven ile ilgili maddeler için matematiğe yönelik güven modeli doğrulayıcı faktör analizi ile doğrulanmıştır. Ardından doğrulanmış modelin, çoklu grup doğrulayıcı faktör analizi (ÇGDFA) ile ülkelere göre ölçme değişmezliği incelenmiştir. Son olarak hizalama metoduna göre ölçme değişmezliği de araştırılmıştır. Ölçme değişmezliği analizleri için Mplus (Muthén & Muthén, 1998-2017) programı, diğer analizler için ise SPSS ve R (R Core Team, 2022) programındaki MVN paketi (Korkmaz, Goksuluk ve Zararsız, 2014) kullanılmıştır.

Doğrulayıcı Faktör Analizi Varsayımları

Analizlere başlamadan önce faktör analizi varsayımları incelenmiştir. Bu varsayımlar; tek ve çok değişkenli normallik, doğrusallık, çoklu bağlantı, teklik incelenmesidir (Tabachnick & Fidell, 2013).

Kayıp veri: Güven ölçeğinde yer alan maddelere yanıt vermeyen bireylere ait veriler, veri setinden çıkarılmış; kalan veriler üzerinden her ülke için Little's MCAR ($p < 0,05$) kayıp veri analizi yapılarak veri setinin tam seçkisiz olmadığı bulunmuştur. Kayıp veriler madde düzeyinde incelendiğinde kayıp verinin % 0,2 - % 2,3 arasında değiştiği görülmektedir. Kayıp verilere beklenti-maksimizasyon (EM) algoritması kullanılarak değer ataması yapılmış analizlere devam edilmiştir. Mplus, veri kümelerindeki kayıp verileri algılar ve beklenti-maksimizasyon (EM) algoritması aracılığıyla eksik verilerin tam bilgi maksimum olabilirlik tahminlerini sağlar (Muthén ve Muthén, 1998-2017).

Uç değerler: Uç değerleri tespit etmek amacıyla tek değişkenli ve çok değişkenli uç değer analizi yapılmıştır. Tek değişkenli uç değerlerin tespitinde değişkenler için standartlaştırılmış z puanları incelenmeli ve z puanları $\pm 3,29$ aralığında bulunmalıdır

(Tabachnick & Fidell, 2013). Veri seti her ülke için analiz edildiğinde maddelere ait z puanlarının -1.49 ile 2.17 arasında değerler aldığı bulunmuştur. z puanları $\pm 3,29$ aralığında olduğu için tek değişkenli uç değerlerin bulunmadığı görülmüştür. Çok değişkenli uç değerler için Mahalanobis uzaklığı incelenmiş aralık dışında kalan toplam 552 kişi analizden çıkarılarak 25.282 veri ile analize devam edilmiştir.

Normallik: Tek değişkenli normallik için veri setinde çarpıklık ve basıklık değerleri incelenmiştir. Tabachnick ve Fidell (2013) çarpıklık ve basıklık katsayılarının -1.5 ile +1.5 arasında değerler alması ile tek değişkenli normalliğin sağlandığını ifade etmiştir. Çarpıklık ve basıklık katsayıları her ülke için incelendiğinde çarpıklık değerinin -0,403 ila 0,869 arasında, basıklık değerinin ise -1,436 ila 0,375 arasında olduğu bulunmuştur. Çarpıklık ve basıklık katsayıları -1,5 ile +1,5 değer aralığında bulunduğundan tek değişkenli normallik sağlanmıştır. Çok değişkenli normallik incelenmiş ve Mardia'nın çarpıklık değerleri 1104,58-2121,28 arasında ve basıklık değerleri 34,08-89,97 arasında değişmektedir. Mardia'nın çarpıklık ve basıklık katsayıları manidar bulunduğu için veri setlerinin çok değişkenli normalliği sağlamadığı sonucuna ulaşılmıştır.

Değişkenler normal dağılmadığında ve eksik veriler bulunduğu sağlam standart hatalara sahip Robust Maximum Likelihood (MLR) tahminleyicisi kullanmak Maximum Likelihood (ML)'ye göre daha iyi sonuç vermektedir (Brown, 2015). Bu nedenle çalışmada dirençli yöntemlerden MLR tahminleyicisi kullanılarak analiz yapılmıştır.

Çoklu bağlantı: Çoklu bağlantı ve tekillik, değişkenler arası korelasyon çok yüksek düzeyde ilişkili olduğunda ortaya çıkan korelasyon matrisiyle ilgili problemlerdir. Çoklu bağlantı ile, değişkenler çok yüksek düzeyde ilişkilidir (.90 ve üstü); tekillikte değişkenler gereksizdir; değişkenlerden biri, diğer değişkenlerden iki veya daha fazlasının birleşimidir (Tabachnick & Fidell, 2013). Çoklu bağlantı problemi olup olmadığını tespit etmek amacıyla değişkenler arası ikili korelasyonlar, Varyans şişkinlik faktörü (VIF) incelenmiştir. Korelasyon matrisi incelendiğinde en büyük değer 0,783 olduğu bulunmuştur. VIF değerlerinin 1,271 ile

3,721 arasında deęiřtięi elde edilmiřtir. Field (2009) oklu baęlantılık problemi iin VIF deęerlerinin 10'un altında olması gerektięini ifade etmiřtir. Deęerler incelendięinde veriler istenen aralıkta bulunduęundan oklu baęlantılılık problemi olmadıęına karar verilmiřtir.

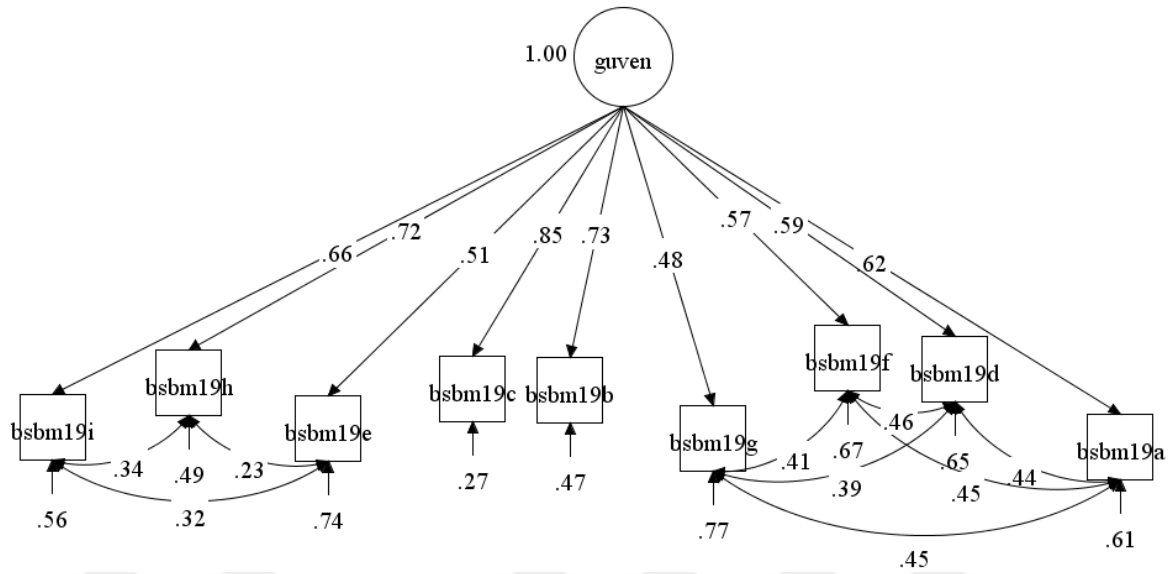
Modelin Oluřturulması

Sekizinci sınıf ğrenci anketlerinden matematięe ynelik gven lęinde yer alan maddeler iin TIMSS uygulamasında kullanılan model baz alınmıřtır. Matematięe ynelik gven modeline her lke iin ayrı ayrı doęrulatoryıcı faktr analizi uygulanmıř daha sonra model tm lkerin yer aldıęı doęrulatoryıcı faktr analizi ile doęrulanmıřtır.

Model'e doęrulatoryıcı faktr analizi uygulandıęında model uyum kriterlerinin saęlanmadıęı bulunmuř nerilen modifikasyonlar sırasıyla uygulanarak model uyumu saęlanmıřtır. Model uyumunun saęlanması iin sırasıyla en byk modifilasyon indeks deęerine sahip maddelerin hata varyansları arasında kovaryans oluřturularak modifikasyon uygulanmıřtır. Bu kovaryansların pozitif maddelerin kendi iinde negatif (ters kodlanmış) maddelerin de kendi aralarında kurulduęu grlmřtir. Pozitif maddelerin hata varyansları arasında 6 adet kovaryans oluřturulmuř, negatif maddelerin hata varyansları arasında ise 3 adet kovaryans oluřturularak model uyumu saęlanmıřtır.

Doęrulatoryıcı Faktr Analizi Sonuları

TIMSS uygulamasının rneklem seiminde iki ařamalı tabakalı rnekleme yntemi kullanıldıęı (Mullis, vd., 2020) iin analizler yapılırken veri yapısı, yani sınıf ve tabakalařma yapısı "TOTWGT" (TOTAL STUDENT WEIGHT) aęırlıklandırılması kullanılarak ele alınmıřtır. Matematięe ynelik gven yapısına iliřkin model řekil 1'de verilmiřtir.



