



**T.C.  
HARRAN ÜNİVERSİTESİ  
LİSANSÜNTÜ EĞİTİM ENSTİTÜSÜ**

**YÜKSEK LİSANS TEZİ**

**TIMSS 2019 MATEMATİK ÖĞRENMEYİ SEVME ÖLÇEĞİ ÖLÇME  
DEĞİŞMEZLİĞİNİN BAYES YAKLAŞIK VE KLASİK FREKANSÇI  
İNCELENMESİ**

**Abdullah YARADILMIŞ**

**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI**

**Şanlıurfa**

**2026**



**T.C.  
HARRAN ÜNİVERSİTESİ  
LİSANSÜNTÜ EĞİTİM ENSTİTÜSÜ**

**YÜKSEK LİSANS TEZİ**

**TIMSS 2019 MATEMATİK ÖĞRENMEYİ SEVME ÖLÇEĞİ ÖLÇME  
DEĞİŞMEZLİĞİNİN BAYES YAKLAŞIK VE KLASİK FREKANSÇI  
İNCELENMESİ**

**Abdullah YARADILMIŞ**

**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI  
Tez Danışmanı: Doç Dr. RAGIP TERZİ**

**Şanlıurfa**

**2026**

## TEŞEKKÜR

Lisansüstü eğitimime başladığım ilk günden itibaren, engin bilgi ve tecrübeleriyle bana yol gösteren; desteğini, iyi niyetini ve sabrını hiçbir zaman esirgemeyen çok değerli danışmanım Doç. Dr. Ragıp TERZİ hocama en içten teşekkürlerimi sunarım. Akademik gelişimime rehberlik eden yapıcı eleştirileri, özverili katkıları ve emeği sayesinde bu süreci daha bilinçli, verimli ve anlamlı bir şekilde tamamlayabildim. Kendisine duyduğum minneti kelimelerle ifade etmek güç; bu yolculuğun en önemli dayanağı olduğu için sonsuz şükranlarımı iletmek isterim.

Yüksek lisans sürecim boyunca tanışma ve ders alma fırsatı bulduğum; akademik katkılarının yanı sıra bana hayat dersleri de kazandıran, derslerinde alanı derinlemesine öğrenmemizi sağlayan, kendisinden çok şey öğrendiğim ve derslerine her zaman büyük bir keyifle katıldığım değerli hocam Prof. Dr. Sedat ŞEN'e teşekkürlerimi sunarım. Tez jürimde yer almayı kabul ederek bu süreci benim için daha da anlamlı ve mutlu kılan kıymetli hocama şükranlarımı ifade etmekten onur duyarım.

Tez jürimde yer almayı kabul eden ve değerli katkılar sağlayan Doç. Dr. Levent YAKAR hocama teşekkür ederim. Yüksek lisans dönemi süresince ders alma fırsatı bulduğum kıymetli bölüm hocalarıma; Dr. Öğr. Üyesi Cahit POLAT'a, Doç. Dr. Derya EVRAN'a ve Dr. Öğr. Üyesi Eda GÖZÜYEŞİL'e şükranlarımı arz ederim. Ayrıca, Harran Üniversitesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı ile hayat yolunda çizgilerimizin kesişmesi benim için çok kıymetli olmuştur.

Birlikte bir yaşamı paylaştığım, hayatta her yükümü hafifleten, tüm başarılarımda yanımda olarak her türlü desteği sağlayan canım eşim Pınar YARADILMIŞ'a; varlıklarıyla kalbime huzur ve yaşamıma kattıkları sevgi ve mutlulukla bana daima güç veren biricik oğullarım Bünyamin'e ve Musab'a en içten sevgilerimi sunarım.

Bu tezi, ömrünü ilim yolunda geçiren ve bana öğrenmenin, öğretmenin değerini miras bırakan **rahmetli babama** ithaf ediyorum...

# İÇİNDEKİLER

ÖZET.....	i
ABSTRACT.....	ii
ŞEKİLLER DİZİNİ.....	iii
TABLOLAR DİZİNİ.....	iv
SİMGELER.....	v
KISALTMALAR.....	vi
1. GİRİŞ.....	1
1.1. Problem Durumu.....	1
1.2. Araştırmanın Amacı.....	5
1.3. Araştırmanın Önemi.....	6
1.4. Sınırlılıklar ve Sayıtlılar.....	7
1.5. Ölçme Değişmezliği.....	8
1.6. Çoklu Grup Doğrulayıcı Faktör Analizi.....	13
1.6.1. Yapısal Değişmezlik.....	14
1.6.2. Metrik Değişmezlik.....	15
1.6.3. Skalar Değişmezlik.....	16
1.6.4. Katı Değişmezlik.....	17
1.7. Frekansçı Uyum İndeksleri.....	20
1.7.1. Ki-Kare Uyum Testi (Chi-Square).....	21
1.7.2. Yaklaşık Hataların Ortalama Karekökü (Root Mean Square Error of Approximation RMSEA).....	22
1.7.3. Standartlaştırılmış Artık Ortalamaların Karekökü (Standardized Root Mean Square Residuals SRMR).....	23
1.7.4. Karşılaştırmalı Uyum İndeksi (Comparative fit index-CFI).....	23
1.7.5. Normleştirilmemiş Uyum İndeksi (Nonnormed fit index-NNF= Tucker Lewis Index - TLI).....	24
1.8. Bayes Teoremi.....	25
1.9. Bayes İstatistikleri.....	26
1.10. Bayes Yapısal Eşitlik Modellemesi.....	28
1.11. Bayes Yaklaşık Ölçme Değişmezliği.....	30
1.12. Bayes Uyum İndeksleri.....	34
1.12.1. Sonsal Tahmin P-Değeri (Posterior Predictive P-Value; PPP).....	34
1.12.2. Güven Aralığı (CI).....	35
1.12.3. Sapma Bilgi Kriteri (DIC).....	35
2. ÖNCEKİ ÇALIŞMALAR.....	38
2.1. ÇGDFA Yöntemi İle Yapılan Ölçme Değişmezliği Çalışmaları.....	38
2.2. BYÖD Yöntemi İle Yapılan Ölçme Değişmezliği Çalışmaları.....	43
3. GEREÇ VE YÖNTEM.....	49
3.1. Araştırmanın Deseni.....	49
3.2. Veri Seti.....	49
3.3. Verilerin Elde Edilmesi.....	50
3.4. Verilerin Analizi.....	51
3.4.1. Kayıp Veri.....	52
3.4.2. Uç Değer.....	53
3.4.3. Normallik.....	53
3.4.4. Çoklu Bağlantı Problemi.....	54
3.5. Doğrulayıcı Faktör Analizi Uygulaması.....	54
3.6. Çoklu Grup Doğrulayıcı Faktör Analizi Uygulaması.....	55
3.7. Bayes Yaklaşık Ölçme Değişmezliği Uygulaması.....	56
4. BULGULAR.....	58
4.1. Analiz Varsayımları.....	58
4.2. Betimsel İstatistikler.....	59
4.3. Doğrulayıcı Faktör Analizi Sonuçları.....	60

4.4. Birinci Alt Amaca Yönelik Bulgu ve Yorumlar.....	61
4.5. İkinci Alt Amaca Yönelik Bulgu ve Yorumlar.....	64
5. TARTIŞMA.....	66
6. SONUÇLAR.....	68
7. ÖNERİLER.....	71
KAYNAKÇA.....	73

## ÖZET

### YÜKSEK LİSANS TEZİ

#### TIMSS 2019 MATEMATİK ÖĞRENMEYİ SEVME ÖLÇEĞİ ÖLÇME DEĞİŞMEZLİĞİNİN BAYES YAKLAŞIK VE KLASİK FREKANSÇI YÖNTEMLERİ İLE İNCELENMESİ

ABDULLAH YARADILMIŞ

HARRAN ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ  
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI

Tez Danışman: Doç Dr. RAGIP TERZİ  
Yıl: 2026, Sayfa: 86

Bu araştırmanın amacı, TIMSS 2019 sekizinci sınıf öğrencilerine uygulanan Matematik Öğrenmeyi Sevme Ölçeği'nin örtük yapısının ölçme değişmezliğini belirlenen ülkelere (Kuveyt, Portekiz, Suudi Arabistan, Singapur, Türkiye ve Tayvan) göre incelemektir. Toplam 27.476 öğrenciden elde edilen veriler doğrultusunda, analiz öncesi varsayımlar test edilmiş ve her ülke için Doğrulamalı Faktör Analizi (DFA) uygulanarak tek faktörlü ölçme modellerinin geçerliliği değerlendirilmiştir. DFA sonuçlarına göre tüm ülkelerde model veri uyumu sağlanmış, bileşik güvenilirlik ve çıkarılan varyans ortalaması değerleriyle geçerlik ve güvenilirlik kanıtları ortaya konmuştur. Ölçeğin örtük yapısına ilişkin kesin ölçme değişmezliği, Çoklu Grup Doğrulamalı Faktör Analizi (ÇGDFA) ile incelenmiş; yapısal ve metrik düzeyde değişmezlik sağlanırken, skalar düzeyde sağlanamamıştır. Bayes kesin değişmezlik yöntemi (faktör yükünde ve madde sabitinde önsel tanımlanmadan) ile yapılan ölçme değişmezliği analiz sonuçları, ÇGDFA yöntemiyle elde edilen sonuçlarla tutarlılık göstermiştir. Bu tutarlılığın ardından, Bayes Yaklaşık Ölçme Değişmezliği yönteminde ölçme modeline ilişkin faktör yükü ve madde sabitleri için farklı önsel varyans düzeyleri dikkate alınmıştır. Bulgular, metrik düzeyde modelin tüm durumlarda veriye uyum sağladığı; skalar düzeyde ise bazı koşullarda uyumun sağlandığı görülmüştür. Sonuç olarak, kesin değişmezlik analizlerinde skalar değişmezlik sağlanmadığından ülkeler arasında madde sabitlerinin eşit olmadığı ve örtük değişkenin tüm ülkelerde aynı şekilde kullanılmadığı sonucuna ulaşılmıştır. Yaklaşık ölçme değişmezliği analizinde ise varyanslara belirli düzeyde esneklik tanındığında, yaklaşık skalar değişmezlik sağlamakta; bu da ülkeler arasında gösterge sabitlerinin eşit olduğu ve örtük değişkenin tüm ülkelerde aynı şekilde kullanıldığı sonucuna ulaşılmaktadır. Bu durumda, gruplar arasında faktör ortalamalarının karşılaştırılması da mümkün hale gelmektedir.

**ANAHTAR KELİMELELER:** Ölçme Değişmezliği, Çoklu Grup Doğrulamalı Faktör Analizi, Bayes Yaklaşık Ölçme Değişmezliği, Kesin Ölçme Değişmezliği ve Yaklaşık Ölçme Değişmezliği.

## **ABSTRACT**

### **MASTER THESIS**

#### **EXAMINING THE MEASUREMENT INVARIANCE OF THE TIMSS 2019 STUDENTS LIKE LEARNING MATHEMATICS SCALE USING BAYESIAN APPROXIMATE AND CLASSICAL FREQUENTIST METHODS**

**Abdullah YARADILMIŞ**

**HARRAN UNIVERSITY  
INSTITUTE OF GRADUATE EDUCATION  
DEPARTMENT OF MEASUREMENT AND EVALUATION IN EDUCATION**

**Thesis Supervisor: Assoc. Prof. Dr. Ragıp TERZİ**

**Year: 2026, Page: 86**

The purpose of this research is to examine the measurement invariance of the Student Like Learning Mathematics Scale administered to eighth-grade students in TIMSS 2019 across selected countries with established latent structures (Kuwait, Portugal, Saudi Arabia, Singapore, Turkey, and Taiwan). Based on data obtained from a total of 27.476 students, pre-analysis assumptions were tested, and the validity of the single-factor measurement models was evaluated by applying Confirmatory Factor Analysis (CFA) for each country. According to the CFA results, model fit was achieved in all countries, and evidence of validity and reliability was established using values for composite reliability and average variance extracted. The measurement invariance concerning the latent structure of the scale was analyzed using Multi-Group Confirmatory Factor Analysis (MGCFA); while structural and metric invariance were achieved, scalar invariance was not. The results of the measurement invariance analyses conducted using the Bayesian exact invariance method (without specifying priors for factor loading and item intercepts) were consistent with the findings obtained via the MGCFA method. Following this consistency, Bayesian Approximate Measurement Invariance analyses were performed by considering different prior variance levels for factor loadings and item intercepts. Findings revealed that the model consistently achieved fit at the metric level, while scalar invariance was supported under certain conditions. In conclusion, strict invariance analyses indicated that scalar invariance was not achieved, suggesting that item intercepts were not equivalent across countries and that the latent construct was not interpreted identically in all contexts. However, approximate measurement invariance analyses demonstrated that when a certain degree of flexibility in variances was allowed, approximate scalar invariance was achieved, implying equivalence of item intercepts and consistent interpretation of the latent construct across countries. Under these conditions, comparisons of factor means across groups became feasible.

**KEYWORDS:** Measurement Invariance, Multi-Group Confirmatory Factor Analysis, Bayesian Approximate Measurement Invariance, Exact Measurement Invariance and Approximate Measurement Invariance

## ŞEKİLLER DİZİNİ

Şekil 1.1. Çoklu Grup Doğrulatoryıcı Faktör Analizi Grafıksel Gösterimi.....	18
Şekil 1.2. Kesin ve Yaklaşık Ölçme Değişmezliđi.....	33

## TABLolar DİZİNİ

<b>Tablo 1.1.</b> Ölçme Değişmezliği Aşamaları.....	19
<b>Tablo 1.2.</b> Uyum İndeksleri Sınıfı.....	20
<b>Tablo 3.1.</b> Araştırma Örneklemi Oluşturan Öğrencilerin Ülkelere Göre Dağılımı.....	50
<b>Tablo 3.2.</b> Matematik Öğrenmeyi Sevme Ölçeği Maddeleri.....	51
<b>Tablo 3.3.</b> Araştırma Grubunu Oluşturan Ülkelerin Silinen ve Kalan Veri Sayıları.....	52
<b>Tablo 4.1.</b> MÖSÖ’de Yer Alan Maddelere Ait Betimsel İstatistikler.....	59
<b>Tablo 4.2.</b> Ölçme Modeline İlişkin Ülkelerin DFA Model Uyum İyiliği İndeksleri.....	60
<b>Tablo 4.3.</b> MÖSÖ’ye Ait Faktör Yükleri, Güvenirlilik Katsayıları ve AVE Değerleri.....	61
<b>Tablo 4.4.</b> Ölçme Değişmezliği Modeline İlişkin Ülkelerin ÇGDFA Yöntemine Göre Model Uyum İyiliği İndeksleri (MLR Kestirimi).....	61
<b>Tablo 4.5.</b> Ölçme Değişmezliği Modeline İlişkin Ülkelerin Bayes Yöntemine Göre Kesin Değişmezlik Model Uyum İyiliği İndeksleri.....	63
<b>Tablo 4.6.</b> Ölçme Değişmezliği Modeline İlişkin Ülkelerin BYÖD Yöntemine Göre Model Uyum İyiliği İndeksleri.....	64

## SİMGELER

$df$	Serbestlik Derecesi
$f$	Fonksiyon
$H_0$	Null Hipotez
$N$	Örneklem Büyüklüğü
$p$	$p$ -Değeri (Anlamlılık Düzeyi)
$\chi^2$	Ki-Kare
$\chi^2/sd$	Ki-Kare/Serbestlik Derecesi
$\Delta$	Fark/Değişim
$\Sigma$	Toplam Sembolü
$/$	Bölme Sembolü
$\cap$	Kesişim
$>$	Büyüktür

## KISATMALAR

<b>AFA</b>	Açımlayıcı Faktör Analizi
<b>AVE</b>	Çıkarılan Varyans Ortalaması
<b>BYEM</b>	Bayes Yapısal Eşitlik Modellemesi
<b>BYÖD</b>	Bayes Yaklaşık Ölçme Değişmezliği
<b>CFI</b>	Comparative Fit Index
<b>CI</b>	Güven Aralığı
<b>CI</b>	Koşul İndeksi
<b>CR</b>	Bileşik Güvenirlik
<b>ÇGDFA</b>	Çoklu Grup Doğrulayıcı Faktör Analizi
<b>DFA</b>	Doğrulayıcı Faktör Analizi
<b>EM</b>	Beklenti Maksimizasyonu
<b>IEA</b>	Uluslararası Eğitim Başarısını Değerlendirme Kuruluşu
<b>MCMC</b>	Markov Chain Monte Carlo
<b>ML</b>	Maksimum Olabilirlik
<b>MLR</b>	Sağlam Maksimum Olabilirlik
<b>MÖSÖ</b>	Matematik Öğrenmeyi Sevme Ölçeği
<b>OECD</b>	Ekonomik İşbirliği ve Kalkınma Örgütü
<b>PIRLS</b>	Uluslararası Okuma Becerileri Araştırması
<b>PISA</b>	Uluslararası Öğrenci Değerlendirme Programı
<b>PPP</b>	Posterior Predictive $p$ -Value (Sonsal Tahmin $p$ -Değeri)
<b>RMSEA</b>	Root Mean Square Error Approximation
<b>SRMR</b>	Standardized Root Mean Square Residual
<b>TD</b>	Tolerans Değeri
<b>TIMSS</b>	Uluslararası Matematik ve Fen Eğilimleri Araştırması
<b>TLI</b>	Tucker Lewis İndeks
<b>VIF</b>	Varyans Artış Faktörü
<b>YEM</b>	Yapısal Eşitlik Modellemesi

## 1. GİRİŞ

### 1.1. Problem Durumu

Küreselleşen dünyada bilgi ve teknolojiye yaşanan hızlı ve sürekli gelişmeler nedeniyle ülkeler, eğitim ihtiyaçlarını ve hedeflerini değiştirmeye ve geliştirmeye yöneltmekte; insan kaynağının bu değişim ve gelişimlere uyum sağlayabilmesi için eğitim öğretim süreci sürekli yenilenmektedir (Millî Eğitim Bakanlığı [MEB], 2015). Bu bağlamda, evrensel değişime ayak uyduramayan, kendini yenilemeyen sistemlerin verimli olması ve sürdürülebilirliği mümkün görülmemektedir. Eğitim sisteminde eğitim programlarının, öğretim yöntemlerinin ve öğretmen yeterliliklerinin sürekli iyileştirilmesi, öğrencilerin başarılarını artırmak için kritik bir öneme sahiptir. Çünkü, bir ülkenin eğitim sisteminin kalitesini ortaya koyan en önemli göstergelerden biri, öğrencilerin performanslarıdır. Öğrencilerin performanslarını değerlendiren analizlerde, ölçme ve değerlendirme uygulamaları kullanılmaktadır. Günümüzde, okul öncesi dönemden başlayarak eğitimin tüm kademelerinde amaçlanan hedeflere ulaşıp ulaşılmadığının değerlendirilmesi, bireylerin sınıflandırılması, karşılaştırılması, istihdam edilmesi gibi çeşitli amaçlarla ölçme ve değerlendirme çalışmaları yapılmaktadır. Ayrıca ülkeler, uluslararası ölçme ve değerlendirme çalışmalarından yararlanarak diğer eğitim sistemlerinden örnekler alabilmekte, bu örnekleri kendi yapısal özelliklerine uyarlamakta ve başarı düzeylerini karşılaştırma imkânı bulmaktadır.

Eğitim ve psikoloji gibi sosyal bilimlerde ölçme, fiziksel bilimlere kıyasla daha karmaşık bir süreçtir. Fiziksel ölçmeler, doğrudan gözlemlenebilen ve belirli bir birimle ölçülen özelliklere dayanırken, sosyal bilimlerde ölçme, soyut kavramları ve davranışları anlamaya yöneliktir. Genel bir tanımla ile ölçme, belirli kurallar çerçevesinde nesnelere veya bireylere ait özelliklerin, bu özelliklerin derecelerine göre sayılar ve sembollerle ifade edilmesidir (Magnusson, 1967; Turgut, 1987). Değerlendirme ise ölçme süreci sonucunda elde edilen verilerin belirli ölçütler doğrultusunda karşılaştırılması yoluyla, birey ya da nesnelere ilgili özellikleri hakkında karar verme sürecidir (Magnusson, 1967; Turgut, 1987).

Eğitimde yapılan ölçme ve değerlendirmeler, öğretmenlerin öğrenci gelişimini takip etmeleri, kullandıkları yöntem ve tekniklerin verimliliğini değerlendirmeleri ve

ülkenin eğitim politikalarının etkinliğini belirlemek açısından önemlidir (MEB, 2007). Ölçme ve değerlendirme süreçleri, eğitimin kalitesinin sürekli olarak artırılmasına ve eğitimdeki güçlü yönlerin yanı sıra zayıf noktaların da belirlenmesini sağlar. Sonuç olarak, ölçme ve değerlendirme süreçlerinin etkinliği, eğitimdeki başarıyı daha da ileriye taşıyacak önemli bir adım olacaktır. Bu nedenle eğitimde kullanılacak ölçme araçlarının geçerlik ve güvenirlik açısından yüksek standartlara sahip olması gerekir.

Geçerlik ve güvenirlik, eğitimde ölçme ve değerlendirmenin sağlıklı bir şekilde işleyebilmesi için temel iki kavramdır. Geçerlik, test sonuçlarının veya ölçme araçlarına dayalı yorum ve uygulamaların yeterliliğini ve uygunluğunu destekleyen deneysel kanıtlar ile teorik temellerin bütünsel bir değerlendirmesidir (Messick, 1995). Daha sade bir ifade ile geçerlik, bir ölçme aracının sadece ölçülmesi hedeflenen özelliği ölçmesi ve bunun dışında herhangi bir başka özelliği ölçmemesi olarak tanımlanmaktadır (Terzi, 2021). Güvenirlik ise ölçme sonuçlarının tesadüfi hatalardan arınık olma derecesi olarak tanımlanabilir (Baykul vd., 2003). Eğitimde bu iki kavram, özellikle büyük ölçekli sınavlar ve uluslararası değerlendirmelerde kritik rol oynar. Geçerli ve güvenilir ölçme araçları, eğitim politikalarının doğru şekilde şekillendirilmesine yardımcı olurken, geçerli ve güvenilir olmayan ölçme araçları ise hatalı sonuçlara yol açar ve bu da eğitimin kalitesini olumsuz etkilemektedir (Bademci, 2019; Turgut ve Baykul, 1992). Bu sebeple, geçerlik ve güvenirlik, yalnızca ölçme araçlarının doğruluğunu değil, aynı zamanda bu araçların eğitimdeki kararlar üzerinde yaratacağı etkileri de doğrudan belirlemektedir.

Geçerlik incelemelerinde, ölçme sonucuna göre yapılması beklenen çıkarımların geçerli olup olmadığını kapsam geçerliği, ölçüt dayanlı geçerlik ve yapı geçerliği olmak üzere üç farklı şekilde incelemek gerekir (Terzi, 2021). Yapı geçerliği, kapsam ve ölçüte dayalı geçerlik türlerine göre daha kapsamlı olduğundan geçerliğin en önemli türüdür (Loevinger, 1957). Hatta, Cohen vd.'ye (2005) göre yapı geçerliği, test puanlarına dayanmadığı, test maddeleriyle ölçülmek istenen örtük değişkenin göstergeleri olup olmadığına bakar, bu ise yapı geçerliğinin kapsam geçerliğini (testin, ölçmek istediği tüm yönleri kapsayıp kapsamadığı) ve ölçüte dayalı geçerliğini (testin sonuçlarının, ilgili ve dış ölçütle ne kadar ilişkili olduğu) de içeren daha geniş bir kavram olduğunu gösterir. Bu geçerlik türü, testin hedeflediği yapıyı ne kadar doğru yansıttığını kanıtlarla güçlendirmekte ve faktör analizi gibi yöntemlerle de

doğrulamaktadır (Büyüköztürk, 2002). Testin yapı geçerliği, yalnızca hedeflenen yapıyı doğru ölçmekle kalmaz, aynı zamanda ölçme aracının farklı gruplarda tutarlılık gösterip göstermediğiyle de ilgilenmektedir. Bu bağlamda ölçme değişmezliği kavramı ön plana çıkmaktadır.

Eğitim bilimleri alanındaki araştırmalar, farklı gruplar arasında yapılan karşılaştırmalarla genellikle ölçme araçlarının geçerliğini test eder ve Koziol (2010), testlerin farklı gruplarda aynı şekilde çalışıp çalışmadığını sorgulamanın önemli olduğunu vurgulamaktadır. Ölçme aracının, farklı gruplarda veya zaman dilimlerinde aynı yapıyı ölçüp ölçmediğini belirlemeye yönelik değerlendirmeye ölçme değişmezliği denir (Vanderberg ve Lance, 2000). Ölçme değişmezliği, özellikle uluslararası karşılaştırmalar ve çok gruplu araştırmalarda önemli bir yere sahiptir. TIMSS gibi uluslararası araştırmalarda, bir ölçme aracının farklı kültürlerde veya ülkelerde aynı şekilde geçerli olması gerekmektedir. Aksi takdirde uluslararası testlerin sonuçları, kültürlerarası karşılaştırmalar için yanıltıcı olabilmektedir (Davidov vd., 2014). Bu nedenle, uluslararası yapılan sınavlardaki veri setlerinin analizlerinde ölçme değişmezliği testlerinin yapılması kritik öneme sahiptir (Vandenberg ve Lance, 2000).

Uluslararası düzeyde eğitim başarılarını değerlendiren birçok büyük ölçekli araştırma bulunmaktadır. Bu sınavlar, genellikle uluslararası organizasyonlar tarafından düzenlenmektedir. Örneğin, Uluslararası Öğrenci Değerlendirme Programı (PISA), Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından yürütülmekte; ayrıca, Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS) ile Uluslararası Okuma Becerileri Araştırması (PIRLS), Uluslararası Eğitim Başarısını Değerlendirme Kuruluşu (IEA) tarafından gerçekleştirilen diğer önemli çalışmalardır (Mullis vd., 2020). Bu sınavlar, ülkelerin eğitim sistemlerini karşılaştırmak, güçlü ve zayıf yönlerini değerlendirmek için kullanılan kritik araçlardır. Yapılan uluslararası sınavlardan en yaygın olanlardan biri TIMSS'tir. TIMSS, dünya genelinde öğrenci başarılarını değerlendiren ve eğitim sistemlerini karşılaştıran önemli bir araştırma olarak öne çıkmaktadır. 1995 yılında ilk kez uygulamaya konulan bu araştırma, dört yıllık periyotlarla yapılmakta ve her dört yılda bir, dördüncü ve sekizinci sınıf düzeyindeki öğrencilerin matematik ve fen bilimlerinde kazandıkları bilgi ve becerileri ölçmektedir (Mullis vd., 2020). Ayrıca TIMSS kapsamında yapılan bağlamsal

anketler, ülkelerin eğitim kaynaklarını, öğretim yöntemlerini ve aile desteğini analiz etmeyi mümkün kılmaktadır (Martin vd., 2020).

Sosyal bilimlerde ölçme araçlarının gelişimi, özellikle küreselleşme ve teknolojik ilerlemeler sayesinde çok daha geniş bir kullanım alanı bulmaktadır (Usta vd., 2013). Ancak bu genişleme, beraberinde bazı zorluklar da getirmektedir. Bu zorluklar; farklı dil, kültür, etnik kimlik, ekonomik durum ve coğrafi koşullara sahip gruplar arasında aynı özelliklerin ölçülmesindeki güçlükten kaynaklanmaktadır. Bu durum, özellikle gözlemlenemeyen ve soyut özelliklerin (zekâ, başarı, tutum gibi) ölçülmesinde belirgin hale gelmektedir. Bu özelliklerin ölçülmesinde kullanılan araçlar, farklı toplumsal bağlamlardan etkilenebilir ve bu da ölçmelerin geçerliğini sorgulatabilmektedir (Özçelik, 2010). Bu noktada, TIMSS bağlam anketlerinin büyük önemi ortaya çıkmaktadır. Bu anketler, öğrenci başarısını etkileyebilecek faktörleri derinlemesine incelemek için tasarlanmaktadır. Matematik ve fen öğrenme deneyimlerinden tutumlara kadar pek çok psikolojik değişkeni ölçen bu anketler, öğrenci başarısını yalnızca bilgi düzeyine değil, aynı zamanda öğrencilerin tutumlarına, motivasyonlarına ve okul kaynaklarına da bağlanmaktadır (Mullis ve Martin, 2017). TIMSS, bu bağlam anketleriyle farklı ülkeler arasındaki başarı farklarını analiz etmeyi ve eğitimin kültürel ve sosyoekonomik yönlerini daha iyi anlamayı amaçlamaktadır (Mullis vd., 2020).

TIMSS 2019 verileri, matematik ve fen başarısını etkileyen en önemli faktörlerden birinin öğrencilerin duyuşsal özellikleri olduğunu ortaya koymaktadır. Öğrencilerin matematiksel başarıları, onların matematiksel tutumları, öz yeterlik inançları ve bu alandaki ilgileriyle güçlü bir şekilde ilişkilidir (Gülleroğlu, 2017). TIMSS 2019 uygulamasında kullanılan ölçekler, öğrencilerin matematiğe yönelik tutumlarını farklı boyutlar üzerinden değerlendirerek söz konusu ilişkinin derinlemesine incelenmesine olanak sağlamıştır. Örneğin, “Matematik Öğrenmeyi Sevme”, “Matematikte Kendine Güven” ve “Matematiğe Değer Verme” gibi ölçekler, öğrencilerin matematikle ilgili duygusal tutumlarını belirlerken, bu tutumların başarı üzerindeki etkisini de gözler önüne sermektedir. TIMSS 2019'da öğrencilerin matematikle ilgili tutumlarının, onların başarısını önemli ölçüde etkilediği ve duyuşsal özelliklerin öğrenci performansını belirleyen kritik faktörlerden biri olduğu gösterilmiştir (Mullis vd, 2020). Bu anketler aracılığıyla toplanan verileri

karşılaştırırken sıklıkla ortaya çıkan metodolojik sorunların ele alınmasına yönelik artan bir ihtiyaç vardır.