Şekil 1 TIMSS 2019 8. Sınıf matematiğe yönelik güven ölçeğine ait ölçme modeli

Şekil 1’deki matematiğe yönelik güven modeli, tek faktör ve bu faktöre ait dokuz maddeden oluşmaktadır. Matematiğe yönelik güven modelinde yer alan maddelere ait standardize edilmiş faktör yüklerinin 0,48 ile 0,85 arasında değerler aldığı görülmektedir.

Model uyum indeksleri incelendiğinde $\chi^2_{(18)} = 219,780$, $p < 0,05$ olarak bulunmuştur. Ayrıca RMSEA= 0.021 ve CFI= 0,994, TLI= 0,989 SRMR= 0,013 olarak bulunmuştur. Hu ve Bentler (1999) tarafından önerilen değerler model uyumu için esas alındığında tüm ülkelerin yer aldığı doğrulayıcı faktör analizi uyum indekslerinin iyi bir uyum sergilediği (RMSEA < 0,10; SRMR < 0,08; CFI > 0,90; TLI > 0,90) sonucuna ulaşılmıştır.

BÖLÜM IV

BULGULAR VE YORUM

Bu bölümde araştırma sorularına ilişkin bulgular ele alınıp yorumlanmıştır.

Birinci Araştırma Sorusuna İlişkin Bulgular

“TIMSS 2019 uygulaması sonucunda elde edilen veri seti için sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeği puanları, Şili, Hong Kong, İsrail, Katar, Rusya, Türkiye ve İngiltere arasında yapısal, metrik, skalar ve katı değişmezliği sağlamakta mıdır?”

Bu soruya yanıt bulmak için her ülke verisine ayrı ayrı doğrulayıcı faktör analizi uygulanmıştır. Ölçme değişmezliği aşamaları çoklu grup doğrulayıcı faktör analizi yöntemi ile test edilmiştir.

Her bir ülke için ayrı ayrı ölçme modelini değerlendirmek amacıyla doğrulayıcı faktör analizi (DFA) uygulanmış ve sonuçlar tablo 4’te verilmiştir.

Tablo 4

Ülkelere İlişkin Doğrulayıcı Faktör Analizi Uyum İndeksleri

Kod	Ülkeler	χ^2	Sd	CFI	TLI	RMSEA	SRMR
152	Şili	108,169*	18	0,988	0,976	0,035	0,023
344	Hong Kong	160,738*	18	0,984	0,968	0,049	0,026
376	İsrail	72,040*	18	0,994	0,988	0,029	0,014
634	Katar	66,713*	18	0,992	0,984	0,022	0,018
643	Rusya Federasyonu	102,868*	18	0,993	0,986	0,035	0,014
792	Türkiye	82,279*	18	0,994	0,988	0,030	0,016
926	İngiltere	59,167*	18	0,995	0,989	0,027	0,014

* $p < 0,05$

Her ülkeye ait modelin doğrulayıcı faktör analizi sonuçları incelendiğinde ülkelerin modele iyi uyum sağladığı görülmektedir (CFI > 0,90; TLI > 0,90; RMSEA < 0,10; SRMR < 0,08). Ölçme modelinin belirlenen ülkelere göre ölçme değişmezliğini sağlayıp sağlamadığı belirlenen yedi ülkenin yer aldığı çoklu grup doğrulayıcı faktör analizi uygulanarak incelenmiştir. Yapısal değişmezlik, metrik değişmezlik ve skalar değişmezlik modelleri tahmin edilmiş tablo 5'te verilmiştir.

Tablo 5

Çoklu Grup Doğrulayıcı Faktör Analizi Sonucu Ölçme Değişmezliği Uyum İndeksleri

Değişmezlik Düzeyleri	χ^2	Sd	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ RMSEA	Δ SRMR
Model 1 (Yapısal Değişmezlik)	655,288	126	0,991	0,983	0,034	0,018			
Model 2 (Metrik Değişmezlik)	1373,265	174	0,981	0,972	0,043	0,060	0,010	0,009	0,042
Model 3 (Skalar Değişmezlik)	4621,984	222	0,929	0,919	0,073	0,080	0,052	0,030	0,020

* $p < 0,05$

İlk aşama olan yapısal değişmezlik için uyum indeksleri incelendiğinde model uyumunun sağlandığı tespit edilmiştir. (CFI > 0,90; TLI > 0,90; RMSEA < 0,10; SRMR < 0,08). Yani, öğrencilerin matematiğe yönelik güvenini ölçmeyi amaçlayan söz konusu ölçeğin tüm

lkelerde yapısal deęiřmezlięi saęladıęı grlmřtr. Yapısal deęiřmezlięin saęlanması llen gven yapısının yedi lke arasında benzer olduęunu ifade etmektedir. Metrik deęiřmezlik ařaması iin de uyum indeksleri model uyumunun deęerlendirilmesinde kabul edilebilir aralıkta bulunmuřtur ($CFI > 0,90$, $TLI > 0,90$; $RMSEA < 0,10$; $SRMR < 0,08$). ΔCFI ve $\Delta RMSEA$ deęerlerinin de istenen aralıkta olduęu bulunmuřtur ($\Delta RMSEA < 0,015$; $\Delta CFI < 0,010$). Matematięe ynelik gven leęinin metrik deęiřmezlięi saęladıęı bulunmuřtur. Metrik deęiřmezlięin saęlanması, gizil deęiřkene ait lme biriminin yedi lke arasında aynı olduęunu ifade etmektedir. Yani, gizil deęiřkendeki bir birimlik deęiřimin gzlenen deęiřkenlerde lkelere gre eřit miktarda deęiřim gstermesi anlamına gelmektedir. Skalar deęiřmezlik ařamasında ise uyum indeksleri deęerlerinin kabul edilebilir aralıkta bulunduęu fakat ΔCFI ve $\Delta RMSEA$ deęerlerinin kabul edilebilir aralıkta bulunmadıęı sonucuna ulařılmıřtır ($\Delta RMSEA > 0,015$; $\Delta CFI > 0,010$). Skalar deęiřmezlik iin ΔCFI ve $\Delta RMSEA$ deęerlerinin kabul edilebilir aralıkta bulunmadıęı yani leęin skalar deęiřmezlięi saęlamadıęı bulunmuřtur. Bu durum gzlenen deęiřkenlerin ortalama farklarının gizil deęiřkenin ortalamasındaki deęiřimden kaynaklanmadıęını ifade etmektedir. Skalar deęiřmezlik saęlanmadıęı iin skalar deęiřmezlięin hangi maddelerde saęlandıęını belirlemek amacıyla kısmi skalar deęiřmezlik incelemesi yapılmıřtır.

Tablo 6

Çoklu Grup Doğrulayıcı Faktör Analizi Sonucu Kısmi Ölçme Değişmezliği Uyum İndeksleri

MODEL	χ^2	Sd	CFI	RMSEA	SRMR	Δ CFI	Δ RMSEA
Model 2 (Metrik Değişmezlik)	1373,265*	174	0,981	0,043	0,042		
Model 3 (Skalar Değişmezlik)	4621.984*	222	0,929	0,073	0,080		
Model 3a	4094.274*	216	0,937	0,070	0,072	0,044	0,027
Model 3b	3516.282*	210	0,946	0,065	0,066	0,035	0,022
Model 3c	2962.800*	204	0,955	0,061	0,066	0,026	0,018
Model 3d	2549.472*	198	0,962	0,057	0,064	0,019	0,014
Model 3e	2310.181*	192	0,966	0,055	0,063	0,015	0,012
Model 3f	1884.934*	186	0,972	0,050	0,060	0,009	0,007
Model 4 (Kısmi Katı Değişmezlik)	2737.330*	204	0,958	0,059	0,080	0,014	0,009

Not: Model 3a=BSBM19G, Model 3b=BSBM19C, Model 3c=BSBM19D, Model 3d=BSBM19I, Model 3e=BSBM19E, Model 3f=BSBM19H maddelerinin kesişim değerleri ülkeler arasında sırayla bir önceki modele ek olarak serbest tahmin edilmiştir.

* $p < 0,05$

Kısmi skalar değişmezlikte kesişimler için en yüksek modifikasyon değerine sahip olan maddeler sırasıyla bir önceki modele eklenerek serbest bırakılıp analiz yapılmıştır. Buna göre önce BSBM19G maddesi için kesişimler serbest bırakılmış ardından diğer maddelerin kesişim parametreleri serbest bırakılıp uygun kriterler sağlanana kadar analize devam edilmiştir. Kısmi skalar değişmezlik kriterleri model 3f için sağlanmış; yani BSBM19A, BSBM19B ve BSBM19F maddeleri için kısmi skalar değişmezlik sağlanmıştır. Ölçekte yer alan dokuz maddeden sadece üç madde için kısmi skalar değişmezlik sağlanmıştır. Daha sonra model 3f için kısmi katı değişmezlik incelenmiş ve kısmi katı değişmezliğin sağlanmadığı bulunmuştur (Δ CFI > 0,010). Kısmi katı değişmezlik sağlanmadığı için bu üç gözlenen değişkenin hata varyansları yedi ülke arasında değişmez değildir.

TIMSS 2019 uygulaması sonucunda elde edilen veri seti için sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeği puanlarının, belirlenen yedi ülke arasında yapısal ve metrik değişmezliği sağladığı bulunmuştur. Skalar değişmezlik sağlanamadığı için kısmi skalar değişmezlik incelenmiş ve üç madde için kısmi skalar değişmezlik sağlanmıştır.

İkinci Araştırma Sorusuna İlişkin Bulgular

“TIMSS 2019 uygulaması sonucunda elde edilen veri seti için sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeği puanları, hizalama metodu kullanıldığında Şili, Hong Kong, İsrail, Katar, Rusya, Türkiye ve İngiltere arasında ölçme değişmezliğini sağlamakta mıdır?”

Bu aşama için öncelikle yapısal değişmezlik koşulunun sağlanması gerekmektedir. Araştırma 1’de model, yapısal değişmezlik koşulunu sağladığı için ölçme değişmezliği hizalama metodu kullanılarak incelenmiştir.

Matematiğe yönelik güven ölçeği puanlarının ülkeler arası gizil ortalamalarını karşılaştırmak için hizalama metodu kullanılmıştır. Bu amaçla Mplus (Muthén & Muthén, 1998-2017) tarafından sunulan serbest hizalama tahmini prosedürü kullanılmıştır.

Tablo 7’de faktör yüklerine ve kesişimlere göre maddeler ayrı ayrı ele alınarak bu maddeler bazında hangi ülkelerin değişmezliği sağladığı hangi ülkelerin değişmezliği sağlamadığı yer almaktadır. Maddelerin karşısındaki değerler ülke kodları olup parantezin olmadığı ülkeler için değişmezliğin sağlandığı, parantez içindeki ve kalın olarak verilen değerler değişmezliğin sağlanmadığı (değişmez olmayan) ülkeleri temsil etmektedir (Muthén & Asparouhov, 2014).

Tablo 7

Hizalama Yönteminde Matematiğe Yönelik Güven Ölçeği Maddelerinin Faktör Yüklerinin ve Kesişimlerinin Değişmezliğinin Sonuçları

Maddeler	Değişmezliği Sağlayan Ülkeler / (Değişmezliği Sağlamayan Ülkeler)	
	Faktör Yükleri	Kesişimler
BSBM19A	152 344 376 (634) (643) (792) 926	(152) (344) 376 634 643 792 926
BSBM19B	(152) (344) (376) (634) 643 792 926	152 344 (376) 634 643 792 (926)
BSBM19C	152 344 376 634 643 792 926	152 (344) (376) 634 (643) (792) (926)
BSBM19D	152 (344) 376 (634) (643) (792) 926	(152) (344) (376) 634 643 (792) 926
BSBM19E	152 344 376 (634) (643) 792 926	152 344 (376) (634) 643 (792) (926)
BSBM19F	152 (344) 376 (634) 643 792 926	152 344 (376) 634 643 792 (926)
BSBM19G	152 (344) 376 (634) (643) (792) (926)	(152) (344) 376 634 (643) 792 (926)
BSBM19H	152 344 376 634 (643) (792) 926	152 344 (376) (634) 643 (792) 926
BSBM19I	(152) (344) 376 634 (643) 792 926	(152) (344) 376 634 (643) 792 926

Not: 152=Şili, 344=Hong Kong, 376=İsrail, 634=Katar, 643=Rusya Federasyonu, 792=Türkiye, 926=İngiltere.

Kalın ve parantez içinde verilen ülkeler değişmezliğin istatistiksel olarak sağlanmadığı ülkelerdir.

Tablo 7, kalın ve parantez içinde kalan ülkeler arasında faktör yükleri ve kesişimlerin anlamlı olarak farklılaştığı; parantez dışındaki ülkelerde ise faktör yükleri ve kesişimlerin ülkeler arasında istatistiksel olarak farklılaşmadığı şeklinde yorumlanır. Tabloyu açıklamak gerekirse BSBM19C maddesinin faktör yüklerinin ülkeler arasında istatistiksel olarak farklılaşmadığı görülmektedir. Bu maddenin faktör yüklerinin tüm ülkeler için değişmezliği sağladığı ve faktör yüklerinin yedi ülke arasında benzer olduğu ifade edilebilir. BSBM19C maddesinin 344 (Hong Kong), 376 (İsrail), 643 (Rusya), 792 (Türkiye) ve 926 (İngiltere) ülkelerine ait kesişim değerleri tüm ülkelere ait kesişim değerlerinin ortalamasından istatistiksel olarak farklılaştığı; 152 (Şili) ve 634 (Katar) için ise bu farklılaşmanın olmadığı görülmektedir. Yani BSBM19C maddesinde kesişimler beş ülke için istatistiksel olarak

farklılaştığından değişmezlik sadece iki ülke için sağlanmıştır. Buna göre BSBM19C maddesinin kesişim değerlerinin Şili ve Katar arasında benzer olduğu ifade edilebilir.

Faktör yükleri incelendiğinde BSBM19C maddesine ait faktör yüklerinin yedi ülke arasında istatistiksel olarak farklılaşmadığı için en değişmez madde olduğu, BSBM19G maddesine ait faktör yüklerinin sadece iki ülke arasında istatistiksel olarak farklılaşmadığı (en fazla sayıda değişen parametreye sahip) için en az değişmez madde olduğu bulunmuştur. Diğer bir ifadeyle BSBM19C maddesi faktör yüklerinde değişmezliği en fazla sağlayan madde, BSBM19G maddesi ise faktör yüklerinde değişmezliği en az sağlayan maddedir.

Kesişimler incelendiğinde BSBM19A, BSBM19B ve BSBM19F maddelerine ait kesişim değerlerinin her biri iki ülke için istatistiksel olarak farklılaşmakta BSBM19C maddesi ise beş ülke için istatistiksel olarak farklılaşmaktadır. BSBM19A, BSBM19B ve BSBM19F maddeleri kesişim değerleri için değişmezliği en fazla sağlayan maddeler, BSBM19C maddesi ise kesişim değerlerinde değişmezliği en az sağlayan maddedir.

Tablo 8

Hizalama Yönteminde Ölçme Parametrelerine Göre Değişmezliği Sağlamayan Ülke Sayıları

Maddeler	Faktör Yükleri	Kesişimler	Toplam
BSBM19A	3	2	5
BSBM19B	4	2	6
BSBM19C	0	5	5
BSBM19D	4	4	8
BSBM19E	2	4	6
BSBM19F	2	2	4
BSBM19G	5	4	9
BSBM19H	2	3	5
BSBM19I	3	3	6
Toplam	25(%39,68)	29(%46,03)	54(%42,86)

Faktör yükleri ve kesişimlere ait değerler birlikte ele alındığında değişmezliği sağlamayan toplam parametre sayısı BSBM19A maddesi için 5, BSBM19B için 6, BSBM19C için 5, BSBM19D için 8, BSBM19E için 6, BSBM19F için 4, BSBM19G için 9, BSBM19H için 5

ve BSBM19I için 6 ülkedir. BSBM19F maddesi değişmezlik düzeyi en yüksek madde; BSBM19G maddesi ise değişmezlik düzeyi en düşük maddedir.

Tablo 7’deki sonuçlar, her maddenin basitlik fonksiyonuna yaptığı katkılarla artırılabilir. Bu fonksiyon aracılığıyla nicel olarak belirlenen değişmezliğe en çok ve en az hangi maddelerin ve hangi grupların katkıda bulunduğu elde edilmektedir. Sonuçlar, kesişim ve faktör yükleri için ayrı ayrı veya birlikte incelenebilir (Muthén & Asparouhov, 2014).

Tablo 9’da gizil yapı için hizalama modelinin maddelerin uyum fonksiyonu değerleri ve R^2 sonuçları yer almaktadır. Tablo 9, optimize edilmiş uyum fonksiyonu her bir maddenin kesişim ve faktör yükü katkısını göstermektedir. Katkı ne kadar küçük olursa, bir parametre o kadar değişmezdir. Bununla birlikte Mplus, uyum fonksiyonunun tersini maksimize ettiği için tabloda 9’da F yerine -F kullanmaktadır.

Tablo 9

Hizalama Yönteminde Kesişim ve Faktör Yüklerine İlişkin Uyum İstatistikleri

Maddeler	Kesişimler		Faktör Yükleri		Uyum Fonksiyonunun Toplam Katkısı
	Uyum Fonksiyonun Katkısı	R^2	Uyum Fonksiyonun Katkısı	R^2	
BSBM19A	-9,433	0,643	-9,962	0,217	-19,395
BSBM19B	-7,736	0,740	-8,336	0,117	-16,073
BSBM19C	-10,745	0,000	-6,929	0,822	-17,673
BSBM19D	-10,900	0,527	-10,870	0,107	-21,770
BSBM19E	-11,428	0,176	-8,844	0,093	-20,271
BSBM19F	-10,111	0,707	-9,626	0,183	-19,737
BSBM19G	-17,173	0,365	-12,405	0,176	-29,578
BSBM19H	-10,155	0,226	-7,601	0,000	-17,756
BSBM19I	-11,113	0,000	-7,820	0,000	-18,934

Tablo 9 incelendiğinde kesişimler için uyum fonksiyonuna göre değişmezlik düzeyi en yüksek madde BSBM19B (-7,736) maddesi iken BSBM19G (-17,173) maddesi kesişimler için değişmezlik düzeyi en düşük maddedir. Faktör yükleri için uyum fonksiyonuna göre

değişmezlik düzeyi en yüksek madde BSBM19C (-6,929)'dir. BSBM19G (-12,405) faktör yükleri için değişmezlik düzeyi en düşük maddedir. Faktör yüklerinde uyum fonksiyonu, kesişimlere göre nispeten daha düşük değerlere sahip olduğu için faktör yükleri kesişimlere kıyasla daha yüksek bir değişmezlik derecesine işaret etmektedir.