Uluslararası karşılaştırmalı çalışmalarda kritik metodolojik sorunlardan biri ölçme değişmezliğinin sağlanmasıdır. Ölçme değişmezliği, bir yapıyı grup üyeliğinden bağımsız olarak aynı şekilde ölçtüğünde geçerlidir ve anlamlı çapraz grup karşılaştırmaları için bir ön koşuldur. Ön koşulun karşılandığını ampirik olarak test etmek yerine varsaymak tehlikeli olabilir ve yanlış sonuçlara yol açabilir. Bu nedenle, bir çalışmada ölçme değişmezliğinin ampirik olarak test edilmesi gerekir. Literatürde çok sayıda ölçme değişmezliği yöntemleri mevcut olup (Van de Schoot vd., 2013), bunlar genel olarak kesin (geleneksel veya frekansçı) ve yaklaşık ölçme değişmezliği olmak üzere iki ana gruba ayrılmaktadır. Kesin ölçme değişmezliğinde parametreler (faktör yükleri, gösterge sabitleri ve hata varyansları) tamamen değişmez kabul edilmekte ve gruplar arasındaki parametre farklılıklarının kesinlikle sıfır olduğu varsayılmaktadır (Brown, 2015; Vandenberg, 2002; Vandenberg ve Lance, 2000). Buna karşılık, yaklaşık ölçme değişmezliğinde, ölçme parametreleri arasındaki grup farklılıkları tam sıfır kısıtlamaları yerine esnek bir biçimde yaklaşık sıfır kısıtlamalarıyla değerlendirilir (Muthên ve Asparouhov, 2012).

## 1.2. Araştırmanın Amacı

Bu araştırmanın genel amacı, matematik öğrenmelerini ve akademik başarıyı etkilediği düşünülen ve TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan Matematik Öğrenmeyi Sevme Ölçeği'nin (MÖSÖ) örtük yapısının ölçme değişmezliğini; belirlenen ülkelere (Kuveyt, Portekiz, Suudi Arabistan, Singapur, Türkiye ve Tayvan) göre incelemektir. Araştırmanın genel amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

1-TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan MÖSÖ'nün örtük yapısı, ÇGDFA yöntemine göre belirlenen ülkeler arasında ölçme değişmezliğini sağlamakta mıdır?

2- TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan MÖSÖ'nün örtük yapısı, BYÖD yöntemine göre belirlenen ülkeler arasında ölçme değişmezliğini sağlamakta mıdır?

### 1.3. Araştırmanın Önemi

TIMSS, uluslararası öğrenci değerlendirme sınavlarından biri olup, temel amacı öğrencilerin matematik ve fen alanlarındaki bilgi ve becerilerini ortaya koymak ve akademik başarılarını yordamalarına destek olmaktır. Bu doğrultuda uygulanan test sonuçları, öğrencilerin güçlü ve zayıf yönlerini belirlemeye olanak tanımaktadır. Öğrenci başarısını kapsamlı bir şekilde değerlendirmek adına uygulanan başarı testlerinin yanı sıra ölçekler de uygulanmaktadır. Bu ölçekler aracılığıyla öğrencilerin akademik performanslarını etkileyen tutum, özgüven ve derslere yönelik değer algısı gibi duyuşsal özelliklerin belirlenmesi amaçlanmaktadır (Mullis vd., 2020). Uluslararası başarı testleri ve ölçekler, ülkeler arasında geçerli karşılaştırmaların yapılabilmesi için ölçme değişmezliğinin test edilmesi gerekmektedir. Ölçme değişmezliği, ülkeler arası karşılaştırmaların doğru ve tutarlı bir biçimde yapılmasını sağlayarak olası yorum farklılıklarını önlemektedir. Yapılan araştırmalar, uluslararası sınavların ölçme değişmezliği konusu sıkça ele alındığı görülmektedir (Atılğan, 2022; Dokur, 2023; Yiğiter, 2024). Öğrencilerin akademik başarılarını belirleyen en önemli faktörlerden biri olan duyuşsal özellikler, bu bağlamda dikkatle incelenmelidir. Bu araştırmada, öğrencilerin başarılarını öngörmeye kritik bir değişken olan duyuşsal özellikler ele alınmış; TIMSS öğrenci anketinde yer alan ve öğrencilerin matematik öğrenmeye yönelik tutumlarını ölçen MÖSÖ'yü tercih edilmesi araştırmanın alan yazına katkı sağlayacağı öngörülmüştür.

Eğitimde ve psikoloji alanında ölçme ve değerlendirmeye konu olan değişkenler (başarı, ilgi ve tutum gibi) doğrudan gözlemlenemeyen örtük yapılar şeklinde karşımıza çıkmaktadır. Bu nedenle, bireylerin sahip olduğu diğer özelliklerden bağımsız olarak bu değişkenleri hassas ve doğru bir biçimde ölçmek oldukça zordur. Özellikle geniş ölçekli uluslararası testlerin (TIMSS ve PISA gibi) yaygınlaşmasıyla birlikte, kullanılan ölçme araçlarının farklı gruplarda aynı yapıyı temsil edip etmediğini belirlemek veya belirli bir grubun lehine sistematik farklılıklar gösterip göstermediğini incelemek için birçok yeni yöntem ve teknik geliştirilmiştir. Alan yazına bakıldığında, araştırmaların genellikle ölçme değişmezliğinin sağlanıp sağlanmadığına odaklandığı görülmektedir (Koçak, 2024; Polat, 2019; Uyar, 2021; Yiğiter, 2023).

Ölçme değişmezliği çalışmalarında en yaygın kullanılan yöntemlerden biri, faktör yükü, gösterge sabiti ve hata varyansları gibi parametrelerin kesin sıfır kısıtlaması ile test edildiği ÇGDFA yöntemidir (Pokropek vd., 2019). Ancak, tam sıfır kısıtlamaları özellikle çok sayıda grup veya zaman noktası olduğunda oldukça katı ve sınırlayıcıdır. Bu katı kısıtlamaları esnetmek amacıyla geliştirilen yeni yaklaşımlardan biri de BYÖD yöntemidir. Bu yöntem, ölçme parametreleri (faktör yükü ve gösterge sabiti) arasındaki grup farklılıklarının tam sıfır kısıtlamaları yerine yaklaşık sıfır kısıtlamalarıyla ele alınmasını sağlayarak, parametrelerin birbirine yaklaştırılmasına imkân tanımaktadır. Böylece, ölçekler arasındaki eşitlik zorunluluğu yerine bir miktar esneklik sunularak daha gerçekçi değerlendirmeler yapılabilmektedir.

#### 1.4. Sınırlılıklar ve Sayıtlar

1. TIMSS 2019 uygulamasına toplamda 64 ülke katılmıştır. Araştırma kapsamında, sekizinci sınıf matematik başarı ortalamalarına göre yüksek, orta ve düşük düzey başarı gösteren toplam 6 ülke (Tayvan, Singapur, Türkiye, Portekiz, Suudi Arabistan ve Kuveyt) ile sınırlandırılmıştır. Bu nedenle elde edilen bulgular tüm ülkeler için genellenemez. Yalnızca seçilen ülkelerin bağlamı çerçevesinde değerlendirilmelidir.

2. Bu çalışmada ölçme değişmezliği analizleri, ÇGDFA ve BYÖD yaklaşımları kullanılarak gerçekleştirilmiştir. Analizler, TIMSS 2019 öğrenci anketlerinde yer alan MÖSÖ'nün örtük yapısı ile sınırlandırılmıştır. Bu nedenle bulgular yalnızca MÖSÖ'ye ilişkin olup, TIMSS'in diğer ölçeklerine veya farklı veri setlerine genellenemez.

3. Araştırmaya başlandığı dönemde TIMSS 2023 verileri henüz yayımlanmadığı için, çalışmada TIMSS 2019'a ait veriler kullanılmıştır. Bu durum, çalışmanın bir sınırlılığı olarak değerlendirilmektedir.

4. Çalışmada kullanılan analizlerde Mplus ve SPSS gibi yazılımların varsayılan ayarları tercih edilmiştir. Farklı yazılım seçenekleri veya alternatif ayarların kullanılması sonuçlarda değişikliklere yol açabileceğinden, elde edilen bulgular yalnızca kullanılan yazılımların varsayılan ayarları çerçevesinde genellenebilir.

5. Araştırmada Türkiye, Tayvan, Singapur, Portekiz, Suudi Arabistan ve Kuveyt örneklerinden elde edilen verilerin, tüm evreni temsil ettiği varsayılmıştır.

6. Araştırmaya Türkiye, Tayvan, Singapur ve Portekiz bilgisayar tabanlı, Suudi Arabistan ve Kuveyt ise kağıt kalem yöntemiyle katılmış olup, iki farklı değerlendirme yöntemi arasında anlamlı bir fark olmadığı varsayılmıştır.

7. TIMSS 2019 uygulamasına katılan sekizinci sınıf öğrencilerinin anket maddelerini samimi ve objektif bir şekilde cevapladıkları, yani öğrenci anketi tutum maddelerini yanıtlarken gerçek tutumlarını yansıttıkları varsayılmıştır.

### 1.5. Ölçme Değişmezliği

Ölçme ve değerlendirme çalışmalarının kökleri antik dönemlere kadar uzansa da (Turgut ve Baykul, 2012), ölçme değişmezliği kavramı akademik yazına ancak yirminci yüzyılın ikinci yarısında girmiştir (Meredith, 1964; Struening ve Cohen, 1963). Son elli yılda ölçme değişmezliği değerlendirme yöntemleri önemli gelişmeler katetmiştir (Leitgöb vd., 2022; Van de Schoot vd., 2012). Bu kavramın gelişimine Karl Jöreskog, Roger Millsap, Michael Meredith, Gregor Muthén ve Peter Bentler gibi bilim insanları büyük katkıda bulunmuştur. Faktör yapılarının değişmezliğini detaylı bir şekilde elen alan ilk bilim insanı Jöreskog (1971) olmuştur. Ölçme değişmezliği kavramı ise, Byrne vd. (1989) tarafından tanıtılmış ve daha sonra test edilmiştir. Son yıllarda ölçme değişmezliği testlerinde kullanılan istatistiksel yöntemlerin daha erişilebilir hale gelmesi, araştırmacılar için bu testleri uygulamayı kolaylaştırmış ve yaygınlaştırmıştır (Putnick ve Bornstein, 2016).

Alan yazında ölçme değişmezliği kavramı ölçme denkliği (measurement equivalence), ölçüm değişmezliği (measurement invariant) ve ölçme karşılaştırılabilirliği (measurement comparability) olarak da isimlendirilir. Bu farklı isimlendirmeler olduğu gibi alan yazında farklı tanımlara da rastlamak mümkündür. Meredith ve Millsap (1992) ile Mellenbergh (1989) ölçme değişmezliğini matematiksel olarak, belirli bir gözlenen puanın elde edilme olasılığının grup üyeliğinden bağımsız olması şeklinde tanımlamıştır. Little'e (1997) göre ölçme değişmezliği, ölçeğin uygulandığı örneklemden gelen farklı yetenek seviyelerindeki grupların, ölçek ile ölçülmek istenen yapının tahminlenen değerlerinin matematiksel olarak yakın değerleri almasıdır. Ölçme değişmezliği, psikolojik bir ölçme aracının güvenilirlik ve yapı geçerliği gibi psikometrik özelliklerinin farklı gruplarda biçimsel olarak eşitliğini değerlendirme süreci olarak tanımlanmaktadır (Herdman vd., 1998). Widaman ve Rice (1997), bu kavramı örtük değişkenlerle gözlenen değişkenler

arasındaki ilişkinin gruplar arasında tutarlılığını ifade eden bir yapı olarak açıklamıştır. Byrne ve Watkins'e (2003) göre ölçme değişmezliği, farklı alt gruplardaki bireylerin ölçme aracındaki maddeleri aynı biçimde algılayıp yorumlamaları durumudur. Horn ve McArdle (1992) ise ölçme değişmezliğini, bir olgunun farklı gözlem ve inceleme koşullarında dahi aynı niteliği ölçüp ölçmediğini belirleme süreci olarak tanımlamaktadırlar.

Psikoloji, eğitim ve sosyal bilimler alanlarında yapılan büyük ölçekli değerlendirmeler, bilişsel (matematik, fen, okuma becerileri) ve duyuşsal (tutum, motivasyon, kaygı) yapılar açısından farklı gruplar arasında karşılaştırmalar yapmayı amaçlamaktadır. Bu karşılaştırmalar, zaman içinde değişimleri veya okul türleri gibi ulusal düzeydeki farkları, ayrıca kültürlerarası farklılıkları incelemek amacıyla yaygın biçimde kullanılmaktadır (Davidov vd., 2014). Psikoloji, eğitim ve sosyal bilimlerde doğrudan gözlemlenemeyen bu yapılar doğrudan da ölçülemez. Alternatif olarak testler veya ölçekler aracılığı ile yaklaşık olarak hesaplanabilir. Bu yaklaşık durumun sebebi yapılan ölçme işlemlerine zorunlu olarak bir miktar ölçme hatasının dahil olmasıdır. Bu durum aslında klasik test kuramının bir sınırlılığıdır. Bu sınırlılık günümüzde de önemli bir konu olan ölçme değişmezliğini gündeme getirmiştir (Crocker ve Algina, 1986). Çünkü, gruplar arası farklılıkları ortaya koymayı amaçlayan bir araştırmada elde edilen anlamlı farkların, davranışsal düzeydeki gerçek farklılıklardan mı yoksa ölçme aracının psikometrik özelliklerinden mi kaynaklandığını ayırt edebilmek, ilgili yapıların gruplar arasında ölçme değişmezliğinin sağlanmasına bağlıdır (Cheung ve Rensvold, 2002; Millsap, 2011).

Ölçme değişmezliği psikoloji, eğitim ve sağlık gibi farklı araştırma alanlarında bir konu olup, ilgili çalışmalar incelendiğinde, bu araştırmaların ortak noktasının gruplar arası karşılaştırma olduğu görülmektedir (Vandenberg ve Lance, 2000). Gruplar arasında anlamlı karşılaştırmalar yapabilmek için ölçülen gizil değişkenin tüm alt gruplarda aynı olması gerekmektedir. Ölçme aracının farklı demografik özelliklere sahip bireyler veya gruplar üzerinde uygulandığında, bu farklılıkları göz önünde bulundurarak karşılaştırmalar yapabilmek için ölçme değişmezliği önemli bir ön koşuldur. Bu durum, ölçülen yapının farklı gruplarda genellenebilirliğini sağlamak açısından büyük önem taşımaktadır (Brown, 2015). Uluslararası verilere erişimin kolaylaşmasıyla, birden fazla grup arasında karşılaştırmalar yapan kültürlerarası ve

uluslararası çalışmaların sayısı artmıştır (Rutkowski ve Svetina, 2014). Örneğin, bir matematik testi birden fazla ülkede uygulanıyorsa ve amaç ülkeler arasında performansı karşılaştırmaksa araştırmacı, aracın tüm ülkelerde geçerli olduğunu ve puanların ülkeler arasında karşılaştırılabileceğini göstermesi gerekir. Benzer şekilde, farklı ülkelerdeki katılımcıların belirli bir konuya yönelik tutumlarını karşılaştıran çalışmalarda, yanıtların karşılaştırılabilirliğini sağlamak ve kullanılan ölçeklerin tüm ülkelerde geçerli olduğunu göstermek için bazı kanıtlar sunulmalıdır. Ancak ölçme değişmezliği, gruplar arasındaki ölçme puanlarının eşitliği ile karıştırılmamalıdır. Puan eşitliği, iki grubun aynı teorik yapıya sahip olduğu ve benzer düzeyde puanlar aldığı anlamına gelir. Örneğin, matematiğe yönelik tutumları ölçen bir çalışmada, iki farklı grubun ortalama puanlarının aynı seviyede olması, bu grupların matematiğe karşı benzer bir tutuma sahip olduğunu gösterir. Ölçme değişmezliği ise ölçülen yapının gruplar arasında güvenilir bir şekilde karşılaştırılabilmesini sağlar. Bu ise grupların aynı puanı alacağı anlamına gelmez, sadece elde edilen puanların, farklı gruplar için anlamlı bir şekilde karşılaştırılabileceğini gösterir. Bu nedenle, karşılaştırılabilir puanlar bazen eşit olabilirken bazen de farklılık gösterebilmektedir.

Alan yazında, temel konusu ölçme değişmezliği olan çalışmalar dışında ölçme değişmezliğinin test edildiği çok nadir görülmektedir. Ölçme değişmezliğinin test edilmemesi, sonuçların hatalı yorumlanmasına sebep olarak (Van de Vijver ve Leung, 2021) çalışmaların geçerliğini önemli derecede olumsuz etkilemektedir (Vandenberg ve Lance, 2000). Eğer ölçme değişmezliği sağlanmazsa, grupların sağlıklı bir şekilde karşılaştırılması mümkün olmamaktadır. Örneğin, iki grubun ortalamaları arasındaki anlamlı farkı test etmek amacıyla bağımsız örneklem  $t$  testi uygulandığını ve istatistiksel olarak anlamlı bir fark elde edildiğini varsayalım. Ancak, söz konusu ölçme aracının her iki grup için aynı yapıyı ölçmemesi, yani ölçme değişmezliğinin sağlanmamış olması durumunda, elde edilen sonuçların geçerliği sorgulanabilir. Bu durumda, karşılaştırılan yapıların özdeş olduğuna dair herhangi bir güvenilir kanıt bulunmadığından, yapılan değerlendirmeler bilimsel açıdan tutarlı bulunmayacaktır. Millsap (2011) ölçme değişmezliğini, ağırlıkları eşit olan ancak şekilleri farklı iki cismin, ölçme yapacak terazinin şekil farklılıkları nedeniyle farklı ağırlıklarda ölçülmesine benzetmektedir. Eğer bir terazi, yalnızca cismin ağırlığını ölçmesi gerekirken şekilden etkilenerek ölçme sonuçlarını değiştirirse, bu durum geçerlik

sorununa yol açmaktadır. Bu eğitim ve psikoloji alanlarında kullanılan ölçeklerde de benzer bir sorun ortaya çıkabilir; ölçme aracının, yalnızca ölçülmek istenen özelliği ölçmesi ve gruplar arası farklılıklardan etkilenmemesi beklenmektedir. Ayrıca ölçme değişmezliği, ölçme araçlarının önemli bir özelliği olan yapı geçerliği test eden bir yaklaşımdır ve bu testin sonuçları, her gruptan elde edilen faktör yapısı ile uyum göstermesi gerekmektedir (Başusta ve Gelbal, 2015). Bir ölçme modelinin tüm gruplar için aynı yapıya sahip olması demek, söz konusu ölçme modelinden elde edilen faktör yüklerinin, gösterge sabitlerinin ve hata varyanslarının aynı olması demektir (Byrne, 2013; Jöreskog ve Sörbom, 1993). Yani, ölçme değişmezliği, bir ölçeğin farklı uygulamalarında aynı parametre değerlerine sahip olmasını gerektirir (Şen, 2023).

Ölçme değişmezliğinin olmaması, bir yapının ölçmesine ilişkin önyargılardan (örneğin, bir yapı tüm katılımcı ülkelerde eşdeğer şekilde ölçülemeyebilir), kullanılan ölçme metodolojisinden (örneğin, Likert tipi maddelere verilen yanıt tarzları ülkeler arasında farklılık gösterebilir) veya belirli maddelere verilen yanıtların farklı olmasından kaynaklanabilir (Van de Vijver ve Leung, 1997). Ölçme değişmezliğine bir örnek olarak, ortaokul öğrencilerinin genel yetenek sınavında şu matematik sorusu verilebilir: Bir sayı, kendisinin üç katının beş fazlasına eşittir. Bu sayıyı bulunuz. Matematik diline hâkim olan bir öğrenci bu ifadeyi doğrudan cebirsel forma ( $x = 3x + 5$ ) dönüştürerek denklemleri çözer ve doğru cevaba ulaşır. Ancak matematik diline aşinalığı olmayan öğrenciler, cebirsel ifadeleri doğrudan matematiksel modele dönüştürmekte zorlanır. Bu durumda, soru matematiksel işlem yapma becerisinden çok, öğrencinin matematiksel dile olan aşinalığını test etmiş olur. Eğer sorular farklı gruplar tarafından aynı şekilde yorumlanmıyorsa, ölçme değişmezliği sağlanamaz ve bu durum, değerlendirme sürecinin geçerliğini olumsuz etkilemektedir.

Millsap'a (2011) göre ölçme değişmezliği, matematiksel olarak koşullu olasılık bağlamında tanımlarken, aynı yetenek düzeyindeki ( $W$ ) bireylerin, ait oldukları gruptan bağımsız ( $V$ ) olarak,  $X$  maddesi için eşit puan alma olasılığına sahip olması esasına dayandırmaktadır. Bu tanım, cebirsel olarak aşağıdaki formülle ifade edilir:

$$P(X | W, V) = P(X | W)$$

Burada  $X$ , gözlemlenmiş ölçmelere ilişkin puanları gösteren  $q \times 1$  boyutlu rassal değişken vektörüdür.  $W$ ,  $X$  için tasarlanan  $r \times 1$  boyutlu gizil değişkenler vektörü olsun.  $V$ ,  $W$  dikkate alındığında  $X$  ile ilgisi olmaması gereken kişi özelliklerini tanımlayan  $s \times 1$  boyutlu ölçülmüş değişkenler vektörüdür.  $V$ , bireylerin cinsiyet ve etnik köken gibi demografik özelliklerine dayalı olarak farklı gruplara ayrılmasını sağlayan tek boyutlu bir değişkendir. Tüm  $X$ ,  $W$  ve  $V$  için  $P(X | W)$ ,  $W$  verildiğinde  $X$ 'in koşullu olasılık fonksiyonudur.  $X$  ayrık bir değişken olduğunda, ayrık koşullu olasılık ve  $X$  sürekli bir değişken olduğunda ise koşullu olasılık yoğunluk fonksiyonu olarak adlandırılır (Millsap, 2011).

Ölçme değişmezliğinin daha somut biçimde açıklanabilmesi amacıyla, tek faktörlü bir ölçme modeline dayalı, basit bir örnek üzerinden konu ele alınmaktadır. Bu modelde, tek bir gizil faktör  $\eta$  için (örneğin matematiksel akıl yürütme yetkinliği) üç gözlenen değişken  $y_1, y_2, y_3$  aracılığıyla ölçülmektedir. Söz konusu göstergeler, katılımcıların problem çözme becerisi  $y_1$ , mantıksal düşünme yeteneği  $y_2$  ve matematiksel kavramları anlama düzeylerini  $y_3$  yansıtmaktadır. Yanıtların 1 (çok düşük) ile 7 (çok yüksek) arasında değişen yedili Likert tipi bir ölçekle toplandığı varsayılmaktadır. Matematiksel yetkinliğe ilişkin ölçmeler, ölçme hatalarına duyarlı olduğundan, her bir göstergeye özgü hata terimleri  $\varepsilon_1, \varepsilon_2, \varepsilon_3$  ve her bir gösterge ile gizil faktör arasındaki ilişkiyi tanımlayan faktör yükleri  $\lambda_1, \lambda_2, \lambda_3$  modele dahil edilmektedir.

Bununla birlikte, söz konusu gizil değişken ortalamasını veya onun kovaryansı olan “öğrencinin günlük matematik çalışma süresi” değişkeni ile ilişkisi bağlamında iki farklı grup (örneğin iki farklı okul sistemi veya ülkeler) arasında karşılaştırmak istediğimizde, ölçme değişmezliği kritik bir önem kazanmaktadır. Gizil değişken, ölçmeleri değişmezlik gösterdiğinde, ortalamalar ve gruplar arasındaki ilişkiler anlamlı bir şekilde karşılaştırılabilir. Aksi takdirde, ortalaması ve kovaryansındaki gruplar arası benzerlikler veya farklılıklar yalnızca gerçek farklılıkların bir sonucu değil, aynı zamanda metodolojik etkilerin bir sonucu da olabilir.

Ölçme değişmezliğini test etmek için çok sayıda yöntem bulunmaktadır (Van de Schoot vd, 2014). Bu yöntemler, araştırmanın amaçları doğrultusunda kesin ölçme değişmezliği ve yaklaşık ölçme değişmezliği olmak üzere iki gruba ayrılır. Önceki

bölmelerde açıklandığı üzere kesin ölçme değişmezliğinde, belirli hipotezlerin sıkı bir şekilde test edilmesini temel alırken, yaklaşık ölçme değişmezliğini daha esnek bir perspektifle değerlendirme yaparak küçük farkları tolere edebilen yaklaşımlar sunmaktadır. Yani kesin ölçme değişmezliğinde parametreler (faktör yükleri, gösterge sabitleri ve hata varyansları) tamamen değişmez kabul edilmekte ve gruplar arasındaki parametre farklılıklarının kesinlikle sıfır olduğu varsayılmaktadır. Buna karşılık, yaklaşık ölçme değişmezliğinde, ölçme parametreleri arasındaki grup farklılıkları tam sıfır kısıtlamaları yerine esnek bir biçimde yaklaşık sıfır kısıtlamalarıyla değerlendirilir. Araştırma doğrultusunda, kesin ölçme değişmezliği kapsamında ÇGDFA'ya ve yaklaşık ölçme değişmezliğinde kapsamında ise BYÖD'ye detaylı olarak yer verilmektedir.

### 1.6. Çoklu Grup Doğrulayıcı Faktör Analizi

Araştırmacılar, gerçek dünyada farklı gruplardan oluşan örneklemelerden veri toplayabilir. Ancak grupların farklılıklarını göz ardı ederek veriyi tek bir grup üzerinden analiz etmek, yanıltıcı olabilir. Eğer yapılan analizlerde incelenen parametreler farklı/alt gruplara göre değişkenlik gösteriyorsa, bu durum elde edilen çıkarımların doğruluğunu sorgulanır hale getirir (Şen, 2023). ÇGDFA, farklı gruplara mensup bireylerin bir ölçme aracına verdiği yanıtların doğru değerlendirilmesini sağlayan istatistiksel bir yöntemdir (Brown, 2015). Başka bir ifadeyle, gözlenen değişkenlerin gizil değişkenlerle nasıl ilişkilendiğini ve oluşturulan ölçme modelinin, farklı gruplarda aynı şekilde işleyip işlemediğini inceleyen bir analiz yöntemidir (Hirschfeld ve Brachel, 2014; Pendergast vd., 2017; Vandenberg ve Lance, 2000). ÇGDFA, Lord ve Novick'in (1968) zihinsel test puanları teorisini bilişsel test verileriyle örneklendiren Jöreskog (1971) tarafından literatüre kazandırılmıştır. Bu yöntem, ardışık ve iç içe geçmiş adımlardan oluşup, Yapısal Eşitlik Modellemesi (YEM) analizlerine dayanarak ölçme değişmezliğini test etmek amacıyla kullanılmaktadır (Whitaker ve McKinney, 2007). Literatür incelendiğinde ÇGDFA, ölçme değişmezliğini belirlemenin en sık tercih edilen yöntemlerindedir. Genel olarak, ölçme modelinin parametrelerine (faktör yükleri, gösterge sabitleri ve hata varyansları) en az iki farklı örnekleme eşitlik kısıtlaması getirerek model uyumunu değerlendirmek amaçlanmaktadır (Şen, 2023). Bu yaklaşım, karşılaştırılan gruplarda

faktör yükleri, gösterge sabitleri ve hata varyanslarının tamamen eşit olduğu varsayımına dayanarak ölçme değişmezliğini test etmektedir (Raykov vd., 2013).

Ölçme değişmezliği, ÇGDFA yöntemi ile aşamalı ve hiyerarşik bir biçimde incelenir. Her aşamada belirlenen hipotezler, model uyum katsayıları aracılığıyla test edilmektedir (Başusta ve Gelbal, 2015). Test sürecinin başlayabilmesi için öncelikle, kullanılan ölçme aracının faktör yapısına dayalı temel modelinin her grupta araştırma verisiyle yeterli uyum göstermesi gerekir. Daha sonra her bir değişmezlik düzeyindeki model uyumunun, daha az kısıtlama içeren önceki model uyumu karşılaştırılması yoluyla belirlenir (Chen, 2007). Ölçme değişmezliği analizleri daha yüksek düzeydeki değişmezlik (gruplar arasında daha fazla parametrenin eşit olmaya kısıtlandığı durum) seviyelerinin kabulü, alt düzeydeki değişmezlik (daha az parametrelerin eşitlendiği durum) seviyelerin sağlanmasını gerektirir. Yani daha düşük değişmezlik seviyeleri, yüksek değişmezlik seviyelerinin ön koşulu olarak değerlendirilebilir (Kankaras ve Moors, 2010; Van de Vijver, 1998). Ölçme değişmezliği, ÇGDFA yöntemi ile hiyerarşik olan dört aşamada test edilir. Bu aşamalar; yapısal (configural) değişmezlik, metrik (metric) değişmezlik, skalar (scalar) değişmezlik ve katı (strict) değişmezliktir (Meredith, 1993). Her bir aşamaya sonraki bölümde detaylı bir şekilde yer verilmiştir.

### 1.6.1. Yapısal Değişmezlik

Ölçme değişmezliğinin ilk hiyerarşik aşamasıdır. En düşük ölçme değişmezlik aşaması olan yapısal değişmezlik, ölçeğin gruplar arasında aynı ölçme modeline sahip olup olmadığını test eder (Kline, 2011). Yani ölçme modeli aynı sayıda faktörü içermeli ve gözlenen değişkenler aynı faktörlere yüklenmelidir. Temel model olması nedeniyle, yapısal değişmezlik genel model uyumuna göre değerlendirilir (Şen, 2023). Bu aşamada faktör yükleri, gösterge sabitleri ve hata varyansları gibi herhangi bir parametreye kısıtlama getirilmez. Yapısal değişmezliğin sağlanması durumunda, ölçülen yapının ve bu yapıya ilişkin maddelerin farklı gruplarda aynı anlamı taşıdığı kabul edilir (Gregorich, 2006). Ancak yapısal değişmezlik sağlanamazsa, grupların farklı yapıları ölçtüğü sonucuna varılır ve ölçme değişmezliğinin sonraki aşamalarına geçilmez.

Yapısal değişmezlik genel model uyumu verileri tarafından desteklense bile, yine de anlamlı karşılaştırmalar yapmamıza izin vermez. Bunun nedeni, anlamlı

karşılaştırmalar için daha ileri değişmezlik (metrik ve skalar) seviyeleri gerektirmesidir. Çünkü bu aşamada elde edilen benzerlikler veya farklılıklar gerçek farklılıklardan ziyade metodolojik sebeplerden kaynaklanıyor olabilir. Bu nedenle yapısal değişmezlik, daha kısıtlayıcı modellerin karşılaştırılabilmesi açısından kritik bir öneme sahiptir ve metrik ile skalar değişmezlik düzeylerinin test edilebilmesi için temel bir ön koşul olarak kabul edilmektedir (Cheung ve Rensvold, 2002; Little, 1997). Literatür incelendiğinde yapısal değişmezlik; örüntü değişmezliği, faktör sayılarının eşitliği, biçimsel ve şekilsel olarak da isimlendirildiği görülmektedir.

### 1.6.2. Metrik Değişmezlik

Aşamalılık gereği yapısal değişmezliğin (faktör yapısının eşitliği) doğrulanması durumunda metrik değişmezliğin (faktör yüklerin eşitliği) testine geçilmektedir. İkinci ölçme değişmezliği seviyesi olan metrik değişmezlik, katılımcıların gruplar arasında gizil yapıya aynı anlamı atfedip atfetmediklerini test eder. O zaman yalnızca gizil yapının gruplar arasında karşılaştırılabilir olduğunu bilmekle kalmayız, aynı zamanda maddelerle temel yapı arasındaki ilişkinin gruplar arasında aynı olduğunu da biliriz. Metrik değişmezlik sağlandığında, ölçme aracına ait maddelerin gruplar arasında aynı yapıyı ölçtüğü yorumu yapılabilir. Ancak değişmezlik sağlanmadığında ise ölçme aracındaki maddeler bir grup için diğer gruptakilere göre daha anlamlı veya dikkat çekici olabilir. Bu durum ölçeğin belirli bir gruba karşı yanlı davrandığı şeklinde yorumlanabilir (Byrne ve Watkins, 2003).

Parametre sınırlamasının ilk kez uygulandığı bu modelde, gösterge sabitleri (kesişimler) ile hata varyansları serbest bırakılırken, faktör yüklerine kısıtlama getirilir (Gregorich, 2006). YEM’de faktör yükleri, gizil değişkenlerden madde yanıtlarını tahmin etmek için kullanılan regresyon katsayılarıdır. Metrik değişmezliğin sağlanması, bir gruptaki faktör yükündeki bir birimlik değişimin, diğer grupta da aynı etkiyi göstermesi anlamına gelir (Chen vd., 2005; Evran, 2012). Bu düzeydeki değişmezliği test etmek için, bir gruptan elde edilen faktör yükleri diğer gruba sabitlenir ve modelin uyum indeksleri değerlendirilir. Metrik değişmezliğinin sağlanıp sağlanmadığı, faktör yüklemeleri kısıtı olan metrik değişmezliği modelin uyumunu herhangi bir kısıt olmayan yapısal değişmezlik modelinin model uyumuyla karşılaştırarak test edilir. Model uyumunda ki-kare ( $\chi^2$ ) testi anlamlı değilse veya CFI

ve SRMR gibi model veri uyumları manidar bir fark yoksa metrik değişmezliğinin veriler tarafından desteklendiği varsayılır. Ancak, yapısal ve metrik değişmezliğin varlığında bile, yapının aynı şekilde ölçüp ölçmediği hala belirsizdir (Şen, 2023). Çünkü model hala ortalamaları güvenle karşılaştırmaya izin vermez. Bu yalnızca skalar değişmezliğin varlığında elde edilebilir (Cheung ve Rensvold 2002; Vandenberg ve Lance, 2000). Metrik değişmezlik sağlanması durumunda, skalar ölçme değişmezliği aşamasına geçilir. Literatür incelendiğinde metrik değişmezlik, zayıf değişmezlik olarak da isimlendirildiği görülmektedir.

### 1.6.3. Skalar Değişmezlik

Ölçme değişmezliğinin değerlendirilmesinde üçüncü aşama, skalar değişmezliktir (Steenkamp ve Baumgartner, 1998). Bu düzeyde, metrik değişmezlikte aranan faktör yüklerinin eşitliğine ek olarak, gösterge sabitlerinin de gruplar arasında eşit olması gerekmektedir (Wu vd., 2007). Dolayısıyla skalar değişmezlik testinde, faktör örüntüsü ve faktör yükleriyle birlikte gösterge sabitlerine kısıtlama getirilirken, hata varyansları serbest bırakılır. Skalar değişmezliğin sağlanıp sağlanmadığı, gösterge sabitlerinin sınırlandırıldığı modelin, metrik değişmezlik modeline göre uyum indeksleri açısından anlamlı bir fark gösterip göstermediğiyle değerlendirilir. Eğer model uyumunda anlamlı bir fark yoksa, skalar değişmezliğe ilişkin kanıt elde edilmiş olur. Bu durum, gözlenen puanların gizil puanlarla tutarlı biçimde ilişkili olduğunu ve gizil yapıda aynı puana sahip bireylerin, grup farkı gözetmeksizin gözlenen değişkende aynı puanı aldıklarını gösterir (Milfont ve Fischer, 2010). Skalar değişmezliğin sağlanması, katılımcıların ölçme aracını tüm gruplarda benzer şekilde kullandıkları varsayımını destekler ve böylece aynı yapının tutarlı biçimde ölçüldüğü ileri sürülebilir (Şen, 2023). Aksi takdirde belirli bir grubun, diğer gruplarla kıyasla maddelere sistematik olarak daha yüksek veya daha düşük cevaplama eğiliminden olduğu söylenir. Örneğin, aynı okulda matematik dersi için yapılan ortak bir sınav ele alınabilir. Eğer A sınıfındaki bir öğrencinin aldığı 70 puan, B sınıfındaki bir öğrencinin 50 puan ile aynı bilgi ve beceri düzeyini temsil ediyorsa, bu durum sınavda bir yanlılık olduğunu gösterir. Yani matematiksel olarak aynı bilgi ve beceri düzeyine sahip iki öğrenci, farklı sınıflarda olmalarından dolayı farklı puan alıyorsa, ölçme aracının gruplar arasında tutarlı sonuçlar vermediği anlamına gelir. Böyle bir durumda, sınav sonuçları gerçekte öğrencilerin matematik bilgisini değil, sadece buldukları sınıfın

değerlendirme sistemine göre şekillenen puanları yansıtıyor olabilir. Böyle bir durumda skalar değişmezliğin sağlanmadığı söylenir.

Yapısal ve metrik değişmezlik, farklı gruplardaki kovaryanslar veya standartlaştırılmamış regresyon katsayılarının anlamlı şekilde karşılaştırılmasını sağlar. Ancak bu seviyelerdeki değişmezliğin sağlanması, gruplar arası ortalamaların doğrudan karşılaştırılabilir olması için yeterli değildir (Şen, 2023). Oysa birçok araştırmada, grupların örtük ortalamalarının karşılaştırılması kritik bir öneme sahiptir (Benedict vd., 1998). Skalar değişmezlik geçerli olduğunda, grupların gözlenen değişken ortalamaları arasındaki farkların, gizli değişkenlerdeki farklardan kaynaklandığı sonucuna varılabilir. Ancak skalar değişmezlik sağlanmadığında, maddelerin faktörleri açıklama düzeyi gruplar arasında değişiklik gösterebilir. Gösterge sabiti, örtük değişkenin sıfır olduğu durumda gözlenen değişkenin aldığı değeri ifade eder. Gruplar arasında gösterge sabiti farklarının olup olmadığının belirlenmesi, güçlü bir geçerlik kanıtı olarak değerlendirilir (Vandenberg ve Lance, 2000). Literatür incelendiğinde skalar değişmezlik; ölçek değişmezliği ve güçlü değişmezlik olarak da isimlendirildiği görülmektedir.

#### **1.6.4. Katı Değişmezlik**

Ölçme değişmezliğinin son aşaması, her bir maddenin açıklanamayan varyansının gruplar arasında aynı olup olmadığını test eder (Davidov vd., 2014; Kim ve Yoon, 2011; Meredith, 1993; Millsap, 2011; Vandenberg ve Lance, 2000). Skalar değişmezlikte, faktör yükleri ve gösterge sabiti eşitliğine ek olarak hata varyanslarının eşitliği de şarttır (Vandenberg ve Lance, 2000). Ancak literatürde katı değişmezlik seviyesinin zorunlu tutulup tutulmaması konusunda görüş ayrılıkları bulunmaktadır (Millsap ve Olivera-Aguilar, 2012). De Shon (2004) her göstergenin maddeye özgü varyansının gruplar arasında eşdeğer olması gerektiği ve bu koşul sağlanmadığında ortalama farklarının karmaşıktırabileceğini öne sürerek katı değişmezlik düzeyinin uygulanmasını önermektedir. Öte yandan, katı değişmezliğinin aşırı kısıtlayıcı olduğu ileri sürülmekte (Bentler, 1990; Byrne, 1988) ve gerçek verilerle sağlanmasının güç olduğu belirtilmektedir (Little, 2013).

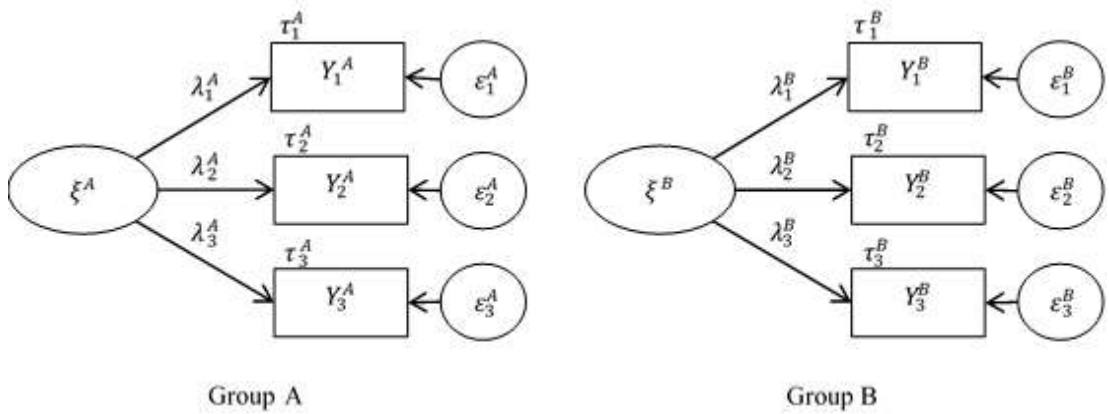
Çoğu araştırmada rastgele hata varyansının gruplar arasında eşdeğer olmasını beklemek mantıksız olduğu çünkü bu varyansın ölçme aracından bağımsız bir şekilde

duruma özgü olduğunu belirtmektedir. Literatürde genel olarak katı değişmezlik seviyesine ulaşmanın zorunlu olmadığı konusunda bir fikir birliği vardır (Brown, 2015; Davidov vd., 2015). Ölçme değişmezliği açısından, skalar değişmezlik seviyesine ulaşmak, yani faktör yüklerini ve gösterge sabitleri gruplar arasında değişmez olduğunu göstermek, ölçek puanlarını geçerli şekilde karşılaştırmak için yeterli görülmektedir. Literatür incelendiğinde katı değişmezlik, kesin değişmezlik olarak da isimlendirilmektedir. ÇGDFA ve aşamaları kavramsal olarak ayrıntılı biçimde açıklandıktan sonra, modelin daha iyi anlaşılabilmesi için matematiksel temeller çerçevesinde cebirsel olarak tanımlanmıştır.

Meredith'e (1993) göre ÇGDFA'yı cebirsel olarak ifade etmek gerekirse, model şu şekilde tanımlanır: Birey  $i$ , madde  $j$  ve grup  $g$  olarak tanımlandığında, gösterge değişkeni  $y_{ij}$ 'ye verilen yanıt, bir veya daha fazla gizil değişken  $\xi_i$ 'nin fonksiyonu biçiminde ifade edilmektedir. Gösterge sabiti  $\tau_j$ , fonksiyonun başlangıç değerini; faktör yükü  $\lambda_j$  ise gizil değişken  $\xi_j$  ile gösterge değişkeni  $y_{ij}$  arasındaki ilişkinin gücünü ifade etmektedir. Modelin cebirsel gösterimi aşağıda sunulmaktadır.

$$y_{ij}^g = \tau_j^g + \lambda_j^g \xi_i^g + \varepsilon_{ij}^g$$

Bu modeldeki ölçme parametreleri gruplar arasında farklı olabilir. İki gruplu ve tek bir örtük değişken içeren ÇGDFA modeli grafiksel olarak Şekil 1.1.'de sunulmaktadır.



**Şekil 1.1.** Çoklu Grup Doğrulamalı Faktör Analizi Grafiksel Gösterimi

Burada,  $\xi$  örtük değişkeni;  $\lambda$  faktör yükünü;  $\tau$  gösterge sabitini;  $Y$  gösterge değişkenini ve  $\varepsilon$  ölçme hatasını temsil etmektedir.

Yapısal değişmezlik, yukarıda belirtildiği gibi faktör yapılarının gruplar arasında eşit olmasını gerektirir. Yani gizil değişkenleri ölçmek için aynı maddelerin kullanılması gerekir. Bunu cebirsel olarak ifade edecek olursak;

$\forall g, h \in 1 \dots G$  için  $\lambda_j^g \approx 0$  (yaklaşık sıfır) ise  $\lambda_j^h \approx 0$  olmalıdır. Aynı şekilde  $\forall g, h \in 1 \dots G$  için  $\lambda_j^g \neq 0$  (yaklaşık sıfır değil) ise  $\lambda_j^h \neq 0$  olmalıdır. Burada  $\lambda_j^g$  yaklaşık sıfır olması  $y_1, y_2, y_3$  göstergelerin örtük değişken  $\xi$  arasındaki ilişkinin zayıf olduğunu göstermektedir. Tersisi ise göstergeler ile örtük değişken arasında güçlü bir ilişki söz konusu olur. Metrik değişmezlik bu yapısal eşitlik durumuna ek olarak faktör yüklerinin de gruplar arasında eşit olmasını gerektirir. Cebirsel olarak,

$$\forall g, h \in 1 \dots G \text{ için } \lambda_j^g = \lambda_j^h$$

şekilde ifade edilir. Skalar değişmezlik, metrik değişmezlik kısıtlamalarına ek olarak gösterge sabitlerinin eşit olmasını gerektirir. Cebirsel olarak,

$$\forall g, h \in 1 \dots G \text{ için } \tau_j^g = \tau_j^h$$

şeklinde ifade edilir. Katı değişmezlik, skalar değişmezlik kısıtlamalarına ek olarak hata varyanslarının da eşit olmasını gerektirir. Cebirsel olarak,

$$\forall g, h \in 1 \dots G \text{ için } \varepsilon_{ij}^g = \varepsilon_{ij}^h$$

şeklinde ifade edilir. ÇGDFA'nın cebirsel olarak tanımlanmasının ardından, yukarıda açıklanan ve belirtilen hususlar, bütüncül bir biçimde Tablo 1.1.'de sunulacak şekilde aşağıda gösterilmektedir.

**Tablo 1.1.** Ölçme Değişmezliği Aşamaları

Değişmezlik Türü	Hipotez	Cebirsel İfade
Yapısal Değişmezlik	Ölçülen modelde aynı sayıda faktör ve gözlenen değişkenlerde aynı faktöre yüklenmelidir.	$\forall g, h \in 1 \dots G$ alt gruplar için $\lambda_j^g \approx 0 \Rightarrow \lambda_j^h \approx 0$ olmalıdır. Tersisi de sağlanmalıdır.
Metrik Değişmezlik	Yapısal değişmezlik modelin kısıtlarına ek olarak faktör yüklerin de eşitliği gerekir.	$\forall g, h \in 1 \dots G$ alt gruplar için gruplar $\lambda_j^g = \lambda_j^h$ olmalıdır.
Skalar Değişmezlik	Metrik değişmezlik model kısıtlarına ek olarak gösterge sabitlerinin de eşitliği gerekir.	$\forall g, h \in 1 \dots G$ alt gruplar için $\tau_j^g = \tau_j^h$ olmalıdır.
Katı Değişmezlik	Skalar değişmezlik model kısıtlarına ek olarak hata varyanslarının da eşitliği gerekir.	$\forall g, h \in 1 \dots G$ alt grupları için $\varepsilon_{ij}^g = \varepsilon_{ij}^h$ olmalıdır.

### 1.7. Frekansçı Uyum İndeksleri

Kline (2016) tarafından belirtildiği üzere, literatürde model uyumunu değerlendirmek amacıyla geliştirilen uyum ölçütleri genel olarak dört temel sınıfta toplanabilir. Şen (2023) ise model uyumunu değerlendirmede kullanılan uyum indeks sınıflarını aşağıdaki şekilde tanımlamaktadır: Mutlak Uyum (Absolute Fit), modelin gözlemlenen verileri ne ölçüde iyi temsil ettiğini ve mükemmel uyumdan ne kadar uzak olduğunu gösteren, indeks değeri arttıkça uyumun kötüleştiğini ifade eden ve bu yönüyle kötülük uyum ölçütleri arasında yer alan bir değerlendirme aracıdır. Artırmalı Uyum (Incremental Fit), tahmini modelin göreceli olarak temel modele ne kadar uyduğunu  $\chi^2$  karşılaştırmasıyla ortaya koyar. Bu temel model, gözlenen tüm değişkenlerin ilişkisiz olduğu ve yalnızca tek bir gösterge ölçeğine sahip bulunduğu varsayımına dayanır. Tutumluluk Düzeltmesi (Parsimony Correction), modelin tutumlu olup olmadığını ve daha az belirgin tahmin yolları üzerinden nasıl geliştirilebileceğini ortaya koymaktadır. Yordayıcı Uyum İndeksleri (Predictive Fit Indexes), modelin aynı popülasyondan rastgele seçilen ve orijinal örnekleme aynı büyüklükteki varsayımsal tekrar örneklemlerine uyum gösterip göstermediğini tahmin eder. Uyum indekslerinin sınıflandırılmasına dair bilgilere Tablo 1.2.'de yer verilmektedir.

**Tablo 1.2.** Uyum İndeksleri Sınıfı

Mutlak Uyum	$\chi^2$ , $\chi^2/df$ , Root Mean Square Error Approximation (RMSEA), Bayes Information Criteria (BIC), Akaike's Information Criteria (AIC), Expected Cross-Validation Index (ECVI), Root Mean Square Residual (RMR), Standardized Root Mean Square Residual (SRMR), Goodness Of Fit Index (GFI) ve Adjusted Goodness Of Fit Index (AGFI)
Artırmalı Uyum	Comparative Fit Index (CFI), Tucker Lewis İndeks (TLI), Normed Fit Index (NFI), Relative Non-Centrality İndeks (RNI) ve Incremental Fit İndeks (IFI)
Tutumluluk Düzeltmesi	PGFI (GFI'nin düzeltilmiş versiyonu), PNFI (NFI'nin düzeltilmiş versiyonu), PNFI2 (IFI'nin düzeltilmiş versiyonu), PCFI (CFI'nin düzeltilmiş versiyonu), PR (Parsonomy ratio - tutumluluk oranı)
Yordayıcı Uyum İndeksi	CAIC (Consistent AIC), SABIC (Sample size Adjusted BIC), BIC (Bayes Information Criteria), AIC (Akaike's Information Criteria)

**Kaynak:** Şen, S. (2023). *Mplus ile yapısal eşitlik modellemesi uygulamaları* (2. basım). Nobel Akademik Yayıncılık.

Tablo 1.2.'den anlaşılacağı üzere bazı indeksler aynı anda farklı sınıflarda (AIC ve BIC gibi) yer alabilirler. Literatürde çok sayıda uyum indeksi geliştirilmiş olmakla birlikte, hangilerinin raporlanması gerektiği konusunda araştırmacılar arasında bir fikir birliği bulunmamaktadır. Şen (2023) göre bazı uygulamalardan kaçınılması gerekmektedir. Bunlar; analizde kullanılan yazılımların çıktı ekranlarında yer alan tüm indekslerin raporlanması, yalnızca en iyi uyumu gösteren indeksin seçilmesi ya da aynı sınıfa ait dört veya beş indeksin birlikte sunulması gibi durumlardır. Model değerlendirmesinde genel olarak önerilen yaklaşım;  $\chi^2$ , RMSEA, CFI ve SRMR değerlerinin raporlanmasıdır (Kline, 2016). Şen'e (2023) göre ise, mutlak ve artırılmış uyum indeksleri sınıfından en az bir indeksin raporlanması gerekmektedir. Bu doğrultuda araştırmada  $\chi^2$ , RMSEA, CFI, TLI ve SRMR değerleri raporlanmaktadır. Raporlanacak uyum indeksleri, aşağıda ayrıntılı biçimde açıklanmaktadır.

### 1.7.1. Ki-Kare Uyum Testi (Chi-Square)

$\chi^2$  testi, beklenen değer (yani teorik olarak ne olmalıydı) ile gözlenen değerleri (ampirik verileri) karşılaştırarak, aralarında istatistiksel olarak anlamlı bir fark olup olmadığını sorgular (Bentler, 1990; Meredith, 1993; Şen, 2023). Örneğin, adil olduğu varsayılan bir madeni paranın 100 kez atılması durumunda beklenen değer her iki yüz için de eşittir (50 yazı, 50 tura). Eğer gözlenen değerler (54 yazı, 46 tura) gibi beklenen değerlere yakınsa, elde edilen  $p$  değeri  $p > 0,05$  olur ve  $H_0$  hipotezi reddedilmez. Bu durumda, gözlenen veriler ile beklenen veriler arasında anlamlı fark bulunmadığı sonucuna ulaşılır. Ancak gözlenen değerler (yazı: 65, tura:35) gibi beklenen değere yakın değilse elde edilen  $p$  değeri  $p < 0,05$  olur ve  $H_0$  hipotezi reddedilir. Dolayısıyla, gözlenen veriler ile beklenen veriler arasında anlamlı bir fark olduğu sonucuna ulaşılır.

Bentler (1990)  $\chi^2$  testini, modelin öngördüğü kovaryans matrisi ile gözlenen kovaryans matrisi arasındaki uyumsuzluk farkını, örneklem büyüklüğünün bir eksiğiyle çarpımı olarak açıklamıştır. Bu açıklamadan  $\chi^2$  testinin örneklem ve korelasyon büyüklüğüne karşı duyarlı olduğu anlaşılmaktadır. Ayrıca  $\chi^2$  normallik varsayımında bulunur ve veri normallikten uzaklaştıkça yanlış sonuçlar verir (Şen, 2023).  $\chi^2$  testinin bu dezavantajlı durumları göz önünde bulundurduğunda bu teste

alternatif oluşturacak birçok indeks geliştirilmiştir. Bunlardan biri de  $\chi^2$  değerinin serbestlik derecesine oranıdır.  $\chi^2/sd$  oranı,  $\chi^2$ 'nin örneklem büyüklüklerinden etkilenme durumunu göze alarak Wheaton vd. (1977) tarafından önerilmiş bir orandır. Tabachnick ve Fidell'e (2007) göre  $2 \leq \chi^2/sd \leq 5$  değeri kabul edilebilir olarak görülmekte iken,  $0 \leq \chi^2/sd \leq 2$  değeri iyi uyum göstergesi olarak değerlendirilmektedir. Ancak  $\chi^2/sd$  model değerlendirmesinde kullanılıyor olsa da bu yöntemin hangi değerlerin iyi uyuma, hangi değerinin kötü uyuma işaret ettiği üzerine literatürde ortak görüş bulunmamaktadır (Şen, 2023).

### 1.7.2. Yaklaşık Hataların Ortalama Karekökü (Root Mean Square Error of Approximation-RMSEA)

Modelin veriye ne kadar yaklaşık uyduğunu gösterip, her bir serbestlik derecesi ( $df$ ) başına kalan hata miktarını ölçer. Modelde bir miktar hata olsa bile, bu hata kabul edilebilir mi? sorusuna cevap verir. Steiger ve Lind (1980) tarafından geliştirilmiş olan bu indeks, model için hesaplanan hata büyüklüğünü test eder. Hatanın derecesine göre bu indeks 0 ile 1 aralığında değer almakta ve değerın sıfıra yaklaşması ise mükemmel uyumu göstermektedir (Çokluk vd., 2021). Hu ve Bentler (1999) bu indeks değeri için 0,06 ve altındaki değerlerin iyi uyuma işaret ettiği; Brown ve Cudeck (1993) 0,05 ve altındaki değerler mükemmel uyuma, 0,10 ve üzeri değerlerin zayıf uyuma işaret ettiği; Maccallum vd. (1996) ise 0,08 altı iyi uyum gösteren ve 0,08 ile 0,10 arasında da kabul edilebilir olarak belirlemiştir. RMSEA'nın en önemli özelliği, bilinen bir dağılıma sahip olması ve bu nedenle Güven Aralığı'nın (GA) hesaplanabilmesidir. GA, nokta tahmininin belirsizliğini gösterir ve model uyumunun yalnızca tek bir değer üzerinden değil, olası değer aralığı üzerinden değerlendirmesine imkân tanır. Bu nedenle RMSEA değerlerinin raporlanması sırasında GA'nın da belirtilmesi metodolojik açıdan önemle tavsiye edilmektedir (Şen, 2023). RMSEA hesaplama formülü;

$$RMSEA = \sqrt{\frac{\chi^2 - df}{df \times (N - 1)}}$$

şeklinindedir. Eğer  $\chi^2 < df$  ise, RMSEA genellikle sıfır kabul edilir (negatif bir kök anlamlı olmadığı için). Burada  $\chi^2$ , modelin uyum istatistiği;  $df$ , serbestlik derecesi ve  $N$  ise örneklem büyüklüğünü temsil etmektedir. Formülden de anlaşılacağı üzere

RMSEA indeksi, model karmaşıklığına (daha fazla parametre içermesi), örneklem büyüklüğüne ( $\chi^2$  gibi büyük örneklemden etkilenmeyeceği ancak küçük örneklemelere duyarlı olabileceği) ve serbestlik derecesinden etkilendiği söylenebilir.

### 1.7.3. Standartlaştırılmış Artık Ortalamaların Karekökü (Standardized Root Mean Square Residuals-SRMR)

Standartlaştırılmamış bir istatistik olan Artık Ortalamaların Karekökü (Root Mean Square Residuals-RMR) indeksi farklı kategori sayısına sahip maddeler içeren ölçeklerde yorumlamayı zorlaştırmaktadır. Alternatif olarak standartlaştırılmış RMR (SRMR) kullanılması tavsiye edilir (Chen, 2007; Hu ve Bentler, 1999; Şen, 2023). SRMR, modelin tahmin ettiği kovaryanslarla gözlemlenen kovaryanslar arasındaki farkların standartlaştırılmış karekök ortalamasını ifade eder. Başka bir deyişle, korelasyon artıklarının mutlak değerlerinin ortalamasının yansıtan bir uyum indeksidir (Kline, 2011). Standartlaştırılmış bir ölçüt olması nedeniyle 0 ile 1 arasında değişir ve değer sıfıra yaklaşması, modelin veriyle iyi düzeyde uyum sağladığını gösterir (Tabachnick ve Fidell, 2001). SRMR değeri ne kadar küçükse, modelin uyumu o denli iyidir. Literatürde, 0,08'in altındaki SRMR değerlerinin kabul edilebilir uyumu, 0,05'in altındaki değerlerin ise mükemmel uyumu temsil ettiği belirtilmektedir (Byrne, 2013; Hu ve Bentler, 1999). SRMR hesaplama formülü;

$$SRMR = \sqrt{\frac{1}{P(P+1)/2} \sum_{i=1}^P \sum_{j=1}^i (r_{ij}^{obs} - r_{ij}^{model})^2}$$

şeklindedir. Burada  $P$ , modeldeki gözlenen değişken sayısı;  $r_{ij}^{obs}$ , gözlenen değişkenler arasındaki standartlaştırılmış korelasyon (veya kovaryans);  $r_{ij}^{model}$ , model tarafından tahmin edilen korelasyon (veya kovaryans) ve  $\sum$ , tüm değişken çiftleri arasındaki farkların toplamını temsil etmektedir.

### 1.7.4. Karşılaştırmalı Uyum İndeksi (Comparative fit index-CFI)

CFI, modelin veriye ne kadar iyi uyum sağladığını gösteren karşılaştırmalı bir uyum indeksidir. Temel amacı önerilen modelin, veriyle hiç uyum sağlamayan boş (null) modele veya bağımsız (independence) modele kıyasla ne kadar daha iyi performans gösterdiğini ölçmektir. Diğer uyum ölçütleri ( $\chi^2$  ve RMSEA gibi) ile karşılaştırıldığında örneklem büyüklüğünden daha az etkilenmektedir. CFI, 0 ile 1

arasında değer almakta ve değerlerin bire yaklaşması modelin veriyle iyi uyum sağladığı belirtilmektedir (Hu ve Bentler, 1995). Hu ve Bentler (1999), CFI değerinin 0,90 ile 0,95 arasında olmasını kabul edilebilir, 0,95 ile 1 arasında olmasını ise iyi bir uyumun göstergesi olarak değerlendirmiştir. CFI hesaplama formülü,

$$CFI = 1 - \frac{\chi_{model}^2 - df_{model}}{\chi_{null}^2 - df_{null}}$$

şeklindedir. Burada,  $\chi_{model}^2$ , önerilen modelin ki-kare değeri;  $df_{model}$ , önerilerin modelin serbestlik derecesi;  $\chi_{null}^2$ , boş modelin ki-kare değeri;  $df_{null}$ , boş modelin serbestlik derecesini temsil etmektedir.

#### 1.7.5. Normlaştırılmamış Uyum İndeksi (Nonnormed fit index-NNF=Tucker Lewis Index-TLI):

Bentler ve Bonett (1980) normlaştırılmamış uyum indeksini (NNFI) geliştirerek bu kavramı kovaryans yapısı analizine genellemiş ve daha sonra Tucker-Lewis İndeksi (TLI) olarak da anılmaya başlanmıştır. Başlangıçta faktör analizi için tasarlanan TLI, zamanla YEM’de kullanılmaya başlanmıştır (Lomax, 2004). TLI, modelin karmaşıklığını dikkate alarak, önerilen model ile bağımsız modelin karşılaştırmasını yapar. TLI hesaplama formülü;

$$TLI = \frac{\frac{\chi_{null}^2}{df_{null}} - \frac{\chi_{model}^2}{df_{model}}}{\frac{\chi_{null}^2}{df_{null}} - 1}$$

şeklindedir. Burada  $\chi_{null}^2$ , boş modelin ki-kare değeri;  $df_{null}$ , boş modelin serbestlik derecesi;  $\chi_{model}^2$ , önerilen modelin ki-kare değeri;  $df_{model}$ , önerilen modelin serbestlik derecesi değerini temsil etmektedir. Formülde belirtilen bağımsız model, değişkenler arasında hiçbir ilişki olmadığını varsayar. Bu model genellikle kötü uyumu gösterir. Önerilen model, araştırmacının kurduğu modeli temsil eder. TLI ise bu iki modelin  $\chi^2$  değerlerini ve serbestlik derecelerini karşılaştırarak modelin iyileşme oranını ölçer. 0 ile 1 aralığında bir değer alır ve değerlerin bire yaklaşması mükemmel uyumu göstermektedir (Tabachnick ve Fidell, 2007). TLI, .95 ve üzerinde iyi uyum gösterirken 0,90 ile 0,95 arasındaki değerlerin kabul edilebilir uyumu sağladığı belirtilmektedir (Hu ve Bentler, 1999).