Toplam uyum fonksiyonu incelendiğinde basitlik fonksiyonuna en az katkısı BSBM19B (-16,073) maddesinin en fazla katkısı ise BSBM19G (-29,578) maddesinin sağladığı görülmektedir. Toplam uyum fonksiyonuna göre BSBM19B maddesi değişmezlik düzeyi en yüksek madde, BSBM19G maddesi ise değişmezlik düzeyi en düşük maddedir.

Kesişim değerleri için BSBM19B ve BSBM19F maddelerinin R^2 değerleri sırasıyla 0,740 ve 0,707 olup yüksek bir değişmezlik seviyesine sahiptir. BSBM19C ve BSBM19I maddelerinin kesişimler için R^2 değeri 0,000 bulunmuştur. R^2 değerinin 0,000 bulunması ülkelerin ortalama farklılıklarının madde parametrelerindeki değişkenliği açıklamadığı anlamına gelmektedir.

Faktör yükleri incelendiğinde, en büyük R^2 değeri 0,822 ile BSBM19C maddesi için gözlemlenmiş olup yüksek bir değişmezlik düzeyine sahiptir. En küçük R^2 değeri BSBM19H ve BSBM19I maddelerine ait olup faktör yükü için bu değer 0,000'dır. Diğer bir deyişle BSBM19I maddesinin yedi ülkeye ait faktör yükü değerleri arasındaki değişkenliği hizalama modelindeki gizil değişkenin ortalama ve varyansı tarafından hiç açıklanmamaktadır.

Maddeler uyum fonksiyonuna katkı, R^2 değerleri ve değişmezlik tablosu bakımından genel olarak değerlendirilirse faktör yüklerine göre BSBM19C (Matematik güçlü yönlerimden biri değildir) maddesi yüksek bir değişmezlik düzeyine, BSBM19G (Öğretmenim bana matematikte iyi olduğumu söyler) maddesi ise düşük bir değişmezlik düzeyine sahiptir. Kesişimlere göre BSBM19B (Matematik birçok sınıf arkadaşına göre bana daha zor gelir) ve BSBM19F (Zor matematik problemlerini çözmede iyiyimdir) maddeleri yüksek bir değişmezlik düzeyine, BSBM19G (Öğretmenim bana matematikte iyi olduğumu söyler) maddesi ise düşük bir değişmezlik düzeyine sahiptir.

Tablo 10’da her madde için parametre tahmin değeri verilmekte ayrıca madde düzeyinde bu değerlerin ortalaması en düşük ve en yüksek parametre değerlerine sahip ülkeler yer almaktadır.

Tablo 10

Hizalama Yöntemi Metriğinde Maddelerin Kesişim ve Faktör Yüklerine İlişkin Parametre Tahminleri

Maddeler	Ortalama	Standart Sapma	Minimum		Maksimum	
			Tahmin	Ülke	Tahmin	Ülke
Kesişimler						
BSBM19A	-0,785	0,173	-0,916	İsrail	-0,430	Hong Kong
BSBM19B	-0,780	0,073	-0,864	Katar	-0,662	İngiltere
BSBM19C	-0,783	0,204	-1,056	Türkiye	-0,506	İsrail
BSBM19D	-0,781	0,220	-1,138	Türkiye	-0,442	Hong Kong
BSBM19E	-0,786	0,243	-1,161	İngiltere	-0,488	İsrail
BSBM19F	-0,781	0,199	-1,132	İsrail	-0,615	Rusya
BSBM19G	-0,781	0,634	-1,408	İsrail	0,299	Hong Kong
BSBM19H	-0,782	0,189	-0,949	Şili	-0,469	Türkiye
BSBM19I	-0,789	0,234	-1,095	Rusya	-0,551	Türkiye
Yükler						
BSBM19A	0,995	0,204	0,559	Katar	1,166	Rusya
BSBM19B	0,999	0,096	0,856	Şili	1,138	Katar
BSBM19C	0,998	0,032	0,939	Şili	1,029	Hong Kong
BSBM19D	0,997	0,219	0,618	Katar	1,320	Rusya
BSBM19E	0,999	0,123	0,771	Rusya	1,149	Katar
BSBM19F	0,997	0,183	0,609	Katar	1,130	Şili
BSBM19G	0,992	0,293	0,571	Katar	1,307	Türkiye
BSBM19H	1,000	0,066	0,874	Rusya	1,072	Katar
BSBM19I	1,001	0,075	0,865	Rusya	1,094	Şili

Ortalama değişmezlik indeksi=0,283 Toplam değişmez olmayan parametre=%42,86

Kesişim değerleri için BSBM19A maddesinde İsrail’e ait değer Hong Kong’a ait değere göre ortalamaya daha yakındır. Tablo 7’de BSBM19A maddesinin kesişim değerleri için Hong Kong diğer ülkelerden istatistiksel olarak farklılaştığından dolayı değişmezliği sağlamamıştı. Bu iki durum birbiriyle tutarlılık göstermektedir. Diğer bir ifadeyle kesişim değerlerinde ortalamadan diğerlerine göre daha uzak olan ülkeler değişmezliği sağlamamıştır. Faktör yükleri için BSBM19A maddesinde en büyük değer Rusya (1,166),

en küçük değer ise Katara (0,559) aittir. BSBM19B maddesinin faktör yükleri incelendiğinde en yüksek değere sahip ülkenin Katar (1,138), en düşük değere sahip ülkenin ise Şili (0,856) olduğu görülmektedir. BSBM19C maddesinin faktör yük değerinde en yüksek değere sahip ülkenin Hong Kong (1,029), en düşük değere sahip ülkenin ise Şili (0,939) olduğu görülmektedir. BSBM19G maddesine ait faktör yükünde en yüksek değere sahip ülke Türkiye (1,307), en düşük değere sahip ülkenin ise Katar (0,571) olduğu sonucuna ulaşılmıştır. BSBM19G maddesinde faktör yükleri için Türkiye ve Katar'a ait değerler ortalamadan uzaktır. Tablo 7'de BSBM19G maddesinde Türkiye ve Katar'a ait faktör yükleri istatistiksel olarak farklılaştığı için değişmezliği sağlamamıştır. BSBM19G maddesinin kesişimlerinde en yüksek değere sahip ülkenin Hong Kong (0,299), en düşük değere sahip ülkenin ise İsrail (-1,408) olduğu bulunmuştur.

Matematiğe yönelik güven ölçeği için ortalama değişmezlik indeksi 0,283'tür. Bu değer tüm maddelere ait R^2 değerlerinin ortalamasını ifade etmektedir. Model için değişmezliği sağlamayan grupların yüzdesi faktör yükleri için %39,68; kesişimler için %46,03'tür. Genel olarak model için değişmezliği sağlamayan grupların ortalama yüzdesi ise %42,86'dır. Muthen ve Asparouhov (2014) madde düzeyinde anlamlılık testlerine ($\alpha = 0,001$) dayanarak önemli ölçüde değişmezlik göstermeyen grupların yüzdesini %25 olarak önermiştir. Faktör yükleri incelendiğinde değişmezliği sağlamayan grupların yüzdesi Muthen ve Asparouhov (2014) tarafından önerilen %25 sınırının ve Flake ve McCoach'un (2018) simülasyon çalışmalarında %29 olarak ifade edilen sınırın üzerinde bulunmuştur. Bu kriterler esas alındığında matematiğe yönelik güven ölçeğinin ülkeler arasında faktör yüklerine göre karıştırılamayacağı sonucuna ulaşılmıştır. Kesişimler için değişmezliği sağlamayan grupların yüzdesi %46,03 olarak hesaplanmıştır. Muthen ve Asparouhov (2014) tarafından önerilen %25 sınırından hem de Flake ve McCoach'un (2018) simülasyon çalışmalarında %29 olarak ifade edilen sınırdan yüksektir. Genel olarak maddelerin kesişim değerlerinin ülkeler arasında değişmezliği sağlamadığı görülmüştür. Kesişimler için elde edilen hizalama sonuçları ÇGDFA'daki skalar değişmezlik sonuçları ile tutarlılık göstermektedir. Model için

değişmezliği sağlamayan grupların ortalama yüzdesi %42,86 olup önerilen sınırların üzerinde olduğu görülmektedir. Buna göre hizalama yöntemi ile kestirilen ortalamalara (Tablo 11) güvenilemeyeceği belirtilebilir. Bununla birlikte istatistiksel olarak farklılaşmayan yani değişmezliği sağlayan (Tablo 7’de ülke kodu parantez içinde olmayan) ülkeler madde bazında karşılaştırılabilir.