ÇGDFA, kısıtlanmış ve kısıtlanmamış modeller arasındaki uyum iyiliği indeksleri karşılaştırılarak ölçme değişmezliği test edilir (Cheung ve Rensvold, 2002). Literatürde, ölçme değişmezliğini değerlendirmek için farklı yaklaşımlar geliştirilmiş olup, bu yaklaşımlar genellikle  $\chi^2$ , CFI ve RMSEA değerlerindeki değişimlere odaklanmaktadır (Cheung ve Rensvold, 2002). Söz konusu değişimler sırasıyla  $\Delta\chi^2$ ,  $\Delta CFI$  ve  $\Delta RMSEA$  biçiminde ifade edilmektedir. Cheung ve Rensvold (2002) ile Chen (2007), büyük örneklemeler üzerinde yapılan analizlerde  $\chi^2$  testinin sınırlılıklarına dikkat çekerek, bunun yerine uyum indeksleri arasındaki farkların değerlendirilmesi önerilmektedir. Bu doğrultuda, araştırmamızda modeller arası geçişlerde genel model uyumu dikkate alınmış ve  $\Delta CFI$  ile  $\Delta RMSEA$  değerlerindeki farklılıklar üzerinden ölçme değişmezliği değerlendirilmiştir.

YEM'in temel ögesi parametrelerdir ve bu parametrelerin değerlerinin tahmin edilmesi parametre kestirimi olarak adlandırılır. Kestirimin amacı, bilinmeyen parametre değerlerini belirlemek ve modele dayalı kovaryans matrisi ile örnekleme dayalı kovaryans matrisini mümkün olduğunca birbirine yaklaştırmaktır. Bu süreçte, iki matris arasındaki farkı en aza indiren özel bir uyum fonksiyonu kullanılır. Uyum fonksiyonunun hesaplanmasına yönelik farklı yaklaşımlar bulunmakta olup, bunlar parametre kestirim yöntemleri olarak tanımlanmaktadır. YEM'de en yaygın kullanılan yöntem Maksimum Olabilirlik (ML) kestirimidir. Ancak ML yöntemi çok değişkenli normallik varsayımını gerektirmektedir. Bu varsayımın sağlanmadığı durumlarda, Yuan-Bentler tarafından geliştirilen Sağlam Maksimum Olabilirlik (MLR) yöntemi tercih edilmektedir (Şen, 2023). Veri sıralı veya kategorik yapıya sahip, normal dağılmayan ve en az dört kategori içerdiğinde MLR yöntemi önerilmektedir (Gana ve Broc, 2019). Bu araştırmada hem DFA hem de ÇGDFA analizlerinde kestirim yöntemi olarak MLR kullanılmıştır.

### 1.8. Bayes Teoremi

Bayes teoremi, adını İngiliz matematikçi ve istatistikçi olan Thomas Bayes'ten (1702-1761) almıştır ve ilk kez, ölümünden sonra arkadaşı Richard Price'in 1763 yılında yayınladığı Şans Öğretisinde Bir Problemin Çözümüne Yönelik Bir Makale (An Essay Towards Solving a Problem in the Doctrine of Chances) adlı çalışmayla ortaya konmuştur (Stigler, 1983). Bu teorem, koşullu ve genel olasılıklar arasındaki ilişkiyi ele alarak, evrensel doğrular ve gözlemler aracılığıyla çeşitli durumları

modellemeyi amaçlamaktadır (Çelebi, 2019). Bayes yaklaşımının temelinde ise elde edilen yeni bilgiler doğrultusunda olasılıkların yeniden hesaplanarak kararların güncellenmesi; yani ilk tahminin ek bilgiler ışığında yeniden değerlendirilmesi gerektiği düşüncesi yer alır (Mlodinov, 2009).

Bayes teoremi, matematiksel olarak koşullu olasılık tanımıyla açıklanır. Cebirsel olarak ifade edecek olursak; Bir örneklem uzayında  $A$  ve  $B$  iki olay olmak üzere,  $A$ 'nın  $B$  koşuluyla gerçekleşme (yani  $B$  olayı verilmişken,  $A$  olayının gerçekleşme) olasılığı ve  $B$ 'nin  $A$  koşuluyla gerçekleşme olasılıkları sırasıyla aşağıdaki gibi tanımlanır.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0 \text{ ve } P(B|A) = \frac{P(A \cap B)}{P(A)}, P(A) > 0$$

Bu iki ifadeyi birbirine eşitlersek,

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

elde ederiz. Daha sonra her iki tarafı  $P(A)$ 'ya bölersek Bayes teoremine ulaşırız (Kim ve Bolt, 2007).

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Eğer  $A$  olayını etkileyebilecek birkaç farklı olay varsa (örneğin  $B_1, B_2, B_3, B_4 \dots B_n$ ), o zaman  $P(A)$  şöyle bulunur:

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i)$$

Bunu önceki denklemde yerine yazarsak Bayes Teoremi'nin genel halini elde ederiz.

$$P(B_j|A) = \frac{P(A|B_j) \cdot P(B_j)}{\sum_{i=1}^n P(A|B_i) \cdot P(B_i)}$$

Burada  $P(B_j)$ , hipotezin önsel (prior) olasılığı;  $P(A|B_j)$ , hipotez doğru ise gözlemi elde etme olasılığı (likelihood);  $P(B_j|A)$ , gözlem yapıldıktan sonra hipotezin güncel olasılığıdır (posterior) (Kim ve Bolt, 2007).

Bayes teoremini somutlaştırmak amacıyla, Mlodinow'un sunduğu koşullu olasılık örneği incelenebilir. Bir kişinin iki çocuğu olduğu bilinsin. Ayrıca bu kişi, "çocuklarımdan en az biri erkektir." bilgisini vermektedir. Bu durumda sorumuz, her iki çocuğunun da erkek olma olasılığının ne olduğudur. Olası cinsiyet kombinasyonları (ilk çocuk, ikinci çocuk) şu şekildedir: (Erkek, Erkek); (Erkek, Kız); (Kız, Erkek);

(K1Z, K1Z). Eğer “en az biri erkek” bilgisi dikkate alınmazsa, her iki çocuğun da erkek olma olasılığı  $P(\text{Erkek}, \text{Erkek}) = 1/4$  olur. Ancak verilen bilgi doğrultusunda (K1Z, K1Z) seçeneği örneklem uzayından çıkarılır. Geriye kalan olasılıklar (Erkek, Erkek); (Erkek, K1Z); (K1Z, Erkek) olup, bu üç durum eşit olasılıklıdır. Dolayısıyla her iki çocuğunun da erkek olma olasılığı:  $P(\text{Erkek}, \text{Erkek}) = 1/3$  olur. Çoğu zaman “en az biri erkekse, diğerinin de erkek olma olasılığı  $\frac{1}{2}$ ’dir.” şeklinde yanlış bir çıkarım yapılmaktadır. Oysa yukarıda belirtildiği üzere, elde edilen bilgiler olasılık uzayını değiştirmekte ve bu durum Bayes mantığının temelini oluşturmaktadır. Özetle, Bayes yaklaşımına göre yeni bilgi geldikçe olasılıklar güncellenmektedir.

### 1.9. Bayes İstatistikleri

Bayes istatistiklerinde üç temel bileşen bulunmaktadır: Ön dağılım (prior), olabilirlik fonksiyonu (likelihood) ve sonsal dağılım (posterior). Ön dağılım, modeldeki parametrelerle ilgili mevcut bilgi ya da varsayımları ifade eder. Olabilirlik fonksiyonu, gözlemlenen verilerden gelen bilgileri yansıtarak model parametrelerine dair olasılık dağılımı oluşturur. Sonsal dağılım ise veriler dikkate alındıktan sonra elde edilen güncellenmiş bilgi düzeyini yansıtır. Bu dağılım, ön dağılım ile olabilirlik fonksiyonu arasında bir denge kurar. Ön dağılım ne kadar bilgilendirici olursa, sonsal dağılım da o kadar ona benzeyecektir.

Bayes istatistiğinde, parametrelerin hangi değerlere sahip olabileceği konusundaki belirsizlik, ön dağılımlar aracılığıyla ifade edilir. Bu yaklaşım, bilinmeyen parametrelerin sabit bir değere sahip olduğu varsayımını reddeder. Bunun yerine, bu parametrelerin olası değerleri bir olasılık dağılımı olarak ele alır. Ön dağılım, veri gözlemlenmeden önceki inançlarımızı yansıtır (Lambert, 2018). Eğer araştırmacının güçlü ön inançları varsa, bu genellikle dar ve bilgilendirici bir ön dağılıma dönüştürülebilir. Öte yandan, belirsiz ya da zayıf inançlar söz konusu ise geniş ve yaygın bir ön dağılım tercih edilir. Bu sayede, frekansçı modelde olduğu gibi deneyin sonsuz kez tekrarlanması varsayımına güvenmek yerine, yeni bilgileri gelecekteki araştırmalara entegre edilebilir (Van de Schoot ve Depaoli, 2014). Ön dağılım, bilgilendirici ön dağılım (geçmiş çalışmalar, önceki deneyimler ya da uzman görüşlerinden toplanan bilgiler gibi) ve bilgilendirici olmayan ön dağılım (geniş varyansa sahip ve geniş, düz ve yaygın bir dağılım, sonsal dağılımlar üzerinde daha az

etkili, frekansçı istatistiklere daha yakın (Asparouhov ve Muthén, 2010) olmak üzere iki tür dağılım vardır.

Bayes modellerde, ön dağılımların belirlenmesi süreci, özellikle Markov Chain Monte Carlo (MCMC) yöntemi ile belirlenmesi kritik öneme sahiptir. Özellikle YEM gibi karmaşık yapılarda, analitik olarak ön dağılımların çıkarılması çoğu zaman mümkün değil ve doğrudan tahmin edilemezler. Bu gibi durumlarda, Bayes çıkarım süreci, MCMC örnekleme algoritmaları kullanılarak yürütülür (Hastings, 1970). Gibbs örnekleme ve Metropolis-Hastings gibi MCMC teknikleri, parametrelerin çok değişkenli sonsal dağılımlarından yinelenmeli olarak örnekler alır ve her parametre için zinciri sırayla günceller. Bu işlem, sonsal dağılımın yaklaşık olarak elde edilene kadar devam eder. MCMC yöntemlerinin en önemli avantajlarından biri, sonsal dağılımı doğrudan üretebilmesi ve bu nedenle normal dağılım gibi klasik, büyük örneklem varsayımlarına ihtiyaç duymamasıdır. Bu özellik sayesinde karmaşık modellerde bile daha esnek ve veri odaklı çıkarımlar gerçekleştirebilir.

Bayes teoreminin üç bileşeninin (önsel, olasılık ve sonsal) nasıl işlediğini anlamak için, ilgi duyulan parametre kümesini  $\theta$ , gözlemlenen veriyi  $y$  ile gösterelim. Verinin,  $\theta$  verildiğinde gözlemlenme olasılığı  $f(y|\theta)$  ile,  $\theta$  hakkında verilerden önceki bilgi ise önsel yoğunluk  $f(\theta)$  ile tanımlanır. Teorem, sürekli veriler için aşağıdaki şekilde ifade edilir:

$$\begin{aligned} f(\theta | y) &= (f(\theta)f(y | \theta))/f(y), \\ f(\theta|y) &\propto f(y|\theta)f(\theta) \end{aligned} \quad (1.1)$$

Bu ifade,  $y$  verisi verildiğinde  $\theta$  parametresinin sonsal dağılımının,  $\theta$  verildiğinde  $y$ 'nin gerçekleşme olasılığı ile önsel dağılımın çarpımına orantılı olduğunu (yani  $\propto$  işaretiyle gösterildiği gibi) ifade eder (Brown, 2015). Ayrıca, (1.1) denkleminde,  $f(y|\theta)$  teriminin olasılık fonksiyonu olan  $L(\theta|y)$  biçimde yeniden yazarak denklem (1.2)'ye dönüştürülebilir.

$$f(y|\theta) \propto L(\theta|y)f(\theta) \quad (1.2)$$

Muthén ve Asparouhov (2012), Bayes teoreminin temel eşitliklerinden biri olan denklem (1.2)'yi şu şekilde açıklamıştır:

$$\begin{aligned} \text{sonsal dağılım} &= (\text{parametreler}|\text{veri}), \\ &= ([\text{veri}/\text{parametreler}]x \text{parametreler})/\text{veri}, \end{aligned}$$

$$= (\text{olasılık} \times \text{önsel dağılım}) / \text{veri},$$

$$\text{sonsal dağılım} \propto \text{olasılık} \times \text{önsel dağılım}. \quad (1.3)$$

(1.2) ve (1.3) denklemleri, Bayes istatistiğinin özünü temsil eder ve alandaki temel yapı taşları arasında yer alır (Brown, 2015; Muthén ve Asparouhov, 2012).

### 1.10. Bayes Yapısal Eşitlik Modellemesi

Bayes Yapısal Eşitlik Modellemesi (BYEM), frekansçı YEM uygulamalarına alternatif olarak geliştirilmiş ve bu yaklaşımlara kıyasla önemli avantajlar sunmaktadır. Kategorik veri, kayıp veri ve örneklem büyüklüğü gibi sınırlılıkların aşılmasına olanak tanıyan BYEM, normal dağılım varsayımına ihtiyaç duymaması sayesinde esneklik sağlamaktadır. Bu yönüyle BYEM, frekansçı YEM uygulamalarının varsayımlarını ortadan kaldırmaya yönelik bir yaklaşım olarak değerlendirilebilir (Kaplan ve Depaoli, 2012). Ayrıca BYEM, frekansçı YEM'den farklı olarak MCMC yöntemlerine dayanmaktadır (Muthén ve Asparouhov, 2012). MCMC, önceki örneklem gruplarını kullanılarak yeni örneklem grupları oluşturulmakta ve gizil değişken ile bilinmeyen parametreler için sonsal dağılımların elde edilmesine olanak tanımaktadır (Dunson vd., 2005). BYEM'in en önemli ayırt edici özelliği ise, model parametreleri için önsel bilgilerin tanımlanabilmesidir. Önsel bilgilerin kullanımı model hatalarını azaltarak daha iyi uyum elde edilmesine katkı sağlamakta; parametrelerin dağılımı, veriden önce önsel bilgi ile belirlenmekte ve ardından Bayes Teoremi aracılığıyla veri ile birleştirilmektedir (Alkış, 2016). Bu süreç sonucunda elde edilen sonsal dağılım, model parametrelerinin ön bilgisi ile veriden gelen bilginin bütünleşmiş halini yansıtmaktadır (Arbuckle, 2012).

Bayes yaklaşımında önsel bilgi önemlidir. Çünkü model çıkarımı model parametrelerinin sonsal dağılımına ve sonsal dağılım ise önsel dağılımlara dayanmaktadır (Alkış, 2016). Daha önce vurgulandığı gibi önsel bilgiler, bilgilendirici olmayan önsel dağılımlar ve bilgilendirici önsel dağılım olmak üzere ikiye ayrılmaktadır. Bilgilendirici olmayan önsel dağılımlar, parametreler hakkında önsel bilginin sınırlı olduğu veya yalnızca veriden elde edilen bilginin yeterli görüldüğü durumlarda kullanılır; bu tür önseller sonsal dağılım üzerinde çok az etkiye sahiptir ve bu nedenle Bayes tahminler ile frekansçı tahminler arasında büyük farklılıklar yaratmaz. Buna karşılık bilgilendirici önsel dağılımlar, parametreler hakkında belirli

bilgileri içerir ve bu bilgileri hiperparametreler (Normal: ortalama, varyans; Beta:  $\alpha, \beta$ ; Inverse-Wishart:  $R, df$ ) aracılığıyla ifade eder ve modelin sonuçlarını daha güçlü biçimde etkiler. Bayes literatüründe en yaygın kullanılan bilgilendirici önsel türlerinden biri eşlenik önseldir (Normal-Normal, Binom-Beta, çok değişkenli normal-Inverse Wishart). Eşlenik önsel kullanıldığında, sonsal dağılımın matematiksel biçimi korunur, yani dağılım türü değişmez ve sadece parametre değerleri güncellenir (Murat, 2012). BYEM analizleri, çeşitli istatistiksel yazılımlar aracılığıyla gerçekleştirilmektedir. Bu kapsamda, Mplus (Muthén ve Muthén, 2017), OpenBUGS (Spiegelhalter vd., 2014), WinBUGS (Spiegelhalter vd., 2003) ve R yazılımı için geliştirilen blavaan paketi (Merkle ve Rosseel, 2018) gibi uygulamalardan gerçekleştirilebilir (Helvacı, 2025).

### 1.11. Bayes Yaklaşık Ölçme Değişmezliği

Gruplar arasında herhangi bir ölçme parametresi üzerinden yapılan ikili karşılaştırmaların sayısı, grup sayısındaki artışla birlikte katlanarak artmaktadır. Bu durum, yapılan karşılaştırma sayısının artması nedeniyle ölçme değişmezliğinin hatalı biçimde reddedilme olasılığını da artırmaktadır (Rutkowski ve Svetina, 2014). Nitekim, çok sayıda grup üzerinde ölçme değişmezliğini test etmek, metodolojik açıdan ciddi zorluklar içermektedir (Kim vd., 2017). Bu bağlamda, bazı araştırmacılar (Davidov vd., 2014; Davidov vd., 2015; Muthén ve Asparouhov, 2013; Van de Schoot vd., 2013), ÇGDFA yönteminin çok sayıda grup arasında ölçme değişmezliği testinde kullanılmasını iki temel gerekçeyle eleştirmiştir. İlk olarak, çok sayıda grup arasında çiftli karşılaştırmaların yapılması Tip I hata oranını artırmakta, yani ölçme değişmezliğini yanlış bir şekilde değerlendirebilir (Kim vd., 2017). İkinci olarak ise, ÇGDFA'da kullanılan uyum indeksleri model değerlendirmesinde ortaya çıkabilir. Çünkü, uyum indeksleri model değerlendirmesinde esas olarak iki grup karşılaştırılması için kullanılması olabilir (Asparouhov ve Muthén, 2014; Kim vd., 2017). Dahası, bu yaklaşımın bir diğer dezavantajı, gruplar arasındaki parametrelerin tam anlamıyla eşitliğini şart koşmasıdır ki bu da yöntemi oldukça katı hale getirmektedir (Davidov vd., 2015). Oysa gerçek veri analizlerinde ölçme parametrelerinin mutlak eşitliği çoğunlukla sağlanamamaktadır. Özellikle örneklem büyüklüklerinin oldukça büyük olduğu uluslararası ölçek araştırmalarında, gruplar arasında önemsiz düzeydeki ölçme farklılıkları bile ölçme değişmezliğinin

sağlanamadığı yönünde hatalı bir sonuca neden olabilir (Meuleman, 2012; Oberski, 2014). Tüm bu metodolojik sınırlamalar göz önünde bulundurulduğunda, çok sayıda grupta ölçme değişmezliğini test etmeye olanak tanıyan alternatif yaklaşımlara ihtiyaç duyulmaktadır.

BYÖD yöntemi ilk kez Muthén ve Asparouhov (2012), ardından Van de Schoot vd. (2013) tarafından literatüre kazandırılmıştır. Bu yaklaşım, kesin ölçme değişmezliği yöntemlerinden (örneğin ÇGDFA) ayrılarak, gruplar arasındaki ölçme parametrelerinin (faktör yükleri ve gösterge sabit değerleri) tamamen eşit (farkların sıfır) olduğunu varsayma zorunluluğunu ortadan kaldırır. BYÖD yöntemi, daha esnek bir modelleme imkânı sunarak sıfır ortalama ve küçük varyanslı bilgilendirici önsellerle, gruplar arasındaki ölçme parametreleri farklılıklarının sıfıra yakın yani önemsiz düzeyde olduğunu varsayarak, araştırmacıların modellerini daha esnek biçimde test etmelerine olanak tanır. Gruplar arasındaki bu küçük parametre farklılıklarının sonraki analiz sonuçları üzerinde anlamlı bir etkisinin olmadığı varsayılır (örneğin grup ortalamalarının karşılaştırılması gibi). Böylece BYÖD hem metodolojik hem uygulamalı açıdan daha esnek bir ölçme değişmezliği değerlendirme yaklaşımı olarak öne çıkmaktadır.

Kim vd. (2017) çalışmalarında, BYÖD testi iki aşamalı bir analiz süreci olarak tanımlamaktadır. İlk aşamada araştırmacılar, BYEM kullanarak değişmez olmayan parametrelerin belirlendiği; ikinci aşamada ise bu değişmez parametreleri serbest bırakılmasıdır. BYÖD'nin uygulanabilmesi için, sıfır olan farklar yerine bilgilendirici ve küçük varyansa sahip, normal dağılımı izleyen yaklaşık sıfır farklara dayalı alternatif parametre tanımlamalarına olanak tanır. Yani, gruplar arasında farklılık gösteren parametreler esnetilmektedir. Eğer BYÖD sonuçları gruplar arası ölçme parametrelerinde sıfıra yakın farklılıklar gösteriyorsa, gizil faktör ortalamalarının karşılaştırılması daha anlamlı hale gelmektedir (Muthén ve Asparouhov, 2013; Van de Schoot vd., 2013). Literatürde bu iki aşama farklı şekillerde uygulanmış ve üç farklı prosedür ortaya çıkmıştır: a) Metrik ve skalar değişmezliği aynı aşamada belirleyerek test etme, b) Sadece iki düzeyin (yapısal ve skalar değişmezliği) test edilmesi, c) Frekansçı ölçme değişmezliği test prosedürü.

BYÖD uygulamasına ilişkin ilk prosedürde, araştırmacılar metrik ve skalar seviyeleri aynı anda test eder. Bu aşamada gruplar arasındaki faktör yükleri ve madde sabitleri için bilgilendirici önsel değerler belirlenir. Ardından, yaklaşık olarak değişmezlik göstermeyen faktör yükleri ve madde sabitleri için yaklaşık kısıtlamalar serbest bırakılır (Muthén ve Asparouhov, 2013; Van de Schoot vd., 2013). Eğer ölçek çok sayıda madde içerdiğinde ve faktör yükleri ile madde sabitleri arasındaki farklar küçük olduğunda, bu farklılıklar gruplar arasında birbirini dengeler. Bu tür durumlarda birinci prosedür, BYÖD testleri açısından en uygun yöntem olarak değerlendirilmektedir (Muthén ve Asparouhov, 2013).

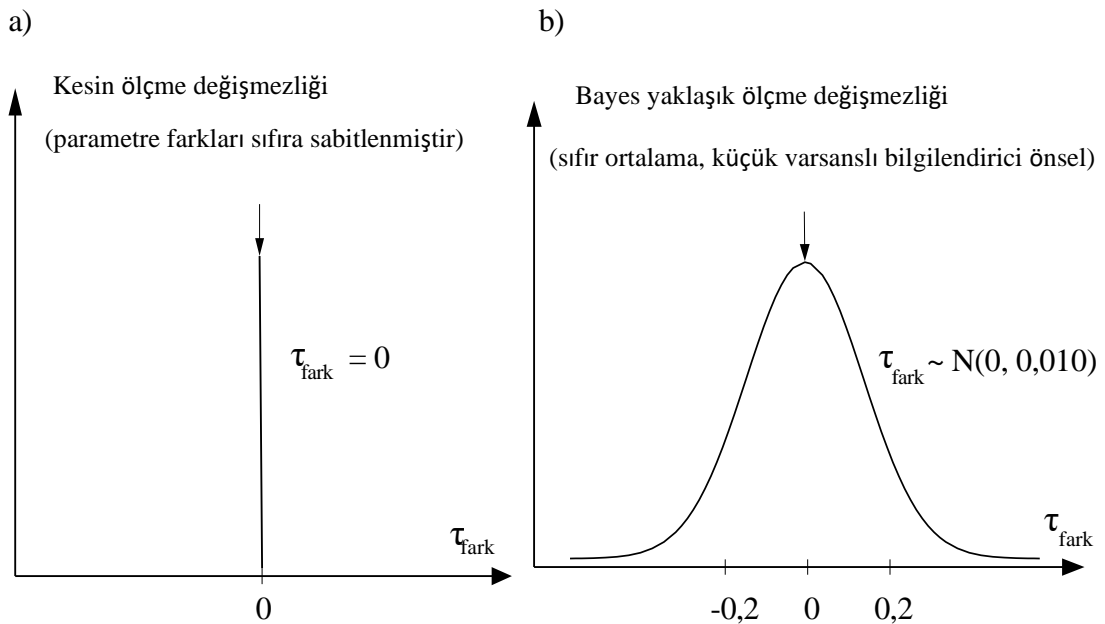
İkinci prosedür, yalnızca iki düzeyin test edilmesini gerektirir: yapısal değişmezlik ve ardından skalar değişmezlik. Bu yaklaşımda araştırmacılar, daha önce uyguladıkları frekansçı ölçme değişmezliği testlerinden elde ettikleri bulgulara dayanarak ölçeğin metrik ya da kısmi skalar değişmezlik gösterdiğini varsayar. Bu nedenle, BYÖD için faktör yüklemeleri ve kalıntı kovaryanslara yönelik bilgilendirici ve küçük varyansa sahip önselleri içeren BYEM kullanılarak yapısal değişmezlik test edilir. Eğer yapısal değişmezlik korunursa, skalar değişmezlik gruplar arasında faktör yükleri ve madde sabitleri için belirlenen bilgilendirici önsellerle test edilir (De Bond ve Van Petegem, 2015). Üçüncü prosedür, ML yerine Bayes tahmin edicisi kullanılarak frekansçı ölçme değişmezliği testine benzer şekilde dört aşamada gerçekleştirilir (Alamri, 2019; Kim vd., 2017).

Üçüncü prosedürün ilk aşamasında, yapısal düzeyde faktör yüklerini ve madde sabitlerini herhangi bir eşitlik kısıtlaması getirmeksizin ÇGDFA'yı Bayes tahminiyle test eder. Bu aşamada yazılımın varsayılan önselleri kullanılır. Örneğin, Mplus'ta faktör yükleri ve madde sabitleri için ortalaması sıfır ve çok yüksek varyansa sahip normal dağılımlar kullanılmaktadır. Model tanımlanırken tüm grupların faktör ortalamaları ve varyansları sırasıyla 0 ve 1 olarak sabitlenmekte, faktör yükleri ve madde sabitleri serbest biçimde tahmin edilmektedir. İkinci adımda ise faktör yükleri ve madde sabitleri arasındaki tutarsızlıkları tanımlamak amacıyla yaklaşık değişmezlik önsel varyansını belirlemektedir. Üçüncü adımda, belirlenen varyans kesme noktası dahil olmak üzere çeşitli önsel varyanslar kullanılarak bir dizi yaklaşık metrik ölçme değişmezliği modeli kurulmakta ve karşılaştırmalar sonucunda en iyi uyum gösteren model seçilmektedir. Seçilen modeldeki önsel varyansın, ikinci adımda

tanımlanan yaklaşık değişmezlik varyansına eşit veya daha küçük olması gerekir; aksi takdirde model reddedilir. Eğer BYÖD metriği geçerliyse, dördüncü adım uygulanır. Dördüncü adımda, üçüncü adımda uygulanan prosedür bu kez madde sabitleri için tekrarlanır. Yaklaşık skalar değişmezliğe ulaşmak için, seçilen modeldeki önsel varyansın önceden belirlenen madde sabitlerine ait yaklaşık değişmezlik varyansına eşit ya da daha düşük olması gerekir (Kim vd., 2017). Araştırmada üçüncü prosedür tercih edilmektedir.

Eğer önsel varyans küçükse, yani model parametreleri sıkıca sınırlandırılmışsa, sonuçlar kesin ölçme değişmezliği varsayımına daha yakın olur. Yani herkes için eşit kabul edilen bir yapı test edilmiş olur. Eğer önsel varyans büyükse, model parametrelerine daha fazla esneklik tanınır. Bu da sonuçların yapısal modele daha uygun hale gelmesini sağlar. Çünkü farklı gruplarda bazı küçük sapmalara izin verilmiş olur. Bu yaklaşım sayesinde, modeller kesin değişmezlik sağlamasa bile, küçük farklılıklarla yaklaşık ölçme değişmezliği varsayımını sürdürülebilir.

Kesin ve yaklaşık ölçme değişmezliğini daha iyi anlaşılabilmesi amacıyla Şekil 1.2. incelenmektedir.



**Şekil 1.2.** Kesin ve Yaklaşık Ölçme Değişmezliği

Kesin (a) ve Bayes yaklaşık (b) ölçme değişmezliği (Muthén ve Asparouhov, 2012)

## 1.12. Bayes Uyum İndeksleri

BYÖD model uyumu, BYEM analitik çerçevesi kapsamında mevcut olan çeşitli model değerlendirme stratejileriyle incelenir. Model uyumu, sonsal tahmin p-değeri (Posterior Predictive p-value; PPP) ve %95 güven aralığının (Confidence Interval; CI) değerlendirilmesi gibi yöntemlerle analiz edilmektedir (Gelman vd., 2004). Araştırmacılar, bu uyum göstergelerine ek olarak model karşılaştırmalarında Sapma Bilgisi Kriteri (Deviation Information Criterion; DIC) (Spiegelhalter vd., 2002), Bayes Bilgi Kriteri (Bayesian Information Criterion; BIC) (Schwarz, 1978) ve Bayes Faktörü (Alamri, 2019) gibi karşılaştırma indekslerinden yararlanabilirler. Bununla birlikte, Bayes modelleme yalnızca bu frekansçı uyum indeksleriyle sınırlı değildir; literatürde önerilen başka yerleşik kriterler de bulunmaktadır. Örneğin, Yaygın Olarak Uygulanabilir Bilgi Kriteri (Widely Applicable Information Criterion; WAIC: Vehtari vd., 2017); Birini Dışarıda Bırakma Çapraz Doğrulaması (Leave-One-Out Cross-Validation; LOO: Vehtari vd., 2017) ve Veri Uyumu Kriteri (Data Agreement Criterion; DAC: Lek ve Van de Schoot, 2019) model karşılaştırması ve uyum değerlendirmesi için önerilen alternatifler arasındadır (Vehtari vd., 2017). Son dönemde ise klasik uyum indekslerinin Bayes versiyonları geliştirilmiştir. Bayes RMSEA (BRMSEA), Bayes CFI (BCFI) ve Bayes TLI (BTLI) gibi göstergeler, gözlenen ve tekrarlanan uyumsuzluk fonksiyonları arasındaki farklara dayanarak hesaplanabilmektedir (Liang, 2020). Bu çalışmada, BYÖD model uyumu değerlendirmesinde PPP ve %95 güven aralığı kullanılmakta; model karşılaştırması için ise DIC tercih edilmektedir. Belirlenen bu indeksler sonraki bölümde ayrıntılı olarak açıklanmaktadır.

### 1.12.1. Sonsal Tahmin p-Değeri (Posterior Predictive p-Value; PPP)

Bayes modelleme bağlamında PPP, modelin gözlemlenen veriye ne derece uyum sağladığını değerlendirmek için kullanılan önemli bir ölçüttür. Her ne kadar PPP, frekansçı istatistiklerdeki  $\chi^2$  testi kadar yaygın olmasa da Bayes modelleme bağlamında 0,5 civarındaki bir PPP değeri, modelin gözlemlenen veriye yüksek düzeyde uyum sağladığını gösteren bir gösterge olarak değerlendirilmektedir (Muthén ve Asparouhov, 2012). Modelin yanlış tanımlanması hâlinde ise aşırı uçta PPP değerleri (örneğin  $PPP < 0,05$  veya 0,01) gözlenmesi beklenir; bu tür değerler,

PPP'nin doğru tanımlanmış bir modelin olasılık dağılımına ait olmadığını ve dağılımın uç noktalarında yer aldığını gösterir (Kim vd., 2017). Kısaca PPP, gözlemlenen verilerle modelin tahmin ettiği (simüle edilen) veriler arasındaki uyumu test eder ve matematiksel olarak şu şekilde hesaplanır.

$$PPP = \frac{1}{N} \sum_{i=1}^N 1[T(y^{rep(i)}, \theta^{(i)}) \geq T(y, \theta^{(i)})]$$

Burada  $N$ , simülasyon sayısı (örnekleme miktarı);  $\sum, i = 1$ 'den  $N$ 'ye kadar toplam;  $\theta^{(i)}$ ,  $i$ . sonsal parametre örneği;  $y^{rep(i)}, \theta^{(i)}$  ile üretilen simüle veri;  $T$ , uyum istatistiği (varyans,  $\chi^2$  vb.); 1, gösterge fonksiyonu (koşul sağlanıyorsa 1, değilse 0 verir) şeklinde tanımlanmaktadır.

### 1.12.2. Güven Aralığı (CI)

Bayes, istatistikte kullanılan bir kavramdır ve frekansçı istatistikteki güven aralığından farklı bir şekilde yorumlanır. Frekansçı güven aralığı, popülasyondan çok sayıda tekrar eden örnekleme yapılacağı varsayımına dayanırken; Bayes güven aralığı, sonsal dağılımdan örnekleme yapılmasına dayanır (Alamri, 2019). Bu yaklaşım sayesinde, dağılımın çeyrek değerleri (quantiles) kolaylıkla kullanılabilir ve bir parametrenin belirli aralıkta yer alma olasılığı olduğu varsayılır. Matematiksel olarak, parametre uzayı  $\theta$ , için %100  $(1 - \alpha)$  Bayes güven aralığı, sonsal dağılım üzerinden aşağıdaki şekilde tanımlanmaktadır.

$$1 - \alpha = \int_{\theta_L}^{\theta_U} F(\theta|y) d\theta$$

Burada  $\theta_L$  ve  $\theta_U$ , güven aralığının alt ve üst sınırlarıdır;  $F(\theta|y)$ , gözlemlenen veri  $y$  verildiğinde parametre  $\theta$ 'nin sonsal olasılık yoğunluk fonksiyonu ve  $\alpha$ , dışta kalan toplam olasılığı ifade eder (örneğin, %5 için  $\alpha = 0,05$ ). Bayes güven aralığı, sonsal olasılık aralığı olarak da bilinir (Alamri, 2019).

### 1.12.3. Sapma Bilgi Kriteri (DIC)

Bayes model değerlendirme ve karşılaştırmalarda, DIC yaygın olarak kullanılan bir uyum ölçütüdür. DIC değeri ne kadar düşükse, modelin veriye o kadar iyi uyum sağladığı kabul edilir. Ancak en düşük DIC değerine sahip modelin mutlak anlamda en iyi model olduğu söylenemez; yalnızca diğer modellere kıyasla daha iyi bir uyum sunduğu anlaşılmaktadır. Asparouhov vd. (2015), büyük örneklemlerde ve

çok sayıda gözlenen gösterge içeren durumlarda DIC'nin, diğer model karşılaştırma indekslerine (örneğin BIC) göre daha güvenilir sonuçlar verdiğini belirtilmektedir. Benzer şekilde, Pokropek vd. (2020), DIC'nin tercih edilen önsel ortalama ve varyansın belirlenmesinde etkili bir gösterge olduğu vurgulanmıştır. Öte yandan, Hoijsink ve Van de Schoot (2018), DIC'nin küçük önsel dağılımlara sahip modellerin değerlendirilmesinde yetersiz kalabileceğini ifade etmiştir. Tüm bu bulgular ışığında, DIC'nin bir uyum indeksi olarak kullanımı umut verici görünse de elde edilen değerlerin dikkatle yorumlanması gerektiği açıktır. En azından, DIC tek başına değil, diğer uyum indeksleriyle birlikte değerlendirilmelidir. DIC, aşağıdaki biçimde hesaplanmaktadır:

$$DIC = \text{Ortalama Uyum Kötülüğü} + \text{Model Karmaşıklığı}$$

görüldüğü üzere DIC iki parçadan oluşmaktadır. Bu iki bileşen, Ortalama Uyum Kötülüğü ( $D$ ): modelin veriye ne kadar kötü uyduğunu gösterir.

$$D(\theta) = -2 \log p(y|\theta)$$

$D(\theta)$ , sonsal dağılımından alınan örnekler üzerinden hesaplanan ortalama sapma değeridir. Burada,  $y$  gözlemlenen veri,  $\theta$  model parametreleri ve  $p(y|\theta)$  ise veri için olasılığı belirtir. Ortalama uyum kötülüğü ise,

$$D = \frac{1}{S} \sum_{s=1}^S D(\theta^{(s)})$$

Burada,  $S$  sonsal dağılım örnek sayısı ve  $\theta^{(s)}$  her örnek için parametre değerini belirtir. Model Karmaşıklığı ( $p_D$ ): modeldeki parametre sayısı ve belirsizlikle ilgilidir.

$$p_D = D - D(\theta)$$

Burada  $\theta$ , genellikle sonsal ortalama parametre değeri,  $D(\theta)$ , bu ortalama parametreye karşılık gelen sapma değeridir. Yukarıda tanımlanan tüm bileşenler bir araya getirildiğinde, DIC aşağıdaki biçimde elde edilmektedir:

$$DIC = D + p_D = 2D - D(\theta)$$

Bayes yaklaşımında ölçme değişmezliği değerlendirmesi, yapısal, metrik ve skalar aşamaların her birinde modelin veriye uyumu değerlendirilmektedir. Bu süreçte, PPP değerinin anlamlı olmaması, CI sıfır değerini içermesi beklenmektedir. Bu durum, gözlenen ve simüle edilen değerler arasında anlamlı bir fark olmadığını ve dolayısıyla modelin veriye uyumlu olduğunu gösterir. Model karşılaştırmalarında ise DIC

kullanılmaktadır. En düşük DIC değerine sahip model, model veri uyumu sağlayan modeller arasında en iyi model olarak kabul edilir. Dolayısıyla ölçme değişmezliğinin tüm aşamaları, PPP, CI ve DIC kriterleri çerçevesinde sistematik olarak değerlendirilmiştir.

Bayes yaklaşımında ölçme değişmezliği süreci ilk olarak yapısal düzeyde başlamaktadır. Bu aşamada faktör yükleri ve madde sabitleri üzerinde herhangi bir önsel tanımlanmadan yazılımın varsayılan önselleri kullanılarak model test edilir. İkinci aşama olan metrik değişmezlikte, faktör yükleri arasındaki tutarsızlıkları değerlendirmek amacıyla yaklaşık değişmezlik önsel varyansı tanımlanır ve modeller karşılaştırılır. En iyi uyum gösteren model seçilir ve seçilen modeldeki önsel varyansın belirlenen kesme noktasına eşit veya daha küçük olması gerekir. Son aşama olan skalar değişmezlikte ise aynı prosedür bu kez madde sabitleri için uygulanmaktadır. Yaklaşık skalar değişmezliğe ulaşmak için seçilen modeldeki önsel varyansın, madde sabitlerine ait yaklaşık değişmezlik varyansına eşit ya da daha düşük olması gerekir. Araştırmada uygulanan yöntemle dair teknik bilgiler BYÖD analizi bölümünde yer verilmiştir.

## 2. ÖNCEKİ ÇALIŞMALAR

### 2.1. ÇGDFA Yöntemi ile Yapılan Ölçme Değişmezliği Çalışmaları

TIMSS verilerine dayalı gerçekleştirilen ölçme değişmezliği çalışmalarında, ÇGDFA yöntemi kullanılarak öğrencilerin matematik duyuşsal özelliklerinin farklı gruplar arasında tutarlılıkla ölçülüp ölçülemediği incelenmiştir. Bu kapsamda yapılan araştırmalarda cinsiyet (Polat, 2019; Atılgan, 2022; Yiğiter, 2023; Uyar, 2021), kültür (Bofah ve Hannola, 2015; He vd., 2018; Koçak, 2024; Polat, 2019; Yiğiter, 2024) ve bölge (Ölçüoğlu ve Çetin, 2016; Polat, 2019) gibi değişkenler temel alınarak geliştirilen modeller test edilmiştir. Cinsiyet değişkenine göre yapılan analizlerde, katı değişmezlik düzeyine ulaşılan çalışmalar (Polat, 2019; Uyar, 2021; Atılgan, 2022; Yiğiter, 2023) sayesinde gruplar arası doğrudan karşılaştırmaların güvenilir olduğu görülmüştür. Ancak kültürlerarası ve bölge değişkenleri ele alındığında genellikle metrik (He vd., 2018; Koçak, 2024; Yiğiter, 2023) veya skalar (Bofah ve Hannola, 2015; Ölçüoğlu ve Çetin, 2016) düzeyde değişmezlik sağlanmış, dolayısıyla bu karşılaştırmalarda daha temkinli yaklaşılması gerektiği anlaşılmıştır. Literatürde, TIMSS'in duyuşsal verileri kullanılarak ÇGDFA yöntemi aracılığıyla ölçme değişmezliğinin incelendiği çalışmalar ulusal ve uluslararası düzeyde oldukça yaygındır. Literatürdeki mevcut araştırmalar, yayın tarihleri dikkate alınarak ayrıntılı biçimde sunulmuştur.

Bofah ve Hannula (2015) araştırmalarında, TIMSS 2011 öğrenci tutum anketinden elde edilen veriler kullanılarak beş Afrika ülkesinden toplam 38.806 sekizinci ve dokuzuncu sınıf öğrencisi incelenmiştir. Çalışmada matematikte özgüven, matematik öğrenmeyi sevme, matematiğe değer verme, öğretmen duyarlılığı ve ebeveyn katılımı/pozitif ebeveynlik olmak üzere beş boyutlu duyuşsal yapılar değerlendirilmiştir. Normallikten sapmaları kontrol etmek amacıyla MLR yöntemi kullanılmıştır. Ölçme değişmezliği analizleri ÇGDFA yöntemi ile test edilmiştir. Bulgular, yapısal, metrik ve kısmi skaler değişmezliğe güçlü destek sağlamıştır. Böylece duyuşsal yapılar için yapılan ortalama karşılaştırmaların gizil ortalama farklılıklarını yansıttığı ortaya konmuştur.

Ölçüoğlu ve Çetin (2016) araştırmalarında, TIMSS 2011 uygulamasında yer alan Türkiye'nin 6.928 sekizinci sınıf öğrencisinin verilerine dayanarak, öğrencilerin

matematik başarısını etkileyen değişkenler modellenmiş ve oluşturulan modelin coğrafi bölgelere göre ölçme değişmezliği incelenmiştir. Araştırmanın ilk aşamasında, öğrenci anketindeki maddeler Açımlayıcı Faktör Analizi (AFA) ile incelenmiş ve bu maddelerin on ikisini matematik ile ilgili duyuşsal özellikler, dokuzunun ev ortamı, beşinin ise okul ortamı boyutlarında toplandığı görülmüştür. Ardından, bu değişkenlerle öğrencilerin matematik başarıları arasındaki ilişkileri test etmek amacıyla YEM kullanılmış ve elde edilen bulgular doğrultusunda, matematik başarısı üzerinde en fazla pozitif etkiyi duyuşsal özelliklerin gösterdiği, bunu ev ortamının izlediği, okul ortamının ise negatif yönde fakat anlamlı olmayan bir ilişki gösterdiği belirlenmiştir. Araştırmanın ikinci aşamasında kurulan modelin tüm bölgelerde geçerli olduğu saptanmış, ancak ÇGDFA ile yapılan ölçme değişmezliği analizinde yalnızca yapısal ve metrik değişmezlik koşullarının sağlandığı, bu nedenle yalnızca bu düzeylerdeki karşılaştırmaların güvenilir kabul edilebileceği sonucuna ulaşılmıştır. Skalar değişmezlik sağlanmadığında, gözlenen farklılıkların bölgelerden mi yoksa ölçme aracından mı kaynaklandığı belirlenememektedir. Bu nedenle, bölgeler arası matematik başarı puanların karşılaştırılmasında dikkatli olunması gerektiği sonucuna ulaşılmıştır.

He vd. (2018) araştırmalarında, PISA 2015 ve TIMSS 2015 sekizinci sınıf öğrenci bağlam anketlerinden elde edilen öz bildirim verileri kullanılmış ve her iki araştırmaya katılan 29 ülke örnekleme dahil edilmiştir. Yöntem olarak, ölçeklerin ölçme değişmezliği ÇGDFA çerçevesinde test edilmiş, metrik düzeyde ölçme değişmezliği sağlanan yapılar için faktör puanları elde edilerek Madde Tepki Kuramı temelli ölçek puanlarıyla karşılaştırılmıştır. Bulgular, araşsal motivasyon ve fen bilimlerinden keyif alma ölçeklerinin kültürlerarası karşılaştırılabilir olduğunu, farklı ölçekleme yöntemleriyle elde edilen puanların benzer desenler sergilediğini ve bu yapıların fen başarısı ile pozitif ilişkili olduğunu göstermiştir. Buna karşılık, PISA'daki okula aidiyet ölçeği zayıf uyum göstermiş ve başarı ile ilişkisi tutarlı bulunmamıştır. Genel olarak, çalışma ölçme değişmezliği testlerinin önemini vurgulamakta ve PISA ile TIMSS'teki bilişsel olmayan değerlendirmelerin geçerliliği üzerine metodolojik ve pratik çıkarımlar sunmaktadır.

Ertürk ve Erdiñ-Akan (2018) çalışmalarında, TIMSS 2015 uygulamasında yer alan Türkiye'nin 6.456 dördüncü sınıf öğrencisinin verilerine dayanarak, matematik

başarısını etkileyen gizil değişkenlerin (matematiği sevme, matematiğe olan ilgi ve matematiğe ilişkin özgüven) ölçme değişmezliği cinsiyet değişkenine göre incelenmiştir. İlişkisel desenli bu çalışmada, ölçme değişmezliği, ÇGDFA yöntemiyle yapısal, metrik, skalar ve katı değişmezlik aşamaları değerlendirilmiştir. Bulgular, tüm değişkenlerin yapısal değişmezliği sağladığı ancak yalnızca matematiği sevme değişkeninin cinsiyet gruplarında katı değişmezlik düzeyini karşıladığını göstermiştir.

Polat (2019) çalışmasında, TIMSS 2015 uygulamasında yer alan sekizinci sınıf öğrenci anket verilerini kullanarak oluşturulan matematik ve fen duyuşsal özellik modellerinin ölçme değişmezliği kültürlerarası (Singapur, Suudi Arabistan ve Türkiye), cinsiyet ve bölge değişkenlerine göre incelenmiştir. Araştırma, toplam 15.954 öğrenci verisiyle yürütülmüş ve öğrenci anketlerinden elde edilen verilerle matematik ve fen dersini sevme, bu derslere yönelik etkinliklere katılım, kendine güven ve derslere verilen değer gibi duyuşsal boyutlar için ayrı ayrı modeller oluşturulmuş ve ÇGDFA yöntemi ile ölçme değişmezliğini değerlendirilmiştir. Bulgulara göre, oluşturulan duyuşsal özellik modelleri kültürlerarası ve bölgelere göre skalar değişmezliğini, cinsiyet değişkenine göre ise katı değişmezliği sağlamıştır. Bu sonuçlar, cinsiyete göre yapılan karşılaştırmaların anlamlı olduğu; kültürlerarası ve bölgelere göre skalar değişmezliğin sağlanması herhangi bir yanlılık bulunmadığını göstermektedir. Bu bağlamda, gözlenen değişkenlerdeki farklılıkların yapıların ortalama farklılıklarından kaynaklandığı sonucuna ulaşılmıştır.

Uyar (2021) çalışmasında, TIMSS 2015 uygulamasında sekizinci sınıf öğrencilerine yönelik gerçekleştirilen Matematik Tutum Anketi'nin Keşifsel Yapısal Eşitlik Modellemesi, DFA modellerine uygunluğunu ve cinsiyete göre ölçme değişmezliği incelenmiştir. Türkiye örneklemini oluşturan 6.079 öğrenci ile yürütülen çalışmada, tutum anketine ait maddeler; öğrencilerin matematiğe olan ilgisi, öğretimin ilgi çekiciliğine dair görüşleri, özgüven düzeyleri ve matematiğe verdikleri değer olmak üzere dört ana başlık altında toplanmıştır. ESEM ile yapılan analiz sonucunda, anketin 5 faktörlü ve 35 maddeli yapısının modele uygun olduğu belirlenmiştir. Ayrıca, anketin cinsiyet değişkenine göre ESEM'e göre yapısal, metrik, skalar, katı, varyans-kovaryans ve gizil ortalama değişmezliği olmak üzere altı aşamalı bir süreç test edilmiş ve tüm değişmezlik koşullarını sağladığı saptanmıştır.

Atılgan (2022) çalışmasında, TIMSS 2019 uygulamasına Türkiye’den katılan sekizinci sınıf öğrencilerine ait matematikle ilişkili duyuşsal özelliklerin cinsiyet değişkenine göre ölçme değişmezliğini incelemeyi amaçlamaktadır. Araştırma kapsamında 4.048 öğrenciden elde edilen verilerde; okula ait olma duygusu, matematikte kendine güven, matematiği sevmeye ve matematiğe değer verme ölçekleri analiz edilmiştir. Geçerlik için AFA ve DFA, güvenilirlik için ise Cronbach Alfa katsayıları hesaplanmıştır. DFA sonuçlarına göre, dört ölçekten sadece matematikte kendine güven ölçeği yapı geçerliğini sağlayamamış ve bu nedenle ölçme değişmezliği analizine dahil edilmemiştir. Diğer üç ölçek için ölçme değişmezliği ÇGDFA yöntemiyle test edilmiş ve matematiği sevmeye ile matematiğe değer verme ölçeklerinde katı değişmezlik, okula ait olma duygusu ölçeğinde ise skalar değişmezlik sağlanmıştır.

Dokur (2023) çalışmasında, TIMSS 2019 uygulamasına katılan sekizinci sınıf öğrencilerinin matematiğe yönelik güven düzeylerini ölçmek amacıyla kullanılan ölçeğin farklı ülkeler arasında ölçme değişmezliğini incelemeyi amaçlamaktadır. Yedi farklı kültürü temsil eden ülke (Şili, Hong Kong, İsrail, Katar, Rusya, Türkiye ve İngiltere) örnekleminde gerçekleştirilen araştırma, ilişkisel tarama modeliyle yürütülmüştür. Çalışmada ölçme değişmezliği, ÇGDFA yöntemi ve Hizalama Yöntemi (HY) kullanılmıştır. ÇGDFA sonuçları, ölçeğin yapısal, metrik ve üç madde açısından kısmi skalar değişmezliğini sağladığını ortaya koyarken, HY verileri, önerilen sınırların üzerinde değişmezlik sapmaları göstermiş ve ölçeğin genel olarak ölçme değişmezliğine ulaşamadığını ortaya koymuştur. Bu doğrultuda, matematiğe yönelik güven ölçeğinin yedi ülke için tam anlamıyla eşdeğer olmadığı ifade edilmekte ve ülkeler arası karşılaştırmalarda bu durumun mutlaka göz önünde bulundurulması gerektiği vurgulanmaktadır.

Yiğiter (2023) çalışmasında, uluslararası değerlendirmelerdeki karşılaştırmaların güvenilirliğini sağlamak amacıyla ölçme değişmezliğini ele alarak, TIMSS 2019 Türkiye örnekleminde sekizinci sınıf öğrencilerine ait matematik duyuşsal özellikleri modelinin cinsiyete göre ölçme değişmezliğini test etmeyi amaçlamıştır. Model, matematik öğrenmeyi sevmeye, öğretimin netliği, dersteki disiplinsiz davranışlar, kendine güven ve matematiğe değer verme ölçeklerinden oluşturulmuş ve 3.658 öğrenci verisiyle analiz edilmiştir. Geliştirilen modelin DFA

sonuçları, modelin veriyle iyi uyum sağladığını göstermiştir. Ardından yapılan ÇGDFA ile modelin yapısal, metrik, skalar ve katı değişmezlik düzeylerini sağladığı görülmüş, yani cinsiyete göre ölçülen yapılar istatistiksel olarak anlamlı karşılaştırmaların yapılabileceği sonucuna ulaşılmıştır. Son olarak, *t* testleri ile yapılan analizlerde erkeklerin öğretim netliğinde daha yüksek puanlara sahip olduğu, kız öğrencilerin ise matematikte kendine güven ve dersteki disiplinsiz davranışlar ölçeklerinde daha avantajlı olduğu gözlemlenmiştir. Buna karşın matematiğe değer verme ve matematik öğrenmeyi sevme ölçeklerinde anlamlı bir cinsiyet farkı saptanmamıştır.

Yiğiter (2024) çalışmasında, farklı ülkeler ve eğitim sistemleri arasında karşılaştırma yapmanın güvenilirliği açısından kritik bir öneme sahip olan ölçme değişmezliğini ele almaktadır. Araştırma kapsamında, TIMSS 2019 döngüsünde yer alan “öğrenciler matematik öğrenmeyi sever” ölçeği üzerinden matematik içsel motivasyonunun 14 ülkede ne ölçüde karşılaştırılabilir olduğu incelenmiştir. Dördüncü sınıf (70192) ve sekizinci sınıf (82800) olmak üzere toplam 152.992 öğrenciye ait verilerle gerçekleştirilen analizlerde, ölçme değişmezliği hem ÇGDFA hem de HY kullanılarak test edilmiştir. Dördüncü sınıf düzeyinde MGDFFA yöntemine göre sadece yapısal değişmezlik, HY’ye göre yaklaşık değişmezlik elde edilmiştir. Sekizinci sınıf düzeyinde ÇGDFA yöntemine göre metrik değişmezliğin sağlandığı, HY’ye göre ise yaklaşık değişmezliğin bulunduğunu göstermiştir.

Koçak (2024) çalışmasında, TIMSS 2019 uygulamasında yer alan sekizinci sınıf matematik başarı testi ile matematikle ilişkili duyuşsal özelliklere dair ölçeklerin Türkiye, Güney Kore ve Finlandiya örneklemelerinde ölçme değişmezliğini incelemektedir. Araştırmada Matematik Öğrenmeyi Sevme, Matematikte Kendine Güven ve Matematiğe Değer Verme ölçeklerinden oluşan bir model ele alınmış ve başarı testiyle birlikte bu yapıların ülkeler arasında ne ölçüde karşılaştırılabilir olduğu araştırılmıştır. Öncelikle her bir ölçek ve model için AFA ve DFA gerçekleştirilmiş, ardından üç ülke verisi birlikte kullanılarak ÇGDFA uygulanmıştır. Çalışmanın bulguları, geliştirilen modelin metrik düzeyde ölçme değişmezliğini sağladığı, ancak skalar değişmezlik sağlamadığını göstermektedir. Bu nedenle, gözlenen farklılıkların ülkelerden kaynaklanmadığı, karşılaştırmalar yapılırken bu durumun dikkate alınması gerekmektedir.

## 2.2. BYÖD Yöntemi ile Yapılan Ölçme Değişmezliği Çalışmaları

Literatürde TIMSS duyuşsal verileri kullanılarak BYÖD yöntemiyle ölçme değişmezliği incelenen çalışmalar oldukça nadir rastlanmaktadır. Bundan dolayı literatür taramasına farklı disiplinlerde yapılan araştırmalar ele alınmıştır. Araştırmalar; zihinsel dayanıklılık (Gucciardi vd., 2016), insan değerleri (Cieciuch vd., 2014; Cieciuch vd., 2018), akademik stres (Winter ve Depaoli, 2020), yaşam memnuniyeti (Seddig ve Leitgöb, 2018), çevre tutumları (Arts vd., 2021), otizm belirtileri (Stevanovic vd., 2021), bilişsel işlevler (Williams vd., 2018), matematik başarısı (Wang, 2017) gibi farklı yapılar üzerine odaklanmıştır. Bayes analizleri, frekansçı yöntemlerin sağlayamadığı metrik ve skalar değişmezliği, küçük sapmalara izin vererek elde etmiş; böylece çapraz kültürel ve boylamsal analizlerde esnek ve uygulanabilir bir çözüm sunmuştur. Literatürde, BYÖD yöntemiyle yapılan araştırmalar ile ÇGDFA ve BYÖD'ü birlikte uygulayan araştırmalar, yayın tarihleri dikkate alınarak ayrıntılı biçimde sunulmuştur. TIMSS verilerinin BYÖD yöntemiyle incelenmesi literatürde sınırlı sayıda araştırma yer aldığından, araştırmancının bu açıdan literatüre önemli bir katkı sağlaması beklenmektedir.

Cieciuch vd. (2014) çalışmalarında, insani değeri ölçmek için geliştirilen PVQ-5X ölçeğinin kültürlerarası geçerliğini test etmeyi ve kesin ölçme değişmezliği analizleri ile yaklaşık ölçme değişmezliği yöntemlerini karşılaştırarak bu alandaki metodolojik bir boşluğu doldurmayı amaçlamaktadır. ÇGDFA ile yapılan klasik testlerin yanı sıra, Muthén ve Asparouhov (2013) tarafından önerilen ve parametreler arasında küçük farklılıklara izin veren BYÖD yöntemini kullanılarak PVQ-5X'in farklı ülkelerdeki geçerliği değerlendirilmiştir. Bu çalışmada BYÖD yöntemi ile 19 değer için yaklaşık skalar değişmezlik elde edilmiştir.

Gucciardi vd. (2016) çalışmalarında, sporcuların zihinsel dayanıklılığa ilişkin öz bildirimlerinin kültürlerarası ölçme değişmezliğini değerlendirmek ve spor psikolojisi alanındaki araştırmacılara BYÖD analizini tanıtmak ve örneklemektir. Avustralya (353), Çin (254) ve Malezya (341) olmak üzere toplam 948 sporcunun katılımıyla gerçekleştirilen araştırmada, zihinsel dayanıklılık envanteri BYÖD yöntemi kullanılarak yapısal, metrik ve skalar değişmezlik düzeylerinde analiz edilmiştir. Bulgular, yaklaşık metrik ve skalar değişmezliğin sağlandığını

göstermektedir. Bu sonuçlar, BYÖD yönteminin kültürlerarası örneklerle yürütülen psikometrik geçerlik analizlerinde esnek ve faydalı bir yöntem olduğunu ortaya koyarken; aynı zamanda zihinsel dayanıklılık gibi bağlamsal olarak hassas yapıları ölçerken kültürel uyarlamaların dikkate alınması gerektiğini vurgulamaktadır.

Wang (2017) çalışmasında, PISA 2009 ve PISA 2012 matematik testlerinde 15 yaşındaki öğrencilerin performanslarını değerlendirerek BYÖD yönteminin zamana ve gruplar arası karşılaştırmalarda ne kadar esnek ve işlevsel olduğunu ortaya koymaktadır. Araştırmanın örnekleme, söz konusu PISA sınavlarına katılan 15 yaş grubu öğrencilerden oluşmaktadır. Yöntem olarak, BYÖD yaklaşımı uygulanmış; modelin istikrarı ve performansı ise Monte Carlo simülasyon çalışmasıyla test edilmiştir. Simülasyon sonuçları, modelin veriye iyi uyum sağladığını ve istikrarlı tahminler sunduğunu göstermiştir. Elde edilen bulgular, çapraz yüklemelerin tam sıfır olmasının gerçekçi olmadığı kültürlerarası çalışmalarda bu yaklaşımın ÇGDFA yöntemine güçlü bir alternatif sunduğunu ortaya koymuş; ayrıca ikinci düzey birimlerdeki örneklem büyüklüğü gibi uygulamaya yönelik pratik konulara dair önemli rehberlik sağlamıştır.

Cieciuch vd. (2018) çalışmalarında, ülkeler arası anlamlı karşılaştırmalar için gerekli olan ölçme değişmezliğini, kesin değişmezlik varsayımı yerine BYÖD yaklaşımıyla test etmektedir. Araştırmada, Avrupa Sosyal Araştırması'na (ESS) katılan 15 ülkeden altı veri toplama turunda yer alan toplam 274.447 katılımcının yanıtladığı 21 maddelik Portre Değerler Anketi (PVQ-21) ölçeği kullanılmıştır. Yöntem olarak, Muthén ve Asparouhov'un geliştirdiği BYÖD tekniği uygulanarak ülkeler arasında parametrelerin tam değil yaklaşık eşitliği temelinde modelleme yapılmıştır. Bulgular, değişime açıklık ve benliği güçlendirme gibi iki üst düzey değer boyutunda yaklaşık ölçme değişmezliğinin neredeyse tüm turlarda sağlandığını; ancak benliği aşma ve koruma değerleri için bu durumun yalnızca bazı ülkelerle sınırlı kaldığını göstermiştir. Sonuç olarak, bu yöntem ölçme modellerinde daha esnek bir yaklaşım sunmakta ve kesin değişmezliğin sağlanamadığı durumlarda dahi ülkeler arası karşılaştırmalar için anlamlı bir temel oluşturmaktadır.

Seddig ve Leitgöb (2018) çalışmalarında, boylamsal DFA kapsamında yaklaşık ölçme değişmezliği yaklaşımı kullanarak zamana bağlı ölçme değişmezliğini

değerlendirmeyi amaçlamaktadır. Genç bireylerin, hayattan keyif alma ve haz peşinde olma eğilimlerine ilişkin panel veriler kullanılarak yürütülen araştırmada, ÇGDFA yöntemi ile skalar değişmezliğin sağlanamaması nedeniyle örtük ortalama karşılaştırmalarının yapılamadığı görülmüştür. Buna karşın, ortalaması sıfır ve küçük varyansa sahip Bayes önsel dağılımlar aracılığıyla yaklaşık skalar değişmezliğin elde edildiği bulunmuştur. Modelin uyumu ve yakınsama süreci için kullanılan izleme prosedürleri, düşük bir önsel varyansın bile zaman içindeki madde sabiti farklarındaki küçük sapmaları açıklamak için yeterli olduğunu göstermiştir. Bu sonuçlar, Bayes yaklaşımın geleneksel sınırlılıkları aşarak zamana bağlı psikolojik yapıları daha esnek ve güvenilir bir biçimde değerlendirmek için etkili bir alternatif sunduğunu ortaya koymaktadır.