Sonuç olarak değişmezliği sağlamayan ülkelerin ortalama yüzdesi önerilen sınırların üzerinde bulunmuştur. Bahsedilen %25 kuralına göre bu durum güvenilir olmayan hizalama sonuçları anlamına gelmektedir. Sonuçlara göre matematiğe yönelik güven ölçeği ölçme değişmezliğini sağlamamıştır. Hizalama yönteminde değişmezlik veya performans derecesini değerlendirmek için kesin kurallar tam anlamıyla oluşturulmamıştır. Hizalama yönteminde parametreler değişmezlik indeksleri gibi değildir ve gerçek model parametreleridir. Değişmezlik sağlanmasa bile parametreler yorumlanabilir.

Matematiğe yönelik güven ölçeğine ait hizalanmış faktör ortalamaları tablo 11’de gösterilmiştir. Tabloda ayrıca ülkeler arasındaki %5 düzeyinde anlamlı faktör ortalama farklılıklarına yönelik sonuçlar yer almaktadır. Bu tablo gizil ortalama karşılaştırmasını vermekte ve her ülkenin gizil ortalaması tablolarındaki sıralamalarına göre azalan sırada sunulmaktadır.

Tablo 11

TIMSS 2019 8. Sınıf Matematiğe Yönelik Güven Ölçeğinin Gizil Ortalama Karşılaştırması ve Sıralaması

Ülke Sırası	Ülke Kodu	Faktör Ortalaması	Anlamlı Derecede Daha Küçük Faktör Ortalamasına Sahip Ülkeler
1	344	1,092	152 643 792 634 926 376
2	152	0,983	634 926 376
3	643	0,937	634 926 376
4	792	0,934	926 376
5	634	0,850	926 376
6	926	0,693	376
7	376	0,363	

Tablo 11 incelendiğinde ülkelerin faktör ortalama puanlarının yüksekten düşüğe doğru sıralandığı ve her bir ülkenin ortalama faktör puanından %5 düzeyinde manidar olarak farklılık gösteren ülkelerin hangisi olduğu görülmektedir. Değişmezliği sağlamayan (değişen) parametre sayısının önerilen sınırı (%25, %29) aşması nedeniyle hizalama yöntemi ile elde edilen faktör ortalamalarına güvenilemeyeceği için Tablo 11'de sunulan karşılaştırma sonuçlarının yorumlanmasının doğru olmayacağı belirtilebilir.



BÖLÜM V

TARTIŞMA, SONUÇ VE ÖNERİLER

Bu bölümde öncelikle araştırma sonuçları verilmiş ve daha sonra bu sonuçlar ilgili literatürle karşılaştırılarak tartışılmıştır. Daha sonra, özellikle politika yapıcılar başta olmak üzere gelecek araştırmalara ışık tutması amacıyla birtakım önerilerde bulunulmuştur.

Birinci Araştırma Sorusuna İlişkin Sonuçlar

Bu çalışmada öncelikle TIMSS 2019 sınavına katılan sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeği puanlarının Şili, Hong Kong, İsrail, Katar, Rusya, Türkiye ve İngiltere arasında ölçme değişmezliği çoklu grup doğrulayıcı faktör analizi ile incelenmiştir. Çalışmada matematiğe yönelik güven ölçeği puanlarına ait veriler için öncelikle matematiğe yönelik güven modeli oluşturulmuş gerekli modifikasyonlar yapılarak doğrulayıcı faktör analizi uygulanmış ve model doğrulanmıştır. Çalışmada daha sonra matematiğe yönelik güven ölçeği puanlarına ait verilerin ülkelere göre ölçme değişmezliğinin sağlanıp sağlanmadığını tespit etmek amacıyla çoklu grup doğrulayıcı faktör analizi uygulanmıştır. Analiz sonuçları incelendiğinde matematiğe yönelik güven ölçeğinin ülkelere göre yapısal değişmezlik ve metrik değişmezlik aşamalarını sağladığı; skalar değişmezlik koşulunu ise sağlamadığı tespit edilmiştir. Bu yüzden kısmi skalar

değişmezlik incelenmiş modelde yer alan dokuz maddeden üç madde (BSBM19A, BSBM19B ve BSBM19F maddeleri) için kısmi skalar değişmezlik sağlanmıştır. Daha sonra bu üç maddenin yer aldığı model için kısmi katı değişmezlik incelenmiş ve değişmezliğin sağlanmadığı bulunmuştur.

PISA, TIMSS gibi geniş ölçekli uygulamalarda matematiğe yönelik duyuşsal özellikler ile ilgili ölçme değişmezliği araştırmaları incelendiğinde kültürlere (ülkelere), cinsiyete, coğrafi bölgelere, okul türüne ve sosyo-ekonomik düzeye göre çalışmalar yapılmıştır.

Literatür incelendiğinde TIMSS uygulamalarında matematiğe yönelik duyuşsal özellikler ile ilgili ölçme değişmezliği çalışmalarında farklı sonuçlar bulunduğu görülmektedir.

Uyar (2021), TIMSS 2015 Türkiye örnekleminde sekizinci sınıf öğrencilerine uygulanan matematiğe yönelik tutum anketinin açımlayıcı yapısal eşitlik modelini kullanarak cinsiyete göre ölçme değişmezliğini incelenmiş, anketin cinsiyete göre ölçme değişmezliğini sağladığı sonucuna ulaşmıştır. Polat (2019), TIMSS 2015 Türkiye, Singapur ve Suudi Arabistan'ın sekizinci sınıf öğrenci anketlerini ele alarak matematik ve fen duyuşsal özellik modeli oluşturmuş, bu modellerin kültürlere, cinsiyete ve bölgelere göre ölçme değişmezliğini incelemiştir. Sonuç olarak matematik ve fen duyuşsal özellik modeli, kültürlere ve bölgelere göre ölçek değişmezliğini, cinsiyete göre katı değişmezliği sağlamıştır. Ertürk ve Erdinç Akan (2018), TIMSS 2015 Türkiye örneklemini dördüncü sınıf düzeyinde matematiği sevmeye, matematiğe olan ilgi ve matematiğe ilişkin özgüven değişkenleri için cinsiyete göre ölçme değişmezliğini incelemiştir. Araştırma sonucunda sadece matematiği sevmeye gizil değişkeninin cinsiyete göre ölçme değişmezliğini sağladığı, diğer değişkenlerin cinsiyete göre sadece yapısal değişmezliği sağladığını tespit etmişlerdir. Pedrero ve Manzi (2020), TIMSS 2011 uygulamasında 58 ülkenin yer aldığı çalışmada, dördüncü sınıf öğrencilerinin matematik ve fen bilimlerindeki öğrenci anketlerinden içsel motivasyon, katılım ve öz inançlar ölçeklerinin ölçme değişmezliğini araştırmışlardır. Sonuç olarak, motivasyon ve katılım yapılarının kültürler arasında ölçme değişmezliğini sağladığını; öz inanç ölçeğinin ise ölçme değişmezliğini sağlamadığını belirlemişlerdir.

TIMSS 2015 uygulamasında Uyar (2021) ve Polat (2019) 8. sınıf düzeyinde matematiğe yönelik tutum anketini ele aldıkları çalışmalarında cinsiyete göre ölçme değişmezliğinin sağlandığı sonucuna ulaşmışlar, fakat Ertürk ve Erdinç Akan (2018), 4. sınıf düzeyinde sadece matematiği sevme gizil değişkeninin cinsiyete göre ölçme değişmezliğini sağladığı sonucuna ulaşmıştır.

Matematiğe yönelik güven anketinin TIMSS 2015 uygulaması Türkiye örnekleminde 8. sınıf düzeyinde cinsiyete göre ölçme değişmezliğini sağladığı, 4. sınıf düzeyinde cinsiyete göre ölçme değişmezliğini sağlamadığı sonucuna ulaşılmıştır. Polat'ın (2019) çalışmasında yer alan ülkelere göre ise ölçek değişmezliğinin sağlandığı belirlenmiştir. Pedrero ve Manzi (2020) de TIMSS 2011 uygulamasında 4. sınıf düzeyinde matematiğe yönelik güven anketinin kültürlere göre ölçme değişmezliğini sağlamadığı sonucuna ulaşmıştır. Literatür incelendiğinde önceki yıllarda uygulanan TIMSS araştırmalarında matematiğe yönelik güven anket puanlarının genellikle ölçme değişmezliğini sağlamadığı yönünde sonuçlara ulaşılmıştır. Yürütülen tez çalışmasında da skalar değişmezlik sağlanmamış bu yüzden kısmi skalar değişmezlik incelenmiştir. Kısmi skalar değişmezlik sonucunda ise “Matematikte genellikle iyiyimdir” (BSBM19A), “Matematik birçok sınıf arkadaşına göre bana daha zor gelir” (BSBM19B) ve “Zor matematik problemlerini çözmede iyiyimdir” (BSBM19F) maddeleri için kısmi skalar değişmezliğin sağlandığı bulunmuştur.