Winter ve Depaoli (2020) çalışmalarında, küçük örneklem boyutuna sahip boylamsal verilerde, BYÖD yaklaşımının (Mplus programı kullanılarak) uygulanabilirliğini gösterilmektedir. Araştırmanın örneklemini, önemli bir vize sınavı öncesi ve sonrasında akademik stres düzeyleri ölçülen üniversite öğrencileri oluşturmaktadır. Yöntem olarak, kesin ölçme değişmezliği testlerinin ötesine geçilerek, zaman içerisindeki parametre farklılıklarına sıfır ortalamalı ve küçük varyanslı önsel dağılımlar kullanarak analizde küçük sapmalara izin veren BYÖD yöntemi uygulanmıştır. Kesin ölçme değişmezliği yöntemleri küçük örneklem nedeniyle yetersiz kalırken, Bayes yaklaşım bazı parametrelerin zaman içinde değiştiğini tespit etmiş ve buna göre kısmen değişmezlik sağlayan bir modelin tahmin edilmesini mümkün kılmıştır. Sonuç olarak, bu yöntem, özellikle örneklem kısıtı olan çalışmalarda gizil yapıların zamana bağlı karşılaştırmalarında değerli ve esnek bir alternatif sunmaktadır.

Sideridis vd. (2020) çalışmalarında, Suudi Arabistan'daki ulusal bir sınav verisini kullanarak BYÖD yöntemiyle ölçme değişmezliğini test etmek ve kesin değişmezlik aşamalarından olan skalar değişmezlik koşulları sağlanamadığında alternatif bir yaklaşım sunmaktır. Yöntem olarak, Bayes temelli değerlendirme çerçevesi kullanılarak yalnızca değişmez parametreler üzerinden örtük ortalamaların karşılaştırılması yapılmış; eşdeğerlik eşiklerinin belirlenmesinde 0,10'luk önsel varyanslar kullanılmıştır. Bulgular, etki büyüklüğü ölçütleri ve önsel-sonsal olasılıklı tahmin p-değeri (PPP) yardımıyla gruplar arasında sıfırdan anlamlı şekilde farklılık

gösteren ortalama farklarının belirlenebildiğini ortaya koymuştur. Sonuç olarak, Bayes yaklaşımları ve PPP istatistiği, ölçme değişmezliğinin tam sağlanamadığı durumlarda bile anlamlı grup karşılaştırmaları yapmak için güçlü ve esnek bir alternatif sunmaktadır.

Arts vd. (2021) çalışmalarında, farklı ülkelerde çevre ve iklim değişikliğine ilişkin tutumları değerlendiren büyük ölçekli sosyal anketlerde karşılaştırma yapılabilirliğinin temel şartı olan ölçme değişmezliğini, BYÖD ve HY'ye göre test etmeyi amaçlamaktadır. Katılımcıların “çevre için fedakârlık yapma istekliliği” ne dair yanıtları, ÇGDFA, HY ve BYÖD modelleriyle analiz edilmiştir. Bayes analizlerde farklı önsel varyanslar kullanılarak bu durumun ülke sıralamaları üzerindeki etkileri değerlendirilmiş; bulgular, modellerin istatistiksel olarak uyumlu görünse bile küçük varyans değişimlerinin ülke sıralamalarını anlamlı biçimde etkileyebileceğini göstermiştir. Sonuçlar, İsviçre ve Güney Kore halkının çevre için fedakârlık yapmaya en istekli gruplar olduğunu, Letonya'nın ise en düşük motivasyonu sergilediğini ortaya koymuştur. Bu da Bayes yaklaşımların kesin ölçme değişmezliği yöntemlerin ötesine geçerek ölçme geçerliği analizlerinde dikkatli biçimde uygulanması gerektiğini göstermektedir.

Stevanovic vd. (2021) çalışmalarında, yaygın olarak kullanılan ve pratik bir tarama aracı olan Çocukluk Otizmi Değerlendirme Ölçeği'nin (CARS) farklı ülkelerde Otizm Spektrum Bozukluğu (OSB) belirtilerini kültürlerarası tutarlı bir şekilde ölçüp ölçmediğini değerlendirmektedir. Araştırma kapsamında Hindistan, Jamaika, Meksika, İspanya, Türkiye ve Amerika Birleşik Devletleri'nden elde edilen, toplamda OSB tanılı 747 çocuğa ait daha önceki çalışmalardan derlenen veriler kullanılmıştır. Ölçme değişmezliği, BYÖD yöntemiyle değerlendirilmiştir. Analiz sonuçları, modelin veriye yeterli uyum göstermediğini ve ölçeğin hiçbir maddesinde kültürlerarası değişmezliğin sağlanamadığını ortaya koymuştur. Sosyal iletişim ve etkileşimle ilgili maddeler (örneğin taklit, duygusal tepki, sözlü/sözsüz iletişim), duygusal hassasiyet ve stereotipik davranışlara ilişkin maddelere göre daha düşük düzeyde farklılık sergilemiştir. Bu bulgular, CARS'ın kültürlerarası karşılaştırmalarda geçerliğinin sınırlı olduğunu ve farklı ülkeler arasında yapılan karşılaştırmalarda hangi maddelerin farklı işleyiş gösterdiğine dikkat edilmesi gerektiğini göstermektedir. Ölçeğin kültürel olarak daha duyarlı hale getirilebilmesi için ileri çalışmalara ihtiyaç duyulmaktadır.

Bagheri vd. (2022) çalışmalarında, dünya genelinde yaygın olarak kullanılan Yaşam Kalitesi, Yaşamdan Zevk Alma ve Yaşam Memnuniyet Ölçeği-Kısa Formu'nun (Q-LES-Q-SF) farklı kültürlerde benzer şekilde çalışıp çalışmadığını, yani kültürlerarası ölçme değişmezliğini incelemeyi amaçlamaktadır. Bangladeş, Brezilya, Hırvatistan, Hindistan, Nepal, Polonya, Sırbistan, Türkiye, Birleşik Arap Emirlikleri ve Vietnam'dan toplam 2.822 üniversite öğrencisinden oluşan örnekleme yürütülen araştırmada, BYÖD yöntemiyle analiz edilmiştir. Sonuçlar, ölçeğin genel olarak yaklaşık ölçme değişmezliğini karşılamadığını, ancak 14 maddeden yalnızca ikisinin ev işleri ve boş zaman etkinlikleriyle ilgili maddeler değişmez olmadığını göstermektedir. Bu bulgular, ülkeler arasında yaşam kalitesi puanlarının doğrudan karşılaştırılmasında dikkatli olunması gerektiğini ortaya koymakta; kültürel farklılıkları azaltmak amacıyla ölçekteki bazı maddelerin yeniden ifade edilmesi veya uyarlanmasının yararlı olabileceğini önermektedir.

Sırgancı (2022) araştırmasında, Avrupa Sosyal Araştırması 2020 uygulamasına katılan on ülke örneklem olarak alınmış ve geniş ölçekli değerlendirmelerde çoklu grup karşılaştırmalarının geçerliğini test etmek amacıyla farklı ölçme değişmezliği yöntemleri kullanılmıştır. Frekansçı yaklaşım olarak ÇGDFA, yaklaşık ölçme değişmezliği için ise BYÖD ve HY uygulanmıştır. ÇGDFA ve HY, ML kestirimiyle yürütülmüştür. Bayes analizde ise faktör yükleri ve madde sabitleri için farklı önsel varyans değerleri (0,001; 0,005; 0,010; 0,025; 0,050) tanımlanmıştır. Bulgular, ÇGDFA'da metrik değişmezlik aşamasında gruplar arasında ölçme değişmezliğinin sağlanmadığını göstermiştir. Buna karşılık BYÖD ve HY analizlerinde gruplar arasında faktör ortalamalarının karşılaştırılabilir olduğu modeller elde edilmiş, ancak tüm ülkeler için karşılaştırılabilirlik yalnızca BYÖD modeliyle sağlanmıştır. Sonuç olarak, ülkeler arası karşılaştırmalarında özellikle grup sayısının fazla olduğu durumlarda Bayes yaklaşımın daha güçlü ve geçerli karşılaştırmalar sunduğu ortaya konmuştur.

Wurster (2022) araştırmasında, 26 ülkenin 2007, 2011 ve 2015 TIMSS döngülerine ait okul müdürü, öğretmen ve öğrenci verilerini kullanarak bilişsel olmayan ölçümlerin ülkeler ve zaman boyutundaki ölçme değişmezliğini incelemiş ve farklı analiz yaklaşımlarını karşılaştırmıştır. Kesin ölçme değişmezliğini test etmek için ÇGDFA uygulanmış, yaklaşık ölçme değişmezliği için ise BYÖD ve HM

yöntemleri de kullanılmıştır. Bulgular, ÇGDFA ve BYÖD analizlerin ülkeler arası tüm incelemelerde yapısal değişmezliğini ve zaman boyutunda çoğunlukla metrik düzeyde değişmezliği gösterdiğini ortaya koymuştur. Buna karşılık, HM yöntemi çoğu ölçek için yaklaşık ölçme değişmezliği varsayımı altında örtük ortalamaların ülkeler arasında karşılaştırılabilir olduğunu göstermiştir. Ayrıca çalışmada BYÖD yaklaşımında sınırlılıklar mevcut olup özellikle yüksek hesaplama gücü ve zaman gerektirmesi nedeniyle bu çalışmada öğrenci verileri için uygulanamamıştır.

### 3. GEREÇ VE YÖNTEM

#### 3.1. Araştırmanın Deseni

Bu araştırma, TIMSS 2019 uygulamasına katılan sekizinci sınıf öğrencilerin MÖSÖ'den elde edilen gerçek verilere dayanmaktadır. ÇGDFA ve BYÖD yöntemleri kullanılarak belirlenen ülkelere göre ölçme değişmezliğini incelemeyi amaçlamaktadır. Mevcut yapının veriler yoluyla, herhangi bir müdahale olmaksızın farklı tekniklerle değerlendirilmesi ve örneklemelerden elde edilen bilgilerle evren hakkında çıkarımlar yapılması nedeniyle, araştırma betimsel araştırma niteliği taşımaktadır (Fraenkel ve Wallen, 2006).

#### 3.2. Veri Seti

TIMSS uygulamasının hedef kitlesi, katılımcı ülkelerde örgün eğitime devam eden dördüncü ve sekizinci sınıf öğrencileridir. Ülkeler, uygulamaya yalnızca bir sınıf düzeyinde ya da her iki düzeyde katılım sağlayabilmektedir. TIMSS 2019 uygulamasına dördüncü sınıf düzeyinde 58, sekizinci sınıf düzeyinde ise 39 ülke katılmıştır (Mullis vd., 2020). TIMSS uygulamalarında iki aşamalı rastgele örnekleme yöntemi kullanılmaktadır. İlk aşamada, ülkelerdeki okullar coğrafi bölge ve okul türlerine göre tabakalı örnekleme ile seçilmektedir. İkinci aşamada ise bu okullardan uygulamaya katılacak sınıflar rastgele belirlenmekte ve seçilen sınıflardaki tüm öğrenciler uygulamaya dahil edilmektedir.

Bu çalışmada, sekizinci sınıf düzeyinde uygulanan MÖSÖ'nün ülkeler arasında ölçme değişmezliğini incelemek amaçlanmıştır. Araştırma evrenini 39 katılımcı ülke oluşturmaktadır. Araştırmanın örnekleme ise, TIMSS 2019 uygulamasında sekizinci sınıf matematik başarı ortalamaları dikkate alınarak belirlenmiştir. Bu doğrultuda; yüksek başarı düzeyine sahip ülkeler Tayvan (612 puan) ve Singapur (616 puan), orta düzeyde başarı gösteren ülkeler Türkiye (496 puan) ve Portekiz (500 puan) ile düşük başarı düzeyine sahip ülkeler Suudi Arabistan (394 puan) ve Kuveyt (403 puan) olmak üzere toplam altı ülke örnekleme dahil edilmiştir. Ülkelerin başarı düzeyleri esas alınarak örneklem seçildiği için, seçkisiz olmayan örnekleme yaklaşımlarından amaçsal örnekleme yöntemi tercih edilmiştir. Bu yöntemin maksimum çeşitlilik yaklaşımıyla da başarı düzeyleri farklı olan ülkeler örnekleme dahil edilerek evrenin çeşitliliği yansıtılmaya çalışılmıştır. Araştırmanın

örneklemini oluşturan öğrencilerin ülkelere göre dağılımına ilişkin bilgiler Tablo 3.1.'de sunulmuştur.

**Tablo 3.1.** Araştırma Örneklemini Oluşturan Öğrencilerin Ülkelere Göre Dağılımı

Ülkeler	<i>N</i>	%
Kuveyt	4.574	16,6
Portekiz	3.377	12,3
Singapur	4.853	17,7
Suudi Arabistan	5.680	20,7
Tayvan	4.915	17,9
Türkiye	4.077	14,8
Toplam	27.476	100

Tablo 3.1. incelendiğinde, altı ülkeden toplam 27.476 öğrencinin araştırmaya dahil edildiği görülmektedir. Öğrenci dağılımı incelendiğinde, en fazla katılımcıya sahip ülkenin 5.680 öğrenci (%20,7) ile Suudi Arabistan olduğu, en az katılımcıya sahip ülkenin ise 3.377 öğrenci (%12,3) ile Portekiz olduğu belirlenmiştir. Diğer ülkelerin katılımcı sayıları ise şu şekildedir: Tayvan 4.915 öğrenci (%17,9), Singapur 4853 öğrenci (%17,7), Kuveyt 4.574 öğrenci (%16,6) ve Türkiye 4.077 öğrenci (%14,8). Bu dağılım, ülkelerin örneklem büyüklüklerinin birbirine yakın olduğunu ve analizlerde karşılaştırılabilirlik açısından dengeli bir yapı sunduğunu göstermektedir.

YEM büyük örneklem gerektirir. Örneklem büyüklüğünün, kestirilen parametre sayısının en az yirmi katı olması gerektiği belirtilmektedir (Şen, 2023). Ayrıca, bazı kaynaklarda parametre sayısının en az 10 katı ya da en az 200 katılımcıdan oluşan örneklemelerin yeterli olabileceği ifade edilmektedir (Kline, 2015). Araştırma örneklemine ilişkin değerlendirmede, her bir grubun önerilen eşik değerleri karşıladığı görülmektedir. Bu nedenle, örneklem büyüklüğünün YEM için yeterli olduğu anlaşılmaktadır.

### 3.3. Verilerin Elde Edilmesi

TIMSS uygulaması kapsamında, öğrencilerin akademik başarılarının yanı sıra başarılarını etkileyebilecek duyuşsal özelliklerin de ölçülmesi amaçlanmakta ve bu özelliklerin değerlendirilmesi, bağlam anketleri aracılığıyla yapılmaktadır. Bu anketler öğrenci, ev, öğretmen ve program anketlerinden oluşur (Yin ve Fishbein, 2020). TIMSS 2019 uygulamasında, dördüncü sınıf düzeyinde 18, sekizinci sınıf düzeyinde ise 16 ölçek yer almaktadır. Bazı ölçekler her iki sınıf düzeyine ortak uygulanırken,

bazıları yalnızca bir düzeye özgüdür. Bu ölçeklerden biri olan MÖSÖ, 2011 yılından itibaren her iki sınıf düzeyinde kullanılmaktadır. Bu çalışmada kullanılan veriler, TIMSS 2019 uygulamasında sekizinci sınıf öğrencilerinin MÖSÖ'ye verdikleri yanıtları içermektedir. Araştırmada kullanılan MÖSÖ'ye ait maddelere Tablo 3.2.'de yer verilmiştir.

**Tablo 3.2.** Matematik Öğrenmeyi Sevme Ölçeği Maddeleri

Madde Kodu	Madde İfadesi
BSBM16A	Matematik öğrenmekten hoşlanırım.
BSBM16B*	Keşke matematik çalışmak zorunda olmasam.
BSBM16C*	Matematik sıkıcıdır.
BSBM16D	Matematikte birçok ilginç şey öğrenirim.
BSBM16E	Matematiği severim.
BSBM16F	Sayıları içeren okul çalışmalarını severim.
BSBM16G	Matematik problemlerini çözmeyi severim.
BSBM16H	Matematik derslerini sabırsızlıkla beklerim.
BSBM16I	Matematik en sevdiğim derslerden biridir.

*Not: \*Ters kodlanmış maddeler*

Öğrencilerin matematik öğrenmeye yönelik tutumlarını ölçmeyi amaçlayan MÖSÖ, 7'si olumlu ve 2'si olumsuz olmak üzere toplam 9 maddeden oluşmaktadır. Ölçek, "1= Kesinlikle Katılıyorum", "2= Biraz Katılıyorum", "3= Biraz Katılmıyorum" ve "4= Kesinlikle Katılmıyorum" şeklinde dört seçenekli Likert tipi yanıt formatıyla uygulanmaktadır.

### 3.4. Verilerin Analizi

Veri analizi süreci aşamalı olarak yürütülmüştür. İlk olarak, verilerin analiz öncesi gerekli sayıtlılar olan kayıp veriler, uç değerler, normallik ve çoklu bağlantılılık ele alınarak, veriler analiz için uygun hâle getirilmiştir. Ardından, TIMSS 2019 döngüsünde sekizinci sınıf öğrencilerinin katıldığı MÖSÖ ile oluşturulan ölçme modeli DFA ile test edilerek doğrulanmıştır. Üçüncü aşamada, doğrulanan MÖSÖ'nün modeli ÇGDFA yöntemiyle incelenmiş ve modelin araştırma kapsamında belirlenen ülkelere göre ölçme değişmezliği analiz edilmiştir. Dördüncü aşamada ise, ÇGDFA'da olduğu gibi herhangi bir önsel tanımlaması yapılmadan Bayes yöntemi kullanılarak kesin ölçme değişmezliği değerlendirilmiştir. Son aşamada, faktör yükleri ve gösterge sabitlerine önseller tanımlanarak MÖSÖ ölçme modeli BYÖD yöntemiyle incelenmiştir. Araştırma verilerinin tüm analizleri SPSS 26.0 (IBM Corp, 2019) ve

Mplus 8.5 (Muthén ve Muthén, 2017) yazılımları kullanılmıştır. Sadece çoklu normallik testi, Erciyes Üniversitesi tarafından geliştirilen ve çevrim içi olarak erişilebilen MVN uygulaması kullanılarak yapılmıştır. Bu uygulama, Mardia testi gibi yöntemlerle verilerin çoklu normalliğe uygunluğunu değerlendirmektedir (Korkmaz vd., 2014).

### 3.4.1. Kayıp Veri

Veri setinde ele alınan ilk inceleme, kayıp verilerin bulunup bulunmadığıdır. Uygulama sırasında ya da verilerin sisteme aktarımı esnasında yaşanan teknik sorunlar nedeniyle veri kaybı yaşanabilir. Bu tür verilerin rastgele dağılmış olması ve belirli bir örüntü içermemesi istenen bir durumdur. Bu nedenle analize başlamadan önce kayıp verilerin sistematik (örüntülü) bir kayıp, gösterip göstermediği mutlaka değerlendirilmelidir. Çünkü sistematik kayıplar, sayıları az olsa bile analiz sonuçlarının genellenebilirliğini olumsuz etkileyebilir (Tabachnick ve Fidell, 2007). Kayıp verileri ele almak için üç temel yöntem bulunmaktadır: verilerin silinmesi, yaklaşık değer atama ve parametre tahminidir (Çokluk vd., 2021). Hangi yöntemin seçileceği ise örneklem büyüklüğüne ve değişkenlerin özelliklerine bağlı olarak değişmektedir. MÖSÖ, bireylerin duyuşsal özelliklerini ölçen bir araç olduğundan, eksik verilerin tamamen rastgele (MCAR) olduğu varsayılmıştır. Bu doğrultuda, eksik verilere değer atamak için Beklenti Maksimizasyon (Expectation-Maximization; EM) yöntemi kullanılmıştır. Ülke bazında silinen ve kalan veri sayıları Tablo 3.3'te sunulmuştur.

**Tablo 3.3.** Araştırma Grubunu Oluşturan Ülkelerin Silinen ve Kalan Veri Sayıları

Ülkeler	Veri Girişi Yapılmayan	Kayıp Veri Girişi Yapılan	Rastgele İşaretlenmiş	Toplam Silinen	%	Kalan
Kuveyt	45	39	98	182	4	4.392
Portekiz	14	15	9	38	1,1	3.339
S. Arabistan	5	41	37	83	1,4	5.597
Singapur	1	2	38	41	0,8	4.812
Türkiye	29	70	31	130	3,2	3.947
Tayvan	2	1	53	56	1,1	4.859
Toplam	96	168	256	530	1,9	26.946

Tablo 3.3. incelenmesi sonucunda, MÖSÖ'ye ait tüm maddelere hiçbir veri girişi yapılmadığı tespit edilen altı ülkeye ait toplam 96 veri silinmiştir. Benzer şekilde, tüm maddelere yalnızca kayıp veri kodu (9) girildiği belirlenen altı ülkeye ait toplam 168 veri de silinmiştir. Ayrıca, MÖSÖ'de ikinci ve üçüncü maddelerin ters kodlanmış olmasına rağmen, tüm maddelere aynı şekilde (örneğin 1-1 şeklinde) işaretleme yapılan verilerin rastgele doldurulduğu varsayılmış ve bu nedenle altı ülkeye ait toplam 256 veri silinmiştir. Tüm bu silme işlemlerinin ardından, güncel veri sayıları Kuveyt (4.392), Portekiz (3.339), Suudi Arabistan (5.597), Singapur (4.812), Türkiye (4.947) ve Tayvan (4.859) olarak belirlenmiştir. Böylece araştırmaya toplamda 26.946 veri ile devam edilmiştir.

### 3.4.2. Uç Değer

Kayıp verilerin silme ve atama işlemi yapıldıktan sonra kalan verilerdeki uç değerlerin tespit edilmesine geçilmiştir. Uç değerler, veri setinde yer alan genel yapıdan farklı olan verilerdir (Kline, 2016). Veri setlerindeki uç değerler tek değişkenli ve çok değişkenli uç değerler olmak üzere iki farklı şekilde belirlenmiştir. Tek değişkenli uç değerlerin belirlenmesi için veri setinde yer alan puanlar  $z$  standart puanlarına dönüştürülmüş ve -3 ile +3 aralığı dışındaki puanlar, uç değer olarak değerlendirilmiştir (Kline, 2016). Çok değişkenli uç değerlerin belirlenmesi için ise Mahalonobis uzaklıkları hesaplanmıştır. Mahalonobis uzaklığı, her bireyin diğer bireyler için hesaplanan tüm değişkenlerinin ortalamasının kesişim noktası olan ağırlık merkezine uzaklığıdır (Tabachnick ve Fidell, 2001). Analiz sonucunda bir verinin uç değer olması, o veriyi silmek için yeterli bir gerekçe değildir. Eğer bu değerlerin hatalı veri girişi gibi nedenlerle oluştuğu belirlenmiş ise silinmeleri uygun olur. Ancak uç değerlerin silinmesi analiz sonuçları olumsuz etkiliyorsa, verileri silmek doğru bir yaklaşım olmaz. Araştırmada belirtilen yöntem ve ilgili sınır değerleri göz önünde bulundurulmuştur.

### 3.4.3. Normallik

Normal dağılım değişken sayısına göre tek değişkenli normal dağılım ve çok değişkenli normal dağılım şeklinde sınıflandırılır. Tek değişkenli normallik, tek bir değişkene ait şeklin normalliği ile ilgili bilgi verirken, çok değişkenli normallik ise değişkenlerin tüm kombisyonları açısından normal dağılım göstermesi anlamına

gelmektedir (Çokluk, vd., 2021). Normal dağılımı değerlendirmenin grafiksel ve istatistiksel birçok yolu bulunmaktadır (Çokluk, vd., 2021). DFA ve ÇGDFA analizlerinde kullanılacak kestirim yönteminin belirlenmesi amacıyla, tek değişkenli normallik çarpıklık ve basıklık değerleri üzerinden; çok değişkenli normallik ise Mardia testi aracılığıyla değerlendirilmiştir (Mardia, 1970). George ve Mallery (2010) ile Tabachnick ve Fidell (2007), çarpıklık ve basıklık katsayılarının -1,5 ile +1,5 arasında bulunması durumunda veri setinin tek değişkenli normallik varsayımını sağlandığı belirtmektedir. Çok değişkenli normallik ise Mardia testi sonuçlarının anlamlılık durumuna göre değerlendirilmektedir.

#### 3.4.4. Çoklu Bağlantı Problemi

Çoklu bağlantı problemi (multicollinearity), değişkenler arasındaki ilişkinin güçlü ve yüksek olması durumudur. Böyle bir problem varsa, nedeni olan değişkenlerin analizden çıkarılması önerilir (Tabachnick ve Fidell, 2007). Çoklu bağlantı probleminin incelenmesi amacıyla kullanılan bazı yöntemler vardır (Çokluk vd., 21). Bunlardan biri, her bir ölçekteki maddeler için Varyans Artış Faktörü (VIF), Tolerans Değeri (TD) ve Koşul İndeksleri (CI) incelenmesidir. Çoklu bağlantı probleminin bulunmadığı varsayımının sağlanması için, VIF değerinin 10'dan küçük, TD'nin .10'dan büyük, CI değerinin ise 30'dan küçük olması gerekmektedir (Tabachnick ve Fidell, 2007). Çoklu bağlantı varsayımının incelenmesinde kullanılan bir başka yöntem ise ölçek maddeleri arasındaki korelasyon değerleridir. Bu değerlerin, 0,90'ı geçmemesi gerektiği belirtilmektedir (Tabachnick ve Fidell, 2007). Bu doğrultuda, araştırmada her iki yöntemle elde edilen veriler incelenmiş ve ilgili sınır değerler göz önünde bulundurulmaktadır.

#### 3.5. Doğrulayıcı Faktör Analizi Uygulaması

Ölçme değişmezliği çalışmaları, her grup için ayrı ayrı DFA yapılarak başlatılmıştır. Normallik testlerinden elde edilecek sonuçlar doğrultusunda ise uygun kestirim yöntemi belirlenmiştir. DFA modelinden elde edilen faktör yükleri kullanılarak, YEM'e dayalı geçerlik ve güvenilirlik kanıtları sunmak amacıyla bileşik güvenilirlik (CR) ve çıkarılan varyans ortalaması (AVE) değerleri hesaplanabilir (Şen, 2023). Hair vd.'ye (2014) göre, AVE değerinin 0,50'nin, CR değerinin ise 0,70'in üzerinde olması kabul edilebilir düzeydir. Ayrıca, ölçeğin iç tutarlılığını

değerlendirmek amacıyla  $a$  katsayısı hesaplanmıştır. Yapılan tüm analizler, araştırmaya dahil edilen her ülke için ayrı ayrı hesaplanmış ve bulgular bölümünde sunulmuştur. DFA analizinde modelin veriyle uyumunu değerlendirmek için çeşitli uyum indeksleri kullanılmaktadır. İyi uyum ölçütleri,  $\chi^2$  değerinin manidar olmaması ( $p > 0,05$ ), RMSEA'nın 0,05'in altında, CFI ve TLI değerlerinin 0,95'in üzerinde ve SRMR'nin ise 0,08'in altında olması beklenir. Kabul edilebilir uyum ölçütleri ise RMSEA'nın 0,05 ile 0,08 arasında, CFI ve TLI değerinin 0,90 ile 0,95 arasında ve SRMR'nin 0,08 ile 0,10 arasında bulunmasıdır (Hu ve Bentler, 1999; Kline, 2016; Şen, 2023).

### 3.6. Çoklu Grup Doğrulayıcı Faktör Analizi Uygulaması

DFA ile doğrulanmış modelin ardından, ÇGDFA ile ülkelere göre ölçme değişmezliği incelenmiştir. Ölçme değişmezliği araştırmalarında model uyumu kritik bir öneme sahiptir. ÇGDFA, kısıtlanmış ve kısıtlanmamış modeller arasındaki uyum iyiliği indeksleri karşılaştırılarak ölçme değişmezliği test edilir (Cheung ve Rensvold, 2002). Literatürde, ölçme değişmezliğini değerlendirmeye yönelik farklı yaklaşımlar geliştirilmiştir. Bu yaklaşımlar genellikle  $\chi^2$ , CFI ve RMSEA değerlerindeki değişimlere odaklanır (Cheung ve Rensvold, 2002). Bu değişiklikler sırasıyla  $\Delta\chi^2$ ,  $\Delta CFI$  ve  $\Delta RMSEA$  olarak temsil edilir. Cheung ve Rensvold (2002) ve Chen (2007) göre büyük örneklem analiz edildiği durumlarda  $\chi^2$  testi yerine uyum indeksleri arasındaki farka bakılması önerilmektedir.

Yukarıda belirtildiği üzere, TIMSS ve PISA gibi büyük ölçekli uluslararası değerlendirmelerde,  $\chi^2$  istatistiğinin model uyumsuzluğuna duyarlılığı nedeniyle genel uyum kriteri olarak uygun bulunmamış, bu nedenle bu araştırmada  $\chi^2$  istatistiği dikkate alınmamıştır. Bunun yerine, metrik ve skalar değişmezliği değerlendirilmesinde CFI ve RMSEA değerlerindeki fark esas alınmıştır. Chen (2007) metrik ve skalar değişmezliğin sağlandığının kabul edilebilmesi için  $\Delta CFI$ 'nin 0,010 veya daha küçük,  $\Delta RMSEA$ 'nin 0,015 veya daha küçük olması gerektiğini belirtmiştir. Rutkowski ve Svetina (2014) ise TIMSS ve PISA gibi büyük örneklem içeren çalışmalarda, metrik ve skalar değişmezliğin sağlanabilmesi için  $\Delta CFI$ 'nin 0,020'nin altında ve  $\Delta RMSEA$ 'nin 0,030'un altında olması gerektiğini vurgulamaktadır. Bu iki farklı görüşten de anlaşılacağı üzere, indeksler arasındaki farkın ne kadar olması

konusunda kesin ölçütler mevcut değildir (Şen, 2023). Bu araştırmada ölçme değişmezliği değerlendirmesi, Chen (2007), Rutkowski ve Svetina (2014) ile Şen (2023) tarafından önerilen kriterler doğrultusunda gerçekleştirilmiştir.

### 3.7. Bayes Yaklaşık Ölçme Değişmezliği Uygulaması

ÇGDFA yöntemiyle kesin ölçme değişmezliği incelendikten sonra, parametrelerde herhangi bir önsel kullanılmadan, Bayes yöntemiyle kesin ölçme değişmezliği değerlendirilmiştir. Bu değerlendirme, ÇGDFA süreciyle benzer bir yol izlenerek gerçekleştirilmiştir. Ancak modelin veri uyumunu belirlemede farklı uyum iyiliği ölçütleri kullanılmıştır. Muthén ve Asparouhov (2012) ile Van de Schoot vd.'ye (2013) göre, Bayes analizinde, bir modelin veriye iyi uyduğunu söyleyebilmek için iki temel koşul dikkate alınmalıdır. Bunlardan ilki, PPP'nin anlamlı olmamalıdır; yani aşırı uç değerlerde yer almamalı ve orta değerlere yakın olmalıdır. İkincisi,  $\chi^2$  farkı gibi uyum ölçütlerinin CI'nın sıfır değerini içermelidir; bu durum, gözlenen ve simüle edilen değerler arasında anlamlı bir fark olmadığını ve dolayısıyla modelin veriye uyumlu olduğunu gösterir. DIC ise, model karşılaştırılmasında önemli bir rol oynar. DIC hem modelin veriye uyumunu hem de karmaşıklığını birlikte ele alır. Daha düşük DIC değeri, görece olarak daha iyi uyum sağlayan ve daha az karmaşık bir modeli işaret eder. Modelin bu üç koşulun sağlanması, modelin gözlemlenen veriyi yeterince iyi temsil ettiğine dair güçlü ve güvenilir bir değerlendirme yapılabilir.

Son olarak, BYÖD yönteminde kesin ölçme değişmezliği için kullanılan tam eşitlik kısıtlamaları yerine, sıfır ortalamalı ve küçük varyanslı önseller kullanarak yaklaşık eşitlik kısıtlamaları uygulanmıştır. İlk olarak, varyansı 0,005 olan bir modelle başlanmıştır; modelin uyumu yetersizse, faktör yükleri ve gösterge sabitleri arasındaki farklılıkları tolere edebilmek için varyansı sırasıyla 0,010; 0,050 ve 0,10'a kadar kademeli olarak artırılmıştır. Süreç boyunca model uyumu (DIC, PPP ve gözlemlenen-tekrarlanan  $\chi^2$  farklarının %95 güven aralıkları) ve yakınsama durumu takip edilmiştir. Modeller arasında anlamlı uyum farkları önemsiz hale geldikten sonra, nihai model bu kriterler doğrultusunda değerlendirilmiştir.

Bayes kesin ve yaklaşık ölçme değişmezliği analizlerinde ardışık bir model test yaklaşımı benimsenmiştir. Frekansçı yöntemlerden ÇGDFA'da olduğu gibi yapısal, metrik ve skalar olmak üzere üç düzeyde ölçme değişmezliği incelenmiştir. ÇGDFA

yöntemi ile elde edilen sonuçlarla tutarlılığı değerlendirmek amacıyla, Bayes tahminleme kapsamında kesin değişmezlik analizleri gerçekleştirilmiştir. Bu analizlerde, araştırmacı tarafından faktör yükleri ve madde sabitlerine ilişkin herhangi bir önsel tanımlama yapılmamış; bunun yerine Mplus yazılımının varsayılan önsel değerleri kullanılmıştır (örneğin, faktör yükü için  $\mu = 0, \sigma^2 = 10^{10}$  olan normal dağılım; Gucciardi vd., 2016). Modeller, Gibbs örnekleyicisi ile MCMC simülasyon prosedürleri kullanılarak uygulanmıştır. Dört MCMC zinciri için 150.000 yineleme belirlenmiş ve ilk yarısı varsayılan olarak “ısınma aşaması” (burn-in phase) olarak değerlendirilmiştir. Model yakınsaması, Potansiyel Ölçek Azaltma Faktörü ( $PSR < 1.1$ ; Asparouhov ve Muthén, 2010) gibi istatistiksel ölçütlerle sınanmış ve böylece birden fazla zincirin benzer hedef dağılıma yakınsadığı doğrulanmıştır (Van de Schoot vd., 2014). Bayes kesin ölçme değişmezliği analizlerinde faktör yükü ve madde sabiti için varsayılan önseller kullanılmıştır. Modelin tanımlanabilirliğini sağlamak amacıyla faktör varyansı 1'e, faktör ortalaması ise 0'a sabitlenmiştir. Artık varyanslar için Inverse-Wishart dağılımı IW (1;15), artık kovaryans için ise IW (0;15) önselleri atanmıştır. Benzer şekilde, BYÖD analizlerinde de artık varyans ve artık kovaryans aynı önseller kullanılmıştır. Bunun yanında, faktör yükleri ve madde sabitleri için sıfır ortalama ve farklı varyans düzeylerinde tanımlanmış Normal dağılım önselleri ( $\sigma^2 = 0,005; 0,01; 0,05$  ve  $0,10$ ) kullanılarak dört farklı yaklaşım düzeyi değerlendirilmiştir.

#### 4. BULGULAR

Bu bölümde, araştırmanın amaçları doğrultusunda TIMSS 2019 uygulamasında yer alan MÖSÖ'den elde edilen veriler analiz edilmiştir. İlk aşamada, analiz sürecine ilişkin varsayımlar (kayıp veri durumu, uç değerlerin kontrolü, normallik dağılımı ve çoklu bağlantı problemi) ile ölçek maddelerine ait betimsel istatistiksel bilgiler sunulmuştur. Devamında, araştırmaya dahil edilen altı ülkeye yönelik DFA sonuçları paylaşılmıştır. Son olarak, araştırma soruları doğrultusunda iki farklı yöntem olan ÇGDFA ve BYÖD modelleri kullanılarak gerçekleştirilen ölçme değişmezliği analizlerine ilişkin bulgulara yer verilmiştir. Bu kapsamda, her iki yöntemin analiz sonuçları ayrıntılı biçimde raporlanmıştır.

##### 4.1. Analiz Varsayımları

Veri seti üzerinde yapılan inceleme sonucunda, MÖSÖ'ye ait tüm maddelere veri girişi yapılmayan, tüm maddeleri kayıp veri kodu (9) ile girilen ve ters kodlanmış maddelere rağmen rastgele yanıtladığı düşünülen altı ülkeye ait toplam 256 veri silinmiştir. Sonrasında, bazı maddelere yanıt girilmiş iken bazıları kayıp veri kodu ile kodlanmış olan veriler için EM yöntemi kullanılarak eksik verilere değer ataması yapılmıştır. Tek değişkenli uç değerlerin belirlenmesinde  $z$  standart puanlar hesaplanmış ve tüm ülkelere ait değerler -2,04 ile +2,04 aralığında yer aldığı görülerek uç değer bulunmadığı sonucuna ulaşılmıştır (Kline, 2016). Çok değişkenli uç değerler için Mahalonobis uzaklıkları hesaplanmış, ancak yüksek sayıda uç değer tespiti model uyumunu olumsuz etkileyebileceği gerekçesiyle bu veriler silinmemiştir. Tek değişkenli normallik analizinde çarpıklık ve basıklık değerleri sırasıyla -0,48 ile +1,24 ve -1,53 ile -0,24 arasında değişmiş, bu durum tüm ülkelerde normalliğin sağlandığını göstermiştir (George ve Mallery, 2010; Tabachnick ve Fidell, 2007); buna karşın Mardia çok değişkenli normallik testi sonuçları, hiçbir ülkenin çok değişkenli normalliği sağlamadığını ortaya koymuştur. Son olarak, çoklu bağlantı probleminin incelenmesi kapsamında VIF, T, CI değerleri ve maddeler arası korelasyon analiz edilmiştir. Analiz sonucunda, VIF (1,304–5,175), T (0,178–0,767) ve CI (15,2–18,8) değerleri ile maddeler arası korelasyon katsayılarının (0,270–0,856) elde edildiği ve bu değerlerin tümü kabul edilebilir sınırlar içinde yer aldığı görülmüştür (Tabachnick ve Fidell, 2007). Bu durum, çoklu bağlantı probleminin bulunmadığını göstermektedir.

## 4.2. Betimsel İstatistikler

MÖSÖ’de yer alan maddelere ait betimsel istatistikler Tablo 4.1.’de verilmiştir.

**Tablo 4.1.** MÖSÖ’de Yer Alan Maddelere Ait Betimsel İstatistikler

Madde No		Kuveyt	Portekiz	Suudi Arabistan	Singapur	Türkiye	Tayvan
M1	O	2,16	2,13	2,04	1,9	1,69	2,46
	S.S	1,07	0,99	0,99	0,88	0,90	0,94
	Ç	0,51	0,49	0,64	0,80	1,24	0,08
	B	-0,99	-0,79	-0,65	-0,04	0,70	-0,88
M2	O	2,46	2,14	2,20	2,26	2,49	2,49
	S.S	1,17	1,10	1,14	1,06	1,19	1,02
	Ç	0,02	0,43	0,36	0,28	-0,04	0,08
	B	-1,48	-1,17	-1,32	-1,17	-1,51	-1,10
M3	O	2,54	2,55	2,32	2,28	2,22	2,43
	S.S	1,13	1,02	1,07	0,95	1,12	0,97
	Ç	-0,10	-0,13	0,14	0,17	0,26	0,09
	B	-1,39	-1,10	-1,25	-0,94	-1,36	-0,97
M4	O	2,14	2,14	1,96	1,96	1,8	2,37
	S.S	1,08	0,92	0,99	0,82	0,97	0,90
	Ç	0,48	0,39	0,71	0,57	1,01	0,20
	B	-1,08	-0,71	-0,61	-0,15	-0,05	-0,71
M5	O	2,36	2,26	2,22	2,05	1,85	2,55
	S.S	1,15	1,08	1,10	0,95	1,01	0,97
	Ç	0,24	0,34	0,42	0,58	0,97	-0,04
	B	-1,38	-1,15	-1,16	-0,59	-0,24	-0,96
M6	O	2,40	2,62	2,28	2,46	2,03	2,81
	S.S	1,11	0,93	1,04	0,90	1,05	0,88
	Ç	0,15	-0,18	0,30	-0,00	0,67	-0,40
	B	-1,32	-0,82	-1,10	-0,78	-0,78	-0,51
M7	O	2,35	2,49	2,23	2,24	2,05	2,58
	S.S	1,13	1,03	1,06	0,97	1,03	0,96
	Ç	0,24	0,00	0,38	0,35	0,66	-0,08
	B	-1,33	-1,15	-1,09	-0,85	-0,74	-0,95
M8	O	2,62	2,67	2,49	2,37	2,46	2,78
	S.S	1,13	0,97	1,08	0,94	1,07	0,91
	Ç	-0,10	-0,21	0,07	0,16	0,15	-0,37
	B	-1,40	-0,93	-1,28	-0,86	-1,21	-0,63
M9	O	2,59	2,74	2,44	2,31	2,26	2,85
	S.S	1,20	1,20	1,17	1,12	1,16	0,98
	Ç	-0,08	-0,31	0,11	0,23	0,36	-0,48
	B	-1,53	-1,46	-1,48	-1,33	-1,33	-0,78

**Not:** O=Ortalama; S.S=Standart Sapma; Ç=Çarpıklık; B=Basıklık.

Tablo 4.1. incelendiğinde, ölçek maddelerine ilişkin ortalama değerlerin 1,69 ile 2,85; standart sapma değerlerinin ise 0,82 ile 1,19 arasında değiştiği görülmektedir.

Çarpıklık katsayıları -0,48 ile +1,24 arasında, basıklık katsayıları ise -1,53 ile -0,24 arasında değişim göstermiştir.

### 4.3. Doğrulayıcı Faktör Analizi Sonuçları

Analiz öncesinde varsayımlar değerlendirilmiş ve uygunlukları doğrulandıktan sonra, araştırmaya dahil edilen her ülkeye ait veriler kullanılarak DFA uygulanmış; bu kapsamda oluşturulan ölçme modelinin geçerliğini değerlendirmek amacıyla model uyum iyiliği indeksleri hesaplanarak Tablo 4.2.'de verilmiştir.

**Tablo 4.2.** Ölçme Modeline İlişkin Ülkelerin DFA Model Uyum İyiliği İndeksleri

Ülkeler	RMSEA (%90 GA)	SRMR	CFI	TLI	$\chi^2$	<i>df</i>	<i>p</i>
Kuveyt	,064 (,059-,069)	,027	,973	,964	509,35	27	,000
Portekiz	,085 (,080-,091)	,024	,964	,952	678,19	27	,000
S.Arabistan	,079 (,075-,083)	,033	,958	,945	965,35	27	,000
Singapur	,075 (,071-,080)	,025	,967	,956	764,05	27	,000
Türkiye	,086 (,081-,091)	,033	,948	,931	819,09	27	,000
Tayvan	,089 (,084-,093)	,032	,954	,939	1062,73	27	,000

Tablo 4.2. incelendiğinde, tüm ülkelerde elde edilen  $\chi^2$  değerlerinin istatistiksel olarak anlamlı olduğu görülmektedir ( $p < 0,05$ ). Bu durum, modellerin verilerle iyi uyum göstermediğini düşündürse de  $\chi^2$  testinin örneklem büyüklüğüne duyarlılığı göz önünde bulundurulduğunda, modelin genel uyumunu değerlendirmek için diğer uyum indekslerine odaklanmak daha anlamlı olacaktır (Şen, 2023). Nitekim RMSEA, CFI, TLI ve SRMR gibi alternatif uyum indeksleri tüm ülkelerde kabul edilen eşik değerler içerisinde ya da bu değerlere oldukça yakın olduğu görülmektedir. Bu bulgular,  $\chi^2$  testi dışında kalan uyum ölçütlerinin modelin veriye yeterli düzeyde uyum sağladığını göstermektedir. Dolayısıyla, söz konusu modele ait parametrelerin güvenle yorumlanabileceği söylenebilir.

DFA'ya ait faktör yükleri kullanılarak hem geçerlik hem de güvenilirlik düzeyleri hesaplanmıştır. Bu kapsamda, güvenilirliğe ilişkin kanıtlar CR değerleriyle; yakınsak geçerliğe ilişkin kanıtlar ise AVE değerleriyle ortaya konmuştur. Hair vd. (2014) tarafından belirtildiği üzere, AVE değerlerinin 0,50'nin, *a* ve CR değerlerinin ise 0,70'in üzerinde olması kabul edilebilir düzeyde geçerlik ve güvenilirlik

sağlandığını göstermektedir. Araştırmada yer alan tüm ülkelere ait standartlaştırılmış faktör yükleri, AVE, CR ve  $\alpha$  hesaplanarak Tablo 4.3’de verilmiştir.

**Tablo 4.3.** MÖSÖ’ye Ait Faktör Yükleri, Güvenirlik Katsayıları ve AVE Değerleri

Ülkeler	Maddelerin Faktör Yükü									$\alpha$	CR	AVE
	M1	M2	M3	M4	M5	M6	M7	M8	M9			
Kuveyt	,80	,44	,63	,75	,89	,75	,82	,83	,86	,92	,93	,59
Portekiz	,88	,71	,76	,74	,92	,79	,86	,85	,84	,95	,95	,67
Suudi Arabistan	,79	,53	,71	,74	,90	,77	,84	,84	,87	,93	,93	,61
Singapur	,87	,73	,72	,67	,91	,72	,85	,77	,88	,94	,94	,63
Türkiye	,78	,61	,73	,54	,88	,71	,80	,81	,87	,92	,92	,57
Tayvan	,90	,67	,75	,80	,93	,83	,88	,74	,85	,95	,95	,67

Tablo 4.3. incelendiğinde, tüm ülkelerde faktör yüklerinin 0,44 ile 0,92 aralığında değerler aldığı görülmektedir. Ayrıca,  $\alpha$ , CR ve AVE değerlerinin ilgili eşik düzeyleri aştığı tespit edilmiştir. Bu doğrultuda, analiz kapsamındaki tüm ülkelerde yakınsak geçerlik ve güvenilirlik açısından yeterli istatistiksel kanıtların bulunduğu ifade edilebilir. DFA sonucunda ölçme modeli tüm ülkelerde doğrulandıktan sonra, ölçme değişmezliğini belirlemek amacıyla ÇGDFA’ya geçilmiştir.

#### 4.4. Birinci Alt Amaca Yönelik Bulgu ve Yorumlar

Araştırmanın birinci alt amacı, TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan MÖSÖ’nün örtük yapısı, ÇGDFA yöntemine göre belirlenen ülkeler arasında ölçme değişmezliğini incelemektir. Bu kapsamda, ÇGDFA yöntemi ile ölçme değişmezliği aşamaları (yapısal, metrik ve skalar) test edilmiştir. Her bir aşamaya ilişkin model uyum indeksleri Tablo 4.4.’te verilmiştir.

**Tablo 4.4.** Ölçme Değişmezliği Modeline İlişkin Ülkelerin ÇGDFA Yöntemine Göre Model Uyum İyiliği İndeksleri (MLR Kestirimi)

Model	RMSEA (CI)	SRMR	CFI	TLI	$\chi^2$	df	p	$\Delta CFI$	$\Delta RMSEA$
Yapısal	,080 (,078-,082)	,029	,961	,948	4847	162	,00	-	-
Metrik	,084 (,082-,086)	,065	,946	,943	6613	202	,00	,015	,004
Skalar	,096 (,094-,097)	,074	,916	,925	10231	242	,00	,030	,012

Tablo 4.4. incelendiğinde, parametre kısıtlaması bulunmayan yapısal değişmezlik modeli aşamasında elde edilen model uyum iyiliği indekslerinin kabul edilen eşik değerler içerisinde yer aldığı görülmektedir. Bu durum,  $\chi^2$  testi dışında kalan uyum ölçütlerinin model veriye yeterli düzeyde uyum sağladığını göstermektedir. Bu bulgular doğrultusunda, ülkeler arasında aynı yapının aynı maddelerle ölçüldüğü sonucuna ulaşılmaktadır. Yapısal değişmezliğin sağlandığına dair kanıt elde edilmesiyle birlikte, analiz süreci metrik değişmezlik aşamasına geçilerek devam edilmiştir.

Metrik değişmezlik modeli aşamasında elde edilen model uyum iyiliği indekslerinin SRMR, CFI ve TLI açısından kabul edilen eşik değerler içerisinde, RMSEA açısından ise bu değerlere oldukça yakın olduğu görülmektedir. Yapısal ve metrik değişmezlik modelleri arasındaki CFI farkı 0,015, RMSEA farkı ise 0,004 olarak hesaplanmıştır. Bu değerler ÇG DFA uygulaması bölümünde belirtilen eşik değerler içerisinde ya da bu değerlere oldukça yakın düzeydedir. Yapısal değişmezlikten, metrik değişmezliğe geçişte model uyum indekslerinde anlamlı bir bozulma gözlenmediğinden metrik değişmezliğin sağlandığı ve ülkeler arası faktör yüklerinin eşit olduğu söylenebilir. Bu durum, örtük değişkenin (matematiği sevmeye) tüm ülkelerde aynı yapıyı temsil ettiğini ve örtük değişkende meydana gelen her bir birimlik değişimin, göstergelerdeki puanları tüm ülkelerde eşit düzeyde etkilediğini varsaymamıza olanak tanımakta olduğunu söyleyebiliriz. Metrik değişmezliğin sağlandığına dair elde edilen bulgular sonrasında analiz süreci skalar değişmezlik aşamasına geçilerek devam edilmiştir.

Skalar değişmezlik modeli aşamasında, elde edilen model uyum iyiliği indekslerinden SRMR, CFI ve TLI kabul edilen eşik değerler içerisinde yer alırken, RMSEA değeri ise bu eşik değerinin dışında kalmaktadır. Metrik ve skalar değişmezlik modelleri arasındaki CFI farkı 0,030, RMSEA farkı ise 0,012 olarak hesaplanmıştır. Bu değerler, CFI farkının belirlenen eşik değeri aşması, RMSEA'nın kabul edilebilir düzeyde olmaması ve metrik değişmezlikten skalar değişmezliğe geçişte, genel uyum indekslerinde kayda değer bir düşüş olması sebebiyle skalar değişmezliğin sağlanmadığını göstermektedir. Bu durum, ülkeler arasında gösterge sabitlerinin eşit olmadığı ve örtük değişkenin tüm ülkelerde aynı şekilde kullanılmadığı varsayılır (Davidov vd., 2014; Vanderberg ve Lance, 2000). Başka bir ifadeyle, bazı ülkelerin

katılımcıları, diğer ülkelere kıyasla maddelere sistematik olarak daha yüksek ya da daha düşük yanıt verme eğiliminde olduğu söylenebilir. Ülkelerin ölçme modeline ilişkin kesin değişmezlik durumu, ÇGDFA yöntemi kullanılarak incelenmiş ve değerlendirilmiştir. Bu bulguları desteklemek amacıyla, aynı ölçme modeline yönelik olarak Bayes kesin değişmezlik (herhangi bir önsel kullanılmadan) yöntemiyle gerçekleştirilen analizlere ait her bir aşamanın model uyum indeksleri Tablo 4.5.'te sunulmuştur. Bayes yöntemi, Bayes uyum indeksleri bölümünde yer verildiği üzere, model uyumu değerlendirmesinde PPP ve %95 CI değerleri ile belirlenmiştir.

**Tablo 4.5.** Ölçme Değişmezliği Modeline İlişkin Ülkelerin Bayes Yöntemine Göre Kesin Değişmezlik Model Uyum İyiliği İndeksleri

Model	PPP	CI	DIC
Yapısal	,503	(-70,80-70,18)	518820
Metrik	,310	(-53,99-90,162)	518842
Skalar	,000	(4722,60-4863,699)	523569

Tablo 4.5. incelendiğinde, yapısal değişmezlik modeli aşamasında elde edilen model uyum iyiliği indekslerinin PPP değerinin 0,503 olması ve güven aralığı sıfır içermesi, model uyumu sağladığını göstermektedir. Yapısal değişmezliğin sağlandığına dair kanıt elde edilmesiyle birlikte, analiz süreci metrik değişmezlik aşamasına geçilerek devam edilmiştir.

Metrik değişmezlik modeli aşamasında, elde edilen model uyum iyiliği indekslerinin PPP değerinin 0,310 olması ve güven aralığı sıfırı içermesi, model uyumun sağladığı anlamına gelmektedir. Metrik değişmezliğin sağlandığına dair elde edilen kanıt doğrultusunda analiz süreci skalar değişmezlik aşamasına geçilerek devam edilmiştir.

Skalar değişmezlik modeli aşamasında, elde edilen model uyum iyiliği indekslerinin PPP değerinin 0,000 olması ve güven aralığının sıfırı içermemesi, modelin bu düzeyde uyum sağlamadığını göstermektedir.

Bayes yöntemiyle herhangi bir önsel tanımlama yapılmaksızın (kesin değişmezlik) elde edilen bu bulgular, ÇGDFA yöntemiyle ulaşılan bulgular ile uyumaktadır. Sonuç olarak, araştırmaya dahil edilen ülkelerle oluşturulan ölçme

modeli yapısal ve metrik düzeyde ölçme değişmezliği sağlarken, skalar düzeyde ölçme değişmezliği sağlamamaktadır.

#### 4.5. İkinci Alt Amaca Yönelik Bulgu ve Yorumlar

Araştırmanın ikinci alt amacı, TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan MÖSÖ'nün örtük yapısı, BYÖD yöntemine göre belirlenen ülkeler arasında ölçme değişmezliğini incelemektir. Bu kapsamda, BYÖD yöntemi kullanılarak gruplar arasında faktör yükleri ve gösterge sabitleri, önceden tanımlanmış sifıra yakın önsel varyans değerleri (0,005; 0,01; 0,05; 0,10) incelenmiş ve elde edilen bulgular Tablo 4.6.'te sunulmuştur. En uygun BYÖD modelinin belirlenmesi amacıyla, farklı önsel varyans değerlerine göre kestirilen modeller karşılaştırılmış; bu modellerin veri uyum düzeyleri DIC, PPP ve %95 güven aralığı ölçütleri üzerinden değerlendirilmiştir.

**Tablo 4.6.** Ölçme Değişmezliği Modeline İlişkin Ülkelerin BYÖD Yöntemine Göre Model Uyum İyiliği İndeksleri

Model	Önsel $\lambda$	Önsel $\vartheta$	PPP	CI	DIC
Metrik	,005	-	,481	(-68,77-72,90)	518824
Metrik	,01	-	,501	(-70,26-71,19)	518822
Metrik	,05	-	,507	(-70,71-70,62)	518821
Metrik	,10	-	,507	(-70,79-70,62)	518821
Skalar	-	,005	,000	(110,29-278,316)	519012
Skalar	-	,01	,006	(22,56-180,27)	518921
Skalar	-	,05	,087	(-22,26-127,26)	518874
Skalar	-	,10	,094	(-24,48-125,26)	518872
Metrik+Skalar	,05	,005	,000	(59,10-218,84)	518958
Metrik+Skalar	,05	,01	,101	(-25,86-123,34)	518869
Metrik+Skalar	,05	,05	,479	(-68,75-73,17)	518823
Metrik+Skalar	,05	,10	,502	(-70,38-71,23)	518822
Metrik+Skalar	,01	,005	,000	(60,27-220,07)	518959
Metrik+Skalar	,01	,01	,097	(-25,30-124,44)	518870
Metrik+Skalar	,01	,05	,473	(-68,32-74,143)	518824
Metrik+Skalar	,01	,10	,495	(-69,75-71,82)	518823

**Not:**  $\lambda$  faktör yükünü ve  $\vartheta$  madde sabitini tanımlamaktadır.

Tablo 4.6. incelendiğinde, faktör yükleri açısından (metrik değişmezlik), dört varyans düzeyinin ( $\lambda = 0,005; 0,01; 0,05; 0,10$ ) tamamında model veriye iyi uyum sağladığı görülmektedir. Bu varyanslar arasında, DIC değerinin en küçük olduğu varyans düzeyi  $\lambda = 0,05$  en iyi uyum sağlayan model olarak kabul edilmektedir.

Gösterge sabitleri açısından (skalar değişmezlik), iki varyans düzeyinde ( $\vartheta = 0,05; 0,10$ ) model veriye iyi uyum göstermiş; iki varyans düzeyinde ( $\vartheta = 0,005; 0,01$ ) ise model veriye uymamıştır. Bu bağlamda, DIC değeri en küçük olan varyans düzeyi ( $\vartheta = 0,10$ ), model uyumu açısından en uygun düzey olarak kabul edilmektedir.

Yaklaşık ölçme değişmezliği hem faktör yüklemelerine hem de gösterge sabitleri açısından değerlendirildiğinde, ( $\lambda = 0,05; \vartheta = 0,01; 0,05; 0,10$ ) ve ( $\lambda = 0,01; \vartheta = 0,01; 0,05; 0,10$ ) varyans düzeylerinde modelin veriye uyum sağladığı; buna karşın ( $\lambda = 0,01; 0,05; \vartheta = 0,005$ ) varyans düzeylerinde ise model veri uyumunun sağlanmadığı anlaşılmaktadır. Bu analizde, DIC değeri ( $\lambda = 0,01; 0,05; \vartheta = 0,10$ ) varyansları en iyi uyum sağlayan modeller olarak kabul edilmektedir.

Model veri uyumunun sağlandığı varyans düzeylerinde yaklaşık skalar değişmezliğin sağlandığı yorumu yapılabilir. Bu koşullar altında, ülkeler arasında gösterge sabitlerinin eşit olduğu ve gizil değişkenin tüm gruplarda aynı şekilde kullanıldığı varsayılır. Böylece ölçme aracının katılımcılar tarafından benzer şekilde kullanıldığı kabul edilir ve aynı yapının tutarlı biçimde ölçüldüğü ileri sürülür (Şen, 2023). Skalar değişmezliğinin sağlanması, gözlenen değişkenlerdeki ortalama farklarının gizil değişkenin ortalama düzeyindeki farklılıklardan kaynaklandığını göstermektedir (Finch ve French, 2015; Steinmetz vd., 2009). Dolayısıyla araştırmaya dahil edilen ülkelerde, faktör (matematik öğrenmeyi sevme) ortalamasının farklılık göstermesi, bu farkın ölçme aracının farklı çalışmasından değil, gerçekten matematik öğrenmeyi sevme düzeyindeki değişimden kaynaklandığını güvenle söylenir.

## 5. TARTIŞMA

Literatür incelendiğinde, TIMSS verileri kullanılarak cinsiyet, kültürlerarası, bölgeler arası gibi değişkenler üzerinden ÇGDFA yöntemiyle ölçme değişmezliği çalışmalarının gerçekleştirildiği görülmektedir. Bu çalışmaların bazılarında ölçme değişmezliği sağlanmış (Atılgan, 2022; Ertürk ve Erdinç Alan, 2018; Polat, 2019; Uyar, 2021; Yiğiter, 2023), bazılarında ise ölçme değişmezliği sağlanmamıştır (Dokur, 2023; Koçak, 2024; Ölçüoğlu ve Çetin, 2016; Polat, 2019; Yiğiter, 2024). Genel olarak değerlendirildiğinde, değişkenlerin grup sayısı iki olduğunda (örneğin cinsiyet), ÇGDFA yöntemiyle ölçme değişmezliği sağlanabilmekte; ancak grup sayısının ikiden fazla olduğu durumlarda (örneğin kültürlerarası karşılaştırmalar) ölçme değişmezliği sağlanamamaktadır. Bu bağlamda, araştırmada ÇGDFA yöntemiyle ölçme değişmezliğinin sağlanamaması, literatürdeki sonuçlarla örtüşmektedir.

TIMSS verileriyle BYÖD yöntemiyle yapılan araştırmalar oldukça sınırlı olmakla birlikte literatürde farklı disiplinlerde kültürlerarası karşılaştırmalarda kullanıldığı görülmektedir. Bu yöntemi kullanan bazı çalışmalar yaklaşık skalar düzeyde ölçme değişmezliğini sağlamış (Cieciuch ve Davidov, 2018; Cieciuch vd., 2014; Gucciordi vd., 2016; Sırgancı, 2022; Sideridis vd., 2020; Wang, 2017; Winter ve Depaoli, 2020), bazıları ise bu düzeyde ölçme değişmezliği sağlayamamıştır (Bagheri vd., 2022; Stevanovic, 2021; Wurster, 2022). Genel olarak değerlendirildiğinde, BYÖD yöntemiyle gerçekleştirilen çalışmaların büyük çoğunluğunda yaklaşık ölçme değişmezliğinin sağlandığı görülmektedir. Bu doğrultuda, çalışmamızda BYÖD yöntemiyle yaklaşık ölçme değişmezliğinin sağlanmış olması, mevcut literatürle tutarlılık göstermektedir.