TIMSS uygulamasında matematik başarıları ile ilgili ölçme değişmezliği çalışmaları incelendiğinde; TIMSS 2015 araştırmasının OECD ülkelerine göre ölçme değişmezliğinin incelendiği çalışmalarda Öncü (2019), sekizinci sınıf matematik başarı testinin, Tavlıca (2019), ise dördüncü sınıf matematik başarı testinin ölçme değişmezliğini incelemiştir. Her iki araştırma sonucunda da TIMSS 2015 matematik başarı testi puanlarının OECD üyesi ülkeler arasında ölçme değişmezliğini sağlamadığı, bu yüzden bu ülkeler arasında karşılaştırma yapılamayacağı sonucuna ulaşılmıştır.

PISA uygulamasında matematiğe yönelik duyuşsal özelliklere ait ölçme değişmezliği çalışmaları incelendiğinde farklı sonuçlar bulunduğu görülmektedir.

Kıbrıslıoğlu (2015), PISA 2012 uygulamasında Türkiye, Endonezya ve Çin-Şangay örneklemine ait matematik öğrenme alt boyutu anketlerinin ülkelere ve cinsiyete göre ölçme değişmezliğini incelemiş, sonuç olarak ülkeler arasında ölçme değişmezliğinin sağlanmadığını, cinsiyet grupları arasında ise ölçme değişmezliğinin sağlandığını tespit etmiştir. Ayvalli (2016), PISA 2012 uygulamasındaki matematik okuryazarlığı testinin OECD ülkelerine; Türkiye’de cinsiyete ve bölgelere göre ölçme değişmezliğini araştırmış, sonuç olarak Türkiye’de cinsiyet ve bölgelere göre ölçme değişmezliğinin sağlandığı, OECD üyesi ülkelere göre ise ölçme değişmezliğinin sağlanamadığını bulunmuştur. Gülleroğlu (2017), PISA 2012 Türkiye örneğinde öğrencilerin matematik okuryazarlığı testine ait duyuşsal özelliklerinin cinsiyete göre ölçme değişmezliğini incelediği araştırmasında matematik kaygısı ve matematiğe yönelik ilgi değişkenleri için elde edilen puanların cinsiyete göre karşılaştırılabilir olduğunu bulmuştur.

Kıbrıslıoğlu (2015) çalışmasında matematik öğrenme anketlerinin cinsiyete göre ölçme değişmezliğinin sağlandığını, Gülleroğlu (2017) ise sadece matematik öz yeterliği değişkenine ait ölçme değişmezliği sağlanmadığı sonucuna ulaşmıştır. Ayvalli (2016) Türkiye örneğini ele aldığı çalışmasında matematik okuryazarlığının cinsiyete göre ölçme değişmezliğinin sağlandığı sonucuna ulaşmıştır.

Kıbrıslıoğlu (2015) ve Ayvalli (2016) PISA 2012 uygulamasını ele aldıkları çalışmalarında kültürlere göre ölçme değişmezliğinin sağlanmadığını tespit etmişlerdir.

İkinci Araştırma Sorusuna İlişkin Sonuçlar

Bu çalışmada TIMSS 2019 sınavına katılan sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeği puanlarının Şili, Hong Kong, İsrail, Katar, Rusya, Türkiye ve İngiltere arasında ölçme değişmezliği farklı bir yöntem olan hizalama metodu kullanılarak incelenmiştir. Hizalama metoduna göre ölçme değişmezliğinin tespiti için ilk aşama olan yapısal değişmezlik koşulunun sağlanması gereklidir. Model yapısal değişmezlik koşulunu sağladığından dolayı ölçme değişmezliği hizalama metodu kullanılarak araştırılmıştır.

Matematiğe yönelik güven ölçeğinde yer alan maddeler uyum fonksiyonu değerleri, R^2 değerleri ve değişmezlik tablosu bakımından genel olarak değerlendirilirse faktör yüklerine göre BSBM19C (Matematik güçlü yönlerimden biri değildir) maddesi yüksek bir değişmezlik düzeyine, BSBM19G (Öğretmenim bana matematikte iyi olduğumu söyler) maddesi ise düşük bir değişmezlik düzeyine sahiptir. Kesişimlere göre BSBM19B (Matematik birçok sınıf arkadaşına göre bana daha zor gelir) ve BSBM19F (Zor matematik problemlerini çözmede iyiyimdir) maddeleri yüksek bir değişmezlik düzeyine, BSBM19G (Öğretmenim bana matematikte iyi olduğumu söyler) maddesi ise düşük bir değişmezlik düzeyine sahiptir.

Matematiğe yönelik güven anketi için ortalama değişmezlik indeksi 0.283'tür. Matematiğe yönelik güven modelinde değişmez olmayan parametrelerin yüzdesi faktör yükleri için %39,68 kesişimler için %46,03 bulunmuş ve değişmez olmayan parametrelerin ortalaması %42,86 olarak hesaplanmıştır. Bu değer Muthen ve Asparouhov (2014) tarafından önerilen %25 sınırının üzerinde bulunmuştur. Bu durum %25 kuralına göre, güvenilir olmayan hizalama sonuçları anlamına gelmektedir.

Hizalama metodu sonuçlarına göre TIMSS 2019 sınavına katılan sekizinci sınıf öğrencilerinin matematiğe yönelik güven ölçeği puanlarının belirlenen ülkelere göre ölçme değişmezliğini sağlamadığı ifade edilebilir. Hizalama metodu kullanılarak her madde için ayrı ayrı değişmezlik incelenebilir. Bu yöntemde parametreler değişmezlik indeksleri gibi olmadığı için değişmezlik sağlanmasa bile parametreler yorumlanabilir.

Hizalama yöntemi kullanılarak yapılan araştırmalar incelendiğinde yöntemin genel olarak daha üst düzey ölçme değişmezliği seviyelerine ulaştığı söylenebilir (Ding, Yang Hansen & Klapp, 2022; Glassow, vd., 2021; Sırgancı, vd., 2020).

Immekus (2021), çalışmasında PISA 2015 uygulamasına katılan 35 OECD ülkesinde 15 yaşındaki öğrencilerin yer aldığı her cinsiyet grubu için toplam 70 grup için okul ile ilgili kaygı ölçeğinin ölçme değişmezliğini hizalama yöntemini kullanarak incelemiştir. Sonuç olarak değişmezliğin tam olarak sağlanamadığı bulunmuştur. Yürütülen tez çalışmasında da

benzer olarak hizalama metodu kullanıldığında daha üst düzey ölçme değişmezliğine ulaşılamamıştır. Fakat güven ölçeği maddeleri ele alındığında değişmezlik düzeyi en yüksek madde faktör yüklerine göre “Matematik güçlü yönlerimden biri değildir” (BSBM19C) maddesi, kesişimlere göre “Matematik birçok sınıf arkadaşına göre bana daha zor gelir” (BSBM19B) maddesidir. Hem faktör yükleri hem kesişimlere göre değişmezlik düzeyi en düşük madde ise “Öğretmenim bana matematikte iyi olduğumu söyler” (BSBM19G) maddesidir.

Geniş ölçekli uygulamalar ele alınarak yapılan ülkeler arası karşılaştırmalarda ölçme değişmezliğinin sağlanmaması bu uygulamalardaki ölçme araçlarının çevirilerinin yol açtığı bir sınırlılık olabilir.

Çoklu grup doğrulayıcı faktör analizi ile kısmi skalar değişmezlik BSBM19A, BSBM19B ve BSBM19F maddeleri için sağlanmıştır. Hizalama metodunda da benzer olarak BSBM19A, BSBM19B ve BSBM19F maddeleri kesişim değerleri için değişmezlik düzeyi en yüksek maddelerdir. Bu sonuca göre iki yöntemin tutarlılık gösterdiği görülmektedir.

Öneriler

Araştırmacılara Yönelik Öneriler

- Araştırma yedi ülke verisi kullanılarak yapılmış olup gelecek araştırmalarda tüm ülkelere ait veri seti ile ölçme değişmezliği incelenebilir.
- Bu çalışma TIMSS 2019 sekizinci sınıf düzeyinde öğrenci anketlerinden matematiğe yönelik güven ölçek verileri üzerinden yürütülmüştür. Gelecek çalışmalarda bu araştırmada yer almayan diğer öğrenci anketleri kullanılarak ölçme değişmezliği araştırmaları yapılabilir.
- Bu araştırma TIMSS 2019 sekizinci sınıf düzeyinde uygulanmış olup yapılacak olan araştırmalarda dördüncü sınıf düzeyindeki veriler de kullanılarak ölçme değişmezliği incelenebilir.
- TIMSS sınavında yer alan başarı testlerine ilişkin ölçme değişmezliği çalışmaları yapılarak ülkeler arasındaki karşılaştırmaların geçerliği artırılabilir.
- Bu çalışmada ölçme değişmezliği çoklu-grup doğrulayıcı faktör analizi ve hizalama metodu ile test edilmiştir. Yapılacak çalışmalarda örtük sınıf analizi gibi farklı yöntemler ile ölçme değişmezliği test edilerek hangi yöntemin en etkili değişmezlik düzeyine ulaştığı tartışılabilir.
- Çok fazla grubun yer aldığı çalışmalarda çoklu-grup doğrulayıcı faktör analizi veri setinde uyumsuzluğa yol açmaktadır. Bu yüzden yapılacak ölçme değişmezliği çalışmalarında çoklu-grup doğrulayıcı faktör analizinin sınırlılıklarını gidermek amacıyla hizalama yöntemi uygulanabilir.
- Ölçme değişmezliği çalışmaları yapılırken çoklu grup doğrulayıcı faktör analizi ve hizalama yöntemi birlikte ele alınarak daha kapsamlı bir sonuç elde edilebilir.