Bu çalışmada, öğrencilerin matematik öğrenmeye yönelik tutumlarının karşılaştırılabilirliğini test etmek amacıyla kesin değişmezlik için ÇGDFA ve yaklaşık değişmezlik için BYÖD yöntemleri kullanılmıştır. TIMSS gibi geniş ölçekli uluslararası sınavlarda değişkenlerin karşılaştırılabilirliğinin incelenmesi kritik öneme sahiptir. Çünkü anlamlı karşılaştırmaların yapılabilmesi için ölçme değişmezliği temel bir ön koşuldur. Frekansçı (ML) yaklaşımda, küçük parametre farklılıkları bile modelin reddedilmesine yol açabilmektedir. Bu durumda Bayes yaklaşımı, parametre eşitliği kısıtlamalarından küçük sapmalara izin vererek karşılaştırılabilirliği korumada

daha esnek bir çözüm sunmaktadır. Özellikle skalar düzeyde ölçme değişmezliği, gruplar arası ortalama karşılaştırmalar için gereklidir. Ancak frekansçı yaklaşımda bu koşul çoğu zaman sağlanamamaktadır. Gruplar arasında tam sıfır farkı varsayımının gerçekçi olmaması model veri uyumunu olumsuz etkileyebileceğinden, küçük farklılıkların kabul edildiği Bayes yaklaşımı daha uygun bir yöntem olabilir. Çalışmamızın sonuçları da bu beklentiyi doğrulamış, yaklaşık sıfır farkı varsayımıyla Bayes yaklaşımı ile skalar düzeyde ölçme değişmezliğini sağladığını göstermiştir. Sonuç olarak, yaklaşık ölçme değişmezliği yaklaşımı, gruplar arası parametre eşitliği gerekliliği ile model uyumu arasında dengeli bir uzlaşma olarak değerlendirilebilir.

## 6. SONUÇLAR

Bu arařtırmada, TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan MÖSÖ'nün örtük yapısı, ÇGDFA ve BYÖD yöntemleri ile belirlenen ülkeler arasında ölçme deęişmezlięi incelenmiştir.

İlk aşamada, ilgili ülkelerin veri setleri üzerinde analiz öncesi varsayımlar deęerlendirilmiş ve betimsel istatistikler hesaplanmıştır. Analiz sonuçları, arařtırmaya dahil edilen tüm ülkelerde tek deęişkenli normallik koşulunun saęlandığı, çok deęişkenli normallięin ise hiçbir ülkede saęlanamamıştır. Bu nedenle, DFA ve ÇGDFA analizlerinde MLR kestirim yöntemi tercih edilmiş; Bayes analizleri ise Bayes kestirim yöntemi kullanılarak gerçekteştirilmiştir. DFA ölçme modeli aracılıęıyla faktör yükleri hesaplanmış ve bu yüklerin kabul edilebilir düzeyde olduęu belirlenmiştir. Elde edilen faktör yükleri kullanılarak her ülke verisine ait hem geçerlik hem de güvenilirlik düzeyleri deęerlendirilmiştir. Bu kapsamda, güvenilirliğe ilişkin kanıtlar CR ve  $\alpha$  deęerleriyle; yakınsak geçerliğe ilişkin kanıtlar ise AVE deęerleriyle ortaya konmuştur. Tüm deęerler belirlenen eşik düzeyin üzerinde olduęu görülmüş, dolayısıyla analiz kapsamındaki ülkelerde geçerlik ve güvenilirlik açısından yeterli istatistiksel kanıtların bulunduęu sonucuna ulaşılmıştır.

Arařtırmada, ölçmeğin tek boyutlu ve dokuz maddeden oluşan ölçme modelinin geçerlięinin test etmek amacıyla DFA uygulanmıştır. DFA sonuçları, elde edilen modelin veri uyum iyilięi indekslerinin tüm ülkelerde kabul edilen eşik deęerler içerisinde yer aldığını göstermektedir. Bu bulgu, ölçme modelinin incelenen tüm ülkelerde doęrulandığını ortaya koymaktadır. Ölçmeğin faktör yapısının belirlenmesinin ardından, ölçme deęişmezlięi analizleri gerçekteştirilmiştir.

Bu arařtırmanın birinci arařtırma sorusu, TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan MÖSÖ'nün örtük yapısı, ÇGDFA yöntemine göre belirlenen ülkeler arasında ölçme deęişmezlięini incelenmesidir. Kesin deęişmezlik yöntemi olan ÇGDFA ile yapılan analizler sonucunda, yapısal ve metrik deęişmezlik saęlandığı; ancak skalar deęişmezlięin saęlanamadığı sonucuna ulaşılmıştır. Aynı ölçme modeline yönelik olarak Bayes kesin deęişmezlik yöntemiyle gerçekteştirilen analizlerde de aynı sonuçlar elde edilmiştir: yapısal ve metrik deęişmezlik saęlanmış, skalar deęişmezlik saęlanamamıştır.

Bu elde edilen sonuç ile ülkeler arasında gösterge sabitlerinin eşit olmadığı ve örtük değişkenin tüm ülkelerde aynı şekilde yorumlanmadığını göstermektedir. Dolayısıyla, ülkeler arası karşılaştırmalarda gözlenen değişken ortalamalarındaki farklılıkların, ölçeğin özelliklerinden kaynaklanabileceği göz önünde bulundurulmalıdır.

Araştırmanın ikinci araştırma sorusu ise TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan MÖSÖ'nün örtük yapısı, BYÖD yöntemine göre belirlenen ülkeler arasında ölçme değişmezliğini incelenmesidir. Yaklaşık değişmezlik yöntemi olan BYÖD yöntemi kullanılarak gruplar arasında faktör yükleri ve gösterge sabitleri, sıfıra yakın önsel varyans değerleri (0,005; 0,01; 0,05; 0,10) ile analiz edilmiştir. Analiz sonuçlarına göre, faktör yükleri açısından tanımlanan dört varyans düzeyinin tamamında metrik değişmezlik sağlanmış; en iyi model uyumu ise 0,05 varyans düzeyinde elde edilmiştir. Skalar değişmezlik açısından, madde sabitlerine tanımlanan 0,05 ve 0,10 varyans düzeylerinde modelin veriye iyi uyum sağladığı; daha düşük varyans değerlerinde ise modelin veriye uyum sağlanmadığı belirlenmiştir. Bu bağlamda, skalar değişmezlik için en iyi model uyumu 0,01 varyans düzeyinde elde edilmiştir. Ayrıca skalar değişmezlik düzeyinde yapılan değerlendirmelerde, hem faktör yüklerine hem de madde sabitlerine önsel varyanslar tanımlanarak analiz gerçekleştirilmiştir. Bu analiz sonucunda, daha önce yalnızca madde sabitleri için tanımlandığında modelin veriye uyum sağlamayan 0,01 varyans düzeyinin, faktör yüklerine tanınan esnekliğin skalar düzeyde yaklaşık ölçme değişmezliğin elde edilmesine olumlu katkı sunduğunu göstermektedir.

Metrik değişmezlik analizinde elde edilen sonuçlar, faktör yüklerinin farklı gruplarda aynı şekilde işlediğini ve ölçülen yapının gruplar arasında aynı anlamı taşıdığını göstermektedir. Skalar değişmezlik düzeyinde ise madde sabitleri açısından yalnızca daha yüksek önsel varyans değerlerinde modelin veriye uyum sağladığı belirlenmiştir. Bu durum, gruplar arasında ortalama karşılaştırmalar yapılabilmesi için gerekli olan koşulun yalnızca belirli varyans düzeylerinde sağlandığını göstermektedir. Daha düşük varyans değerlerinde uyum sağlanmaması, modelin katı kısıtlamalar altında esnekliğini kaybettiğini düşündürür. Sonuçlar, BYÖD yönteminin avantajını vurgulamaktadır. Kesin değişmezlik sağlanamadığında, faktör yükleri ve madde sabitlerine küçük varyanslar tanımlayarak yaklaşık değişmezlik elde

edilebilmektedir. Ölçme deęişmezlięi testlerinde esnek varyans tanımlamalarının kritik rol oynadıęı ve özellikle skalar düzeyde karşılaştırmaların güvenilirliğini artırdıęını göstermektedir. Bu bağlamda, araştırmacılara kesin deęişmezlik varsayımlarına baęlı kalmadan, daha gerçekçi ve uygulanabilir karşılaştırmalar yapma imkânı sunmaktadır.

## 7. ÖNERİLER

Araştırma süreci ve elde edilen sonuçlar değerlendirildiğinde, uygulayıcılara ve araştırmacılara yönelik öneriler yer almaktadır.

1. Bu çalışmada, TIMSS 2019 uygulamasına ait sekizinci sınıf öğrenci anket verileri kullanılmıştır. Bu çalışmada ele alınan problem, farklı eğitim kademeleri, yıllar ve sınav türlerine ait veriler üzerinde de yürütülebilir.

2. Bu çalışmada, TIMSS 2019'a katılım sağlayan Kuveyt, Portekiz, Suudi Arabistan, Singapur, Türkiye ve Tayvan'a ait matematik öğrenci anket verileri kullanılmıştır. Ülke seçimi, TIMSS 2019 matematik başarı sıralamaları dikkate alınarak yapılmıştır. Benzer araştırmalar farklı disiplinlerde (alanlarda) ve çeşitli ülke gruplarıyla tekrarlanabilir.

3. Bu çalışma, ölçme değişmezliği ülkeler arası (kültürlerarası) değişkenine göre incelenmiştir. Gelecek araştırmalarda cinsiyet, bölgesel farklılıklar, sosyoekonomik düzey ve dil gibi farklı değişkenler dikkate alınarak ölçme değişmezliği analizleri incelenebilir.

4. Araştırma kapsamında, öğrencilerin matematiği öğrenmeyi sevme boyutuna ilişkin ölçme değişmezliği değerlendirilmiştir. Benzer araştırmalar diğer boyutlar için de gerçekleştirilebilir.

5. Ölçme değişmezliği analizleri, ölçeklerden elde edilen toplam puanlar üzerinden de gerçekleştirilebilir.

6. Ölçme değişmezliği analizleri, ÇGDFA ve BYÖD yöntemleriyle yürütülmüştür. Benzer araştırmalar farklı analiz yaklaşımları (örtük sınıf analizi ve HM gibi) kullanarak tekrarlanabilir; ayrıca birden fazla yöntemle yapılan analizler arasında yöntem karşılaştırmaları yapılabilir.

7. İki'den fazla grubun yer aldığı araştırmalarda, ÇGDFA yöntemiyle genelde ölçme değişmezliği sağlanamamaktadır. Bu nedenle, ölçme değişmezliği çalışmalarında BYÖD ve HM gibi yaklaşık ölçme değişmezliği yöntemlerinin kullanımı önerilmektedir.

8. Araştırma sonucumuz, ÇGDFA yöntemi ile kullanılan modelin ölçme değişmezliğini sağlamadığını göstermektedir. Bu durum, bazı ölçek maddelerinin

belirli gruplara karşı yanlılık içerdiği şeklinde yorumlanabilir. Sonraki arařtırmalarda yanlılık gösteren maddelerin belirlenmesi amacıyla Deęiřen Madde Fonksiyonları (DMF) analizleri yapılabilir; ayrıca kısmi deęiřmezlik arařtırmaları ile devam edilebilir.

9. Arařtırmada örneklem büyüklüğünün fazla olması ve  $\chi^2$  istatistiğinin bu duruma duyarlılığı nedeniyle, ölçme deęiřmezlięi deęerlendirmesinde model karşılařtırma indeksi olarak tercih edilmemiřtir. Bu etkinin azaltılması için daha küçük örneklemlemlerle yeniden uygulanabilir.

10. YEM kapsamında, analiz varsayımlarının saęlanıp saęlanmaması durumlarının sonuçlar üzerindeki etkisi ayrıca incelenebilir. Bu tür farklılıkların model çıktıları üzerindeki yansımaları arařtırılabilir.

11. Ölçme deęiřmezlięi çalışmalarında, yalnızca bireysel düzeydeki toplam puanlar deęil; öęrencilerin sınıf ve okul gibi hiyerarřik yapılar içinde yer aldığı dikkate alınarak çok düzeyli analizler de gerçekteřtirilebilir. Bu yaklaşım, verinin yapısal özelliklerini daha doęru yansıtması bakımından önerilebilir.

## KAYNAKÇA

- Alamri, A. A. S. (2019). *Exploring the behavior of model fit criteria in the Bayesian approximate measurement invariance: A simulation study* (Doctoral dissertation, University of South Florida). Available from ProQuest Dissertations & Theses Global. <https://digitalcommons.usf.edu/etd/8327/>
- Alkış, N. (2016). Bayes yapısal eşitlik modellemesi: Kavramlar ve genel bakış. *Gazi İktisat ve İşletme Dergisi*, 2(3), 105–116. <https://doi.org/10.30855/gjeb.2016.2.3.005>
- Arbuckle, J. L. (2012). IBM SPSS Amos 21 user's guide. *Amos Development Corporation*. <http://www.mrc-bsu.cam.ac.uk/bugs>
- Arts, I., Fang, Q., Meitinger, K., & Van de Schoot, R. (2021). Approximate measurement invariance of willingness to sacrifice for the environment across 30 countries: the importance of prior distributions and their visualization. *Frontiers in Psychology*, 12, Article 624032. <https://doi.org/10.3389/fpsyg.2021.624032>
- Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using mplus. *Structural Equation Modeling*, 21(3), 329-341. <https://doi.org/10.1080/10705511.2014.915181>
- Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation* (Version 3). Muthén & Muthén. <https://www.statmodel.com/download/Bayes3.pdf>
- Asparouhov, T., Muthén, B., & Morin, A. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: comments on Stromeyer et al. *Journal of Management*, 41(6), 1561–1577. <https://doi:10.1177/0149206315591075>
- Atılgan, M. (2022). *TIMSS 2019 matematik başarısına ilişkin duyuşsal özelliklerin cinsiyete göre ölçme değişmezliğinin incelenmesi* [Yüksek lisans tezi, Ankara Üniversitesi]. <https://avesis.ankara.edu.tr/yonetilen-tez/dce954bd-58cf-4bc3-96ad-994820aff421>
- Bademci, V. (2019). Geçerlik: Nedir? Ne Değildir? *Eğitim ve Toplum Araştırmaları Dergisi*, 6(2), 373-385. <https://doi.org/10.29228/etad.646773>
- Bagheri, Z., Chamanpara, P., Jafari, P., Balhara, Y. P. S., Arya, S., Ransing, R., Nishi, D., & Stevanović, D. (2022). Cross-cultural measurement invariance of the Quality of Life Enjoyment and Satisfaction Questionnaire-Short form across ten countries: the application of Bayesian approximate measurement invariance. *BMC Psychology*, 10(1), Article 160. <https://doi.org/10.1186/s40359-022-00864-y>
- Başusta, N. B. ve Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(4), 80-90. <https://doi.org/10.21031/hunefd.94663>

- Baykul, Y., Gelbal, S. ve Kelecioğlu, H. (2003). *Anadolu Lisesi öğretmenleri için eğitimde ölçme ve değerlendirme*. Pegem Akademi Yayıncılık.
- Benedict, R. H., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test–Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist*, 12(1), 43-55. <https://doi.org/10.1076/clin.12.1.43.1726>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bentler, P., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in analysis of covariance structures. *Psychological Bulletin*, 88(3), 558–606. <https://doi.org/10.1037/0033-2909.88.3.588>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications. <https://doi.org/10.1521/978.1462544646>
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2), 230-258. <https://doi.org/10.1177/0049124192021002005>
- Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamada Eğitim Yönetimi*, 32(32), 470-483. <https://dergipark.org.tr/pub/kuvey/issue/10365/126871>
- Byrne, B.M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross Cultural Psychology*, 34(2), 155–175. <https://doi.org/10.1177/002202210225>
- Byrne, B. M. (1988). Measuring adolescent self-concept: Factorial validity and equivalency of the SDQ III across gender. *Multivariate Behavioral Research*, 23(3), 361-375. [https://doi.org/10.1207/s15327906mbr2303\\_4](https://doi.org/10.1207/s15327906mbr2303_4).
- Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press. <https://doi.org/10.4324/9780203774762>
- 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), 456-466. <https://doi.org/10.1037/0033-2909.105.3.456>
- 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. <https://doi.org/10.1080/10705510701301834>
- Chen, F. F., Sousa, K.H., & West, S. G. (2005). Teacher's corner: testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(3), 471-492. <https://doi.org/10.1080/10705510701301834>

- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233-255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)
- Cieciuch, J., Davidov, E., Algesheimer, R., & Schmidt, P. (2018). Testing for approximate measurement invariance of human values in the European Social Survey. *Sociological Methods & Research*, 47(4), 665-686. <https://doi.org/10.1177/0049124117701480>
- Cieciuch, J., Davidov, E., Vecchione, M., Beierlein, C., & Schwartz, S.H. (2014). The cross-national invariance properties of a new scale to measure 19 basic human values: A test across eight countries. *Journal of Cross-Cultural Psychology*, 45 (5), 764–779. <https://doi.org/10.1177/0022022113520070>
- Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education*. Routledge.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.
- Çakici Eser, D. (2021). Investigation of measurement invariance according to home resources: TIMSS 2015 mathematical affective characteristics questionnaire. *International Journal of Assessment Tools in Education*, 633–648. <https://doi.org/10.21449/ijate.817168>
- Çelebi, V. (2019). Bayes teoremi bağlamında olasılıkçı Bayes epistemolojisinin kapsamı üzerine bir inceleme. *FLSF Felsefe ve Sosyal Bilimler Dergisi*, (28), 319-343. <https://doi.org/10.5281/zenodo.622357>
- Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2021). *Sosyal Bilimler İçin Çok Değişkenli İstatistik: SPSS ve LISREL Uygulamaları* (6. baskı). Pegem Akademi.
- Davidov, E., Cieciuch, J., Meuleman, B., Schmidt, P., Algesheimer, R., & Hausherr, M. (2015). The comparability of measurements of attitudes toward immigration in the European Social Survey: Exact versus approximate measurement equivalence. *Public Opinion Quarterly*, 79(S1), 244-266. <https://doi.org/10.1093/poq/nfv008>
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, 40(1), 55–75. <https://doi.org/10.1146/annurev-soc-071913-043137>
- De Bondt, N., & Van Petegem, P. (2015). Psychometric evaluation of the Overexcitability Questionnaire-Two applying Bayesian structural equation modeling (BSEM) and multiple-group BSEM-based alignment with approximate measurement invariance. *Frontiers in Psychology*, 6, 1963. <https://doi.org/10.3389/fpsyg.2015.01963>
- DeShon, R. P. (2004). Measures are not invariant across groups without error variance homogeneity. *Psychology Science* 46(4), 451-468. <https://psycnet.apa.org/record/2004-19990-008>

- Dokur, Y. (2023). *TIMSS 2019 sekizinci sınıf matematiğe yönelik güven ölçeğinin hizalama metoduna göre ölçme değişmezliğinin incelenmesi* [Yüksek lisans tezi, Gazi Üniversitesi]
- Dunson, D. B., Palomo, J., & Bollen, K. (2005). *Bayesian structural equation modeling* (Technical Report). Statistical and Applied Mathematical Sciences Institute.
- Ertürk, Z. ve Erdinç-Akan, O. (2018). TIMSS 2015 matematik başarısı ile ilgili bazı değişkenlerin cinsiyete göre ölçme değişmezliğinin incelenmesi. *Journal of Theoretical Educational Science*, 11(1), 204-226.  
<https://dergipark.org.tr/en/pub/akukeg/issue/40520/412604>
- Evran, D. (2012). *Assessing measurement invariance: Multiple group confirmatory factor analysis for differential item functioning detection in polytomous measures of Turkish and American students* [Master's thesis, University of Florida]  
[https://ufdcimages.uflib.ufl.edu/UF/E0/04/47/22/00001/EVRAN\\_D.pdf](https://ufdcimages.uflib.ufl.edu/UF/E0/04/47/22/00001/EVRAN_D.pdf)
- Finch, W. H., & French, B. F. (2015). *Latent variable modeling with R*. Routledge.  
<https://doi.org/10.4324/9781315881812>
- Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). McGraw-Hill.
- Gana, K., & Broc, G. (2019). *Structural equation modeling with lavaan*. Wiley-ISTE.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Chapman & Hall/CRC.
- George, D., & Mallery, M. (2010). *SPSS for Windows step by step: a simple guide and reference, 17.0 update (10a ed.)*. Pearson
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, 44(11 S3), S78-S94.  
<https://doi.org/10.1097/01.mlr.0000245454.12228.8f>
- Gucciardi, D. F., Peeling, P., Ducker, K. J., & Dawson, B. (2016). When the going gets tough: Mental toughness and its relationship with behavioural perseverance. *Journal of Science and Medicine in Sport*, 19(1), 81-86.  
<https://doi.org/10.1016/j.jsams.2014.12.005>
- Gülleroğlu, H. D. (2017). PISA 2012 matematik uygulamasına katılan Türk öğrencilerin duyuşsal özelliklerinin cinsiyete göre ölçme değişmezliğinin incelenmesi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 37(1), 373-396.  
<https://dergipark.org.tr/tr/pub/gefad/issue/31116/337717>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Exploratory factor analysis. Multivariate data analysis*, (7th ed., pp. 100). Pearson Education.

- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97-109. <https://doi.org/10.1093/biomet/57.1.97>
- Helvacı, İ. (2024). *Sosyal ve duygusal beceriler arasındaki ilişkilerin incelenmesinde sıklıkçı ve Bayesçi yapısal eşitlik modellerinin karşılaştırılması* [Yüksek lisans tezi, Harran Üniversitesi]. DSpace. <http://hdl.handle.net/11513/4034>
- Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Quality of Life Research*, 7(4), 323-335. <https://doi.org/10.1023/A:1024985930536>
- Hirschfeld, G., & Von Brachel, R. (2014). Improving Multiple-group confirmatory factor analysis in R—A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research, and Evaluation*, 19(1), Article 7. <https://doi.org/10.7275/qazy-2946>
- Hoijtink, H., & Van de Schoot, R. (2018). Testing small variance priors using prior-posterior predictive P-values. *Psychological Methods* 23(4), 561–569. <https://doi.org/10.1037/met0000131>
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3-4), 117-144. <https://doi:10.1080/03610739208253916>
- Hu, L. T., & Bentler, P. M. (1995). *Evaluating model fit*. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 76-99). Sage.
- 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), 1–55. <https://doi.org/10.1080/10705519909540118>
- IBM Corp. (2019). *IBM SPSS Statistics for Windows* (Version 26.0). IBM Corp.
- Jöreskog, K.G. (1971) Simultaneous factor analysis in several populations. *Psychometrika* 36(4), 409–426. <https://doi.org/10.1007/BF02291366>
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Scientific Software International. <https://psycnet.apa.org/record/1993-97878-000>
- Kankaras, M., & Moors, G. (2010). Researching measurement equivalence in cross-cultural studies. *Psihologija*, 43(2), 121-136. <https://doi.org/10.2298/PSI1002121K>
- Kaplan, D., & Depaoli, S. (2013). *Bayesian statistical methods*. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods: Vol. 1. Foundations* (pp. 407–437). Oxford University Press.

- Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling*, 18(2), 212-228. <https://doi.org/10.1080/10705511.2011.557337>
- Kim, E.S., Cao, C., Wang, Y., & Nguyen, D.T. 2017. Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling*, 24 (4), 524–544. <https://doi.org/10.1080/10705511.2017.1328074>
- Kim, J. S., & Bolt, D. M. (2007). Estimating item response theory models using Markov Chain Monte Carlo methods. *Educational Measurement: Issues and Practice*, 26(4), 38-51. <https://doi.org/10.1111/j.1745-3992.2007.00107.x>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling*. (4th ed.). The Guilford Press.
- Kline, R. B. (2011). *Convergence of structural equation modeling and multilevel modeling*. In M. Williams and W. P. Vogt (Eds.), *The SAGE handbook of innovation in social research methods* (pp. 562–589). SAGE. <https://doi.org/10.4135/9781446268261>
- Koçak, F. (2024). *TIMSS 2019 Matematik başarı testi ve matematiğe ilişkin duyuşsal özelliklerin ölçme deęişmezlięinin incelenmesi* [Yüksek lisans tezi, Hacettepe Üniversitesi].
- Korkmaz, S., Goksuluk, D. & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2), 151-162. <https://journal.r-project.org/archive/2014/RJ-2014-031/index.html>
- Koziol, N. A. (2010). *Evaluating measurement invariance with censored ordinal data: A Monte Carlo comparison of alternative model estimators and scales of measurement*. [Master's thesis, University of Nebraska]. <https://digitalcommons.unl.edu/cehsdiss/83/>
- Lambert, B. (2018). *A student's guide to Bayesian statistics*. SAGE. <https://uk.sagepub.com/en-gb/eur/book/student%E2%80%99s-guide-bayesian-statistics>
- Leitgöb H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., & Van de Schoot, R. (2022). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research*, 102, Article 102805. <https://doi.org/10.1016/j.ssresearch.2022.102805>
- Lek, K., & Van de Schoot, R. (2019). How the choice of distance measure influences the detection of prior-data conflict. *Entropy*, 21(5), Article 446. <https://doi.org/10.3390/e21050446>

- Liang, X. (2020). Prior sensitivity in Bayesian structural equation modeling for sparse factor loading structures. *Educational Psychological Measurement*, 80, 1025–1058. <https://doi.org/10.1177/0013164420906449>
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32 (1), 53-76. <http://doi.org/10.1207/s15327906mbr3201n3>
- Little, T. D. (2013). *Longitudinal structural equation modeling*. The Guilford Press.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3(3), 635-694. <https://doi.org/10.2466/pr0.3.7.635-694>
- Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Psychology Press.
- Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- MacCallum, R. C., Browne, M.W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>
- Magnusson, D. (1967). *Test theory*. Addison–Wesley.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530. <https://doi.org/10.1093/biomet/57.3.519>
- Martin, M. O., Von Davier, M., & Mullis, I. V. (2020). Methods and procedures: TIMSS 2019 technical report. *International Association for the Evaluation of Educational Achievement*. <https://timssandpirls.bc.edu/timss2019/methods/>
- Millî Eğitim Bakanlığı. (2007). *TIMMS ulusal raporu*. Millî Eğitim Bakanlığı Eğitim Araştırma ve Geliştirme Genel Müdürlüğü.
- Millî Eğitim Bakanlığı. (2015). *PISA 2012 ulusal nihai raporu*. Millî Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), 127-143. [https://doi.org/10.1016/0883-0355\(89\)90002-5](https://doi.org/10.1016/0883-0355(89)90002-5)
- Meredith, W. (1964) Notes on factorial invariance. *Psychometrika* 29(2), 177–185. <https://doi.org/10.1007/BF02289699>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543. <https://doi.org/10.1007/BF02294825>

- Meredith, W., & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, 57(2), 289-311. <https://doi.org/10.1007/BF02294590>
- Merkle, E. C., & Rosseel, Y. (2018). Blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, 85(4), 1–30. <https://doi.org/10.18637/jss.v085.i04>
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 58-65. <https://doi.org/10.1111/j.1745-3992.1995.tb00881.x>
- Meuleman, B. (2012). When are item intercept differences substantively relevant in measurement invariance testing? In H. Best & C. Wolf (Eds.), *Methods, theories, and empirical applications in the social sciences*, (pp. 97-106). VS Verlag Für Sozialwissenschaften. [http://dx.doi.org/10.1007/978-3-531-18898-0\\_13](http://dx.doi.org/10.1007/978-3-531-18898-0_13)
- Milfont, T.L., & Fischer, R. (2010) Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111-121. <https://doi.org/10.21500/20112084.857>
- Millsap, R. E., & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 380–392). Guilford Press.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge/Taylor & Francis Group.
- Mlodinow, L. (2009). *Ayyaş yürüyüşü*, (M. Kayı, Çev.). April Yayıncılık.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2019/frameworks/>
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/internationalresults/>
- Murat, N. (2012). *Yapısal eşitlik modellerde parametre tahminlerinde sıklıkçı ve Bayesçi bir yaklaşım* [Doktora Tezi, Ondokuz Mayıs Üniversitesi].
- Muthén, B., & Asparouhov, T. (2013). BSEM measurement invariance analysis. *Mplus Web Notes: No. 17*. <https://www.statmodel.com/examples/webnotes/webnote17.pdf>
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313–335. <https://doi.org/10.1037/a0026802>

- Muthén, L. K., & Muthén, B. O. (2017). *Mplus: Statistical analysis with latent variables; User's guide*. [version 8]. Muthén & Muthén.
- Oberski, D. (2014), Evaluating sensitivity of parameters of interest to measurement invariance in latent variable models. *Political Analysis*, 22(1), 45-60.  
<https://doi.org/10.1093/pan/mpt014>
- Ölçüoğlu, R. ve Çetin, S. (2016). TIMSS 2011 sekizinci sınıf öğrencilerinin matematik başarısını etkileyen değişkenlerin bölgelere göre incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 7(1), 202-220.
- Özçelik, D. A. (2010). *Ölçme ve değerlendirme*. Pegem Akademi.
- Pendergast, L. L., Von der Embse, N., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology*, 60(1), 65-82.  
<https://doi.org/10.1016/j.jsp.2016.09.003>
- Pokropek, A., Davidov, E., & Schmidt, P. (2019). A Monte Carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(5), 724–744. <https://doi.org/10.1080/10705511.2018.1561293>
- Pokropek, A., Schmidt, P., & Davidov, E. (2020). Choosing priors in Bayesian measurement invariance modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(5), 750–764.  
<https://doi.org/10.1080/10705511.2020.1749099>
- Polat, M. (2019). *TIMSS 2015 matematik ve fen duyuşsal özellik modellerinin kültürlere, cinsiyete ve bölgelere göre ölçme değişmezliğinin incelenmesi* [Yüksek lisans tezi, Hacettepe Üniversitesi]. <http://hdl.handle.net/11655/8068>
- 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(2), 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Raykov, T., Marcoulides, G. A., Lee, C. L., & Chang, C. (2013). Studying differential item functioning via latent variable modeling: A note on a multiple-testing procedure. *Educational and Psychological Measurement*, 73(5), 898-908.  
<https://doi.org/10.1177/0013164413478165>
- 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. <https://doi.org/10.1177/0013164413498257>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.  
<https://doi.org/10.1214/aos/1176344136>

- Seddig, D., & Leitgöb, H., 2018. Approximate measurement invariance and longitudinal confirmatory factor analysis: concept