Uygulayıcılara Yönelik Öneriler

- Bu araştırmada TIMSS 2019 8. sınıf öğrencilerine uygulanan matematiğe yönelik güven ölçeğinin skalar değişmezliği sağlamadığı görülmüştür. Bu yüzden uygulayıcıların anketlere veya başarı testlerine yönelik ölçme değişmezliği çalışmaları yapıldıktan sonra bu sonuçlara dayalı olarak ülkeleri karşılaştırmaları daha geçerli sonuçlar verecektir.
- Bu çalışmada TIMSS 2019 öğrenci anketi kullanılmış olup diğer çalışmalarda, yönetici, öğretmen ve veli görüşlerine dayalı anketlere ait ölçme değişmezliği çalışmaları uygulanabilir. Uygulayıcılar (politika yapıcılar) eğitimin farklı parametrelerini de sürece dahil ederek daha geniş perspektifte veri elde edilebilmek için çeşitli anket çalışmaları yürütebilirler. Bu uygulanan ölçeklerin de ölçme değişmezliği araştırmaları yapılarak daha sağlıklı ve çözüm odaklı sonuçlara ulaşılması sağlanabilir.

KAYNAKLAR

- Akyüz, G. (2014). TIMSS 2011’de öğrenci ve okul faktörlerinin matematik başarısına etkisi. *Eğitim ve Bilim*, 39(172).
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME) (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Arıkan, S., van de Vijver, F. J. R. & Yağmur, K. (2016). Factors contributing to mathematics achievement differences of Turkish and Australian students in TIMSS 2007 and 2011. *EURASIA Journal of Mathematics, Science and Technology Education*, 12(8), 2039-2059.
- Asar, E. (2019). *PISA 2015 Matematik okuryazarlığı testinin farklı dilleri konuşan ülkeler arasında ölçme değişmezliğinin incelenmesi*. Yüksek Lisans Tezi, Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü, Antalya.
- Asparouhov, T. & Muthén, B. O. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling*, 21, 1–14. doi:10.1080/10705511.2014.919210
- Ayvallı, M. (2016). *PISA 2012 matematik okuryazarlığı testinin ölçme değişmezliğinin incelenmesi*. Yüksek Lisans Tezi, Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü, Antalya.

- Baker, D. P. & LeTendre, G. K. (2005). *National differences, global similarities – World culture and the future of schooling*. Stanford, CA: Stanford University.
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality And Individual Differences*, 42(5), 815-824
- Başusta, N. B. & Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(4), 80-90.
- Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238–246.
- Bloom, B. S. (2016). *İnsan nitelikleri ve okulda öğrenme* (Üçüncü baskı). (D. A. Özçelik, Çev.). Ankara: Pegem. (1976)
- Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin*, 107(2), 256.
- Bong, M. & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really?. *Educational psychology review*, 15(1), 1-40.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. (Second Edition). New York: Guilford, Inc.
- Büyüköztürk, Ş. (2012). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* (17. Baskı). Pegem Akademi: Ankara.
- Byrne, B. M. & van de Vijver, F. J. R. (2017). The maximum likelihood alignment approach to testing for approximate measurement invariance: A paradigmatic cross-cultural application. *Psicothema*.
- Byrne, B. M. & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal Of Cross-Cultural Psychology*, 34(2), 155-175.

- Byrne, B. M. (2011). *Structural equation modeling with AMOS Basic concepts, applications, and programming (Multivariate Applications Series)*. Routledge, New York.
- Byrne, B. M. (2015). Testing across nations and cultures: issues and complexities. *International Journal of Business and Globalisation*, 14(2), 170-186.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456.
- Chen, F. F. (2007). Sensitivity of Goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464– 504. doi:10.1080/10705510701301834
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth: Holt, Rinehart and Winston. New York: Harcourt Brace Jovanovich College.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin* 52, 281-302.
- Çakan, M. (2003). Geniş ölçekli başarı testlerinin eğitimindeki yeri ve önemi. *Eğitim ve Bilim*, 28(128), 19-26.
- Çiftçi, S. K. & Yıldız, P. (2019). The effect of self-confidence on mathematics achievement: The Metaanalysis of Trends in International Mathematics and Science Study (TIMSS). *International Journal of Instruction*, 12(2), 683-694.
- Demir, İ. & Kılıç, S. (2010). Using PISA 2003, examining the factors affecting students' mathematics achievement. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 38(38), 44-54.
- Ding, Y., Yang Hansen, K., & Klapp, A. (2022). Testing measurement invariance of mathematics self-concept and self-efficacy in PISA using MGCFA and the alignment method. *European Journal of Psychology of Education*, 1-24.

- Doğan, N. & Barış, F. (2010). Tutum, değer ve özyeterlik değişkenlerinin TIMSS-1999 ve TIMSS-2007 sınavlarında öğrencilerin matematik başarılarını yordama düzeyleri. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(1), 44-50.
- Ertürk, Z. & Erdiñç Akan, O. (2018). TIMSS 2015 matematik başarısını etkileyen değişkenlerin yapısal eşitlik modeli ile incelenmesi. *Ulusal Eğitim Akademisi Dergisi*, 2(2), 14-34.
- Ertürk, Z. & Erdiñç Akan, O. (2018). TIMSS 2015 matematik başarısı ile ilgili bazı değişkenlerin cinsiyete göre ölçme değişmezliğinin incelenmesi. *Kuramsal Eğitimbilim Dergisi*, UBEK-2018, 204-226, Aralık 2018.
- Eryılmaz, E. (2021). *Öğretmenlere yönelik iş tatmini ölçeğinin ülkeler arası ölçme değişmezliğinin incelenmesi (PISA 2015 örneği)*. Yüksek Lisans Tezi, Bolu Abant İzzet Baysal Üniversitesi Lisansüstü Eğitim Enstitüsü, Bolu.
- Eryılmaz, N. & Sandoval Hernández, A. (2021). Improving cross-cultural comparability: does school leadership mean the same in different countries?. *Educational Studies*, 1-22.
- Eryılmaz, N., Rivera Gutiérrez, M. & Sandoval Hernández, A. (2020). Should different countries participating in PISA interpret socioeconomic background in the same way?: a measurement invariance approach. *Revista iberoamericana de educación*.
- Eryılmaz, N. & Sandoval-Hernandez, A. (2023). Is Distributed Leadership Universal? A Cross-Cultural, Comparative Approach across 40 Countries: An Alignment Optimisation Approach. *Education Sciences*, 13(2), 218. <https://doi.org/10.3390/educsci13020218>
- Ferla, J., Valcke, M. & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and individual differences*, 19(4), 499-505.
- Field, A. (2009). *Discovering statistics using SPSS* (Third Edition). Sage: London.

- Fishbein, B., Foy, P. & Yin, L. (2021). *TIMSS 2019 User Guide for the International Database*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/international-database/>
- Flake, J. K. & McCoach, D. B. (2018). An investigation of the alignment method with polytomous indicators under conditions of partial measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(1), 56-70.
- Galbraith, P. & Haines, C. (1998). Disentangling the nexus: Attitudes to mathematics and technology in a computer learning environment. *Educational Studies in Mathematics*, 36(3), 275–290.
- Glassow, L. N., Rolfe, V. & Hansen, K. Y. (2021). Assessing the comparability of teacher-related constructs in TIMSS 2015 across 46 education systems: an alignment optimization approach. *Educational Assessment, Evaluation and Accountability*, 33, 105– 137.
- Gülleroğlu, H. D. (2017). PISA 2012 matematik uygulamasına katılan Türk öğrencilerin duyuşsal özelliklerinin cinsiyete göre ölçme değışmezliğinin incelenmesi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 37(1), 151-175.
- Güngör, M. (2019). *Fen motivasyonu ve özyeterliği modeli'nin ölçme değışmezliğinin incelenmesi: PISA 2015 Türkiye örneğı*. Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Hannula, M. S., Maijala, H. & Pehkonen, E. (2004). Development of understanding and self-confidence in mathematics; grades 5-8. *International Group for the Psychology of Mathematics Education*, 3, 17-24.
- He, J., Barrera-Pedemonte, F. & Buchholz, J. (2019). Cross-cultural comparability of noncognitive constructs in TIMSS and PISA. *Assessment in Education: Principles, Policy & Practice*, 26(4), 369-385.

- 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.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <https://doi.org/10.1080/10705519909540118>.
- Immekus, J. C. (2021). Multigroup CFA and alignment approaches for testing measurement invariance and factor score estimation: Illustration with the schoolwork-related anxiety survey across countries and gender. *Methodology*, 17(1), 22-38.
- İmrol, F. (2017). *PISA 2012 Türkiye örnekleminde matematiğe yönelik motivasyon ve öz-inanç yapılarının ölçme değişmezliğinin incelenmesi*. Yüksek Lisans Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Jöreskog, K. G. & Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Scientific Software International. Mooresville, IL.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological bulletin*, 112(3), 527.
- Karakoç Alathı, B. (2016). *Uluslararası öğrenci değerlendirme programı (PISA - 2012) okuryazarlık testlerinin ölçme değişmezliğinin incelenmesi*. Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Karasar, N. (2012). *Bilimsel araştırma yöntemi* (23. baskı). Ankara: Nobel.
- Karasu, S. (2020). *TIMSS 2015 Türkiye örnekleminde erken matematik aktiviteleri modelinin bölgelere ve cinsiyete göre incelenmesi*. Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Kıbrıslıoğlu, N. (2015). *PISA 2012 matematik öğrenme modelinin kültürlere ve cinsiyete göre ölçme değişmezliğinin incelenmesi: Türkiye-Çin, (Şangay)-Endonezya örneği*. Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

- Kline, P. (2000). *The handbook of psychological testing* (Second Edition). London: Taylor & Francis Group.
- Kline, R. B. (2011). *Principles and Practices of Structural Equation Modelling*. 3rd ed. New York: The Guilford.
- Koç, O. (2019). *4. ve 8. sınıf öğrencilerinin TIMSS 2015 matematik başarısını yordayan değişkenlerin belirlenmesi*. Yüksek Lisans Tezi, Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü, Antalya.
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2):151-162.
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and individual differences*, 19(3), 355-365.
- Liou, P. Y. & Lin, J. J. (2021). Comparisons of science motivational beliefs of adolescents in Taiwan, Australia, and the United States: Assessing the measurement invariance across countries and genders. *Frontiers in Psychology*, 12.
- Lomazzi, V. (2018). Using alignment optimization to test the measurement invariance of gender role attitudes in 59 countries. *Methods, data, analyses : a journal for quantitative methods and survey methodology (mda)*, 12(1), 77-103.
- Lubke, G. H., & Muthén, B. O. (2009). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling*, 11, 514–534. doi:10.1207/s15328007sem1104
- MacCallum, R. C., Browne, M. W. & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149.

- Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B. & Dicke, T. (2018). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, 23(3), 524.
- Marsh, H.W. & Craven, R.G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1(2), 133–163.
- Martin, M. O., von Davier, M. & Mullis, I. V. S. (Eds.). (2020). Methods and Procedures: TIMSS 2019 Technical Report. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/methods>
- McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64.
- McIntosh, C. (2006), Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007). *Personality and Individual Differences*, 42(5), 859-67.
- MEB. (2020). *TIMSS 2019 Türkiye ön raporu*. Eğitim Analiz ve Değerlendirme Raporları Serisi No:15. Ankara: MEB.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.
- Milfont, T.L. & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3, 111-130.

- Millsap, R. E. & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis In: Hoyle RH (Ed), *Handbook of Structural Equation Modelling*, 380–392. New York, NY: Guilford.
- Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. New York, Routledge.
- Morony, S., Kleitman, S., Lee, Y. P. & Stankov, L. (2013). Predicting achievement: Confidence vs self-efficacy, anxiety, and self-concept in Confucian and European countries. *International Journal of Educational Research*, 58, 79-96.
- Mullis, I. V. S. & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/frameworks/>
- Mullis, I. V. S., Martin, M. O., Foy, P. & Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/pirls2016/international-results/>
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/international-results/>
- Muthén, B. & Asparouhov, T. (2013). New methods for the study of measurement invariance with many groups. *Mplus Web Notes*, (18).
- Muthén, B. & Asparouhov, T. (2014). IRT studies of many groups: The alignment method. *Frontiers in Psychology*, 5, 978. doi:10.3389/fpsyg.2014.00978
- Muthén, B. & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociological Methods & Research*, 47(4), 637-664.

- Muthén, L. K. & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Odell, B., Gierl, M. & Cutumisu, M. (2021). Testing measurement invariance of PISA 2015 mathematics, science, and ICT scales using the alignment method. *Studies in Educational Evaluation*, 68, 100965.
- OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD.
<http://dx.doi.org/10.1787/9789264190511-en>
- OECD (2019). *PISA 2018 assessment and analytical framework*. Paris: OECD. doi: <https://doi.org/10.1787/b25efab8-en>
- Oliveri, M. E. & von Davier, M. (2011). Investigation of model fit and score scale comparability in international assessments. *Psychological Test and Assessment Modeling*, 53(3), 315.
- Oral, I. & McGivney, E. (2013). Student performance in math and science in Turkey and determinants of success. *Education Reform Initiative, İstanbul*.
- Ölçüoğlu, R. (2015). *TIMSS 2011 Türkiye sekizinci sınıf matematik başarısını etkileyen değişkenlerin bölgelere göre incelenmesi*. Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Öncü, Ö. (2019). *TIMSS 2015 Sekizinci sınıf matematik başarı testinin OECD ülkelerine göre ölçme değişmezliğinin incelenmesi*. Yüksek Lisans Tezi, Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü, Antalya.
- Pedrero, V. & Manzi, J. (2020). Self-beliefs, engagement and motivation in science and mathematics: Are they universal?. *International Journal of Educational Research*, 101, 101562.

- Polat, M. (2019). *TIMSS-2015 Matematik ve fen duyuşsal özellik modellerinin kültürlere, cinsiyete ve bölgelere göre ölçme değişmezliğinin incelenmesi*. Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- 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.
- Qadach, M., Erdogan, U., Schwabsky, N., Schechter, C. & Tschannen-Moran, M. (2021). Collective teacher efficacy beliefs: testing measurement invariance using alignment optimization among four cultures. *Journal of Educational Administration*.
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rolfe, V. (2021). Tailoring a measurement model of socioeconomic status: Applying the alignment optimization method to 15 years of PISA. *International Journal of Educational Research*, 106, 101723
- Rutkowski, L. & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and psychological measurement*, 74(1), 31-57. doi:10.1177/ 0013164413498257
- Sari, M. H., Arıkan, S. & Yıldızlı, H. (2017). Factors predicting mathematics achievement of 8th graders in TIMSS 2015. *Journal of Measurement and Evaluation in Education and Psychology*, 8(3), 246-265.
- Senemoğlu, N. (2013). *Gelişim, öğrenme ve öğretim kuramdan uygulamaya*. (23. Baskı). Ankara: Yargı.
- Shevlin, M. & Miles, J. N. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual differences*, 25(1), 85-90.

- Sırgancı, G., Uyumaz, G. & Yandı, A. (2020). Measurement invariance testing with alignment method: Many groups comparison. *International Journal of Assessment Tools in Education*, 7(4), 657–673.
- Stankov, L. & Lee, J. (2014). Quest for the best non-cognitive predictor of academic achievement. *Educational Psychology*, 34(1), 1-8.
- Stankov, L. & Lee, J. (2017). Self-beliefs: Strong correlates of mathematics achievement and intelligence. *Intelligence*, 61, 11-16.
- Stankov, L. (2013). Noncognitive predictors of intelligence and academic achievement: An important role of confidence. *Personality and Individual Differences*, 55(7), 727-732.
- Stankov, L., Morony, S. & Lee, Y. P. (2014). Confidence: The best non-cognitive predictor of academic achievement? *Educational Psychology*, 34(1), 9-28.
- Steenkamp, J. B. E. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of consumer research*, 25(1), 78-90.
- Suhr, D. D. (2006). Exploratory or confirmatory factor analysis?. 1-17. In Cary: SAS Institute.
- Tabachnick, B. G. & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Allyn and Bacon.
- Tavlıca, A. (2019). *TIMSS 2015 Dördüncü sınıf matematik testinin ölçme değişmezliğinin ülkelere göre incelemesi*. Yüksek Lisans Tezi, Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü, Antalya.
- Tekin, H. (2010). *Eğitimde ölçme ve değerlendirme* (20. Baskı). Ankara: Yargı.
- Tiryaki, F. (2019). *PISA 2015 öğrenci tutum anketlerinin değişen madde fonksiyonu ve ölçme değişmezliğinin incelenmesi*. Yüksek Lisans Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü. Ankara.

- Van de Schoot, R., Lugtig, P. & Hox, J. (2012) A checklist for testing measurement invariance, *European Journal of Developmental Psychology*, 9(4), 486-492.
- Van de Vijver, F. & Leung, K. (1997). Methods and data analysis of comparative research. *Handbook of cross-cultural psychology*, 1, 257-300.
- Vandenberg, R. J. & Lance, C. E. (2000). A Review and synthesis of the measurement invariance literature: Suggestions practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.
- Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, 5(2), 139-158.
- Wilkins, J. L. M. & Ma, X. (2003). Modeling change in student attitude toward and beliefs about mathematics. *The Journal of Educational Research*, 97(1), 52-63.
- Yoon, M. & Kim, E. S. (2014). A comparison of sequential and nonsequential specification searches in testing factorial invariance. *Behavior Research Methods*, 46, 1199-1206.
- Zakariya, Y. F. (2021). Application of an innovative alignment optimisation method to a cross-cultural mean comparison of teacher self-efficacy: A cross-country study. *Heliyon*, 7(10), e08212.



GAZİLİ OLMAK AYRICALIKTIR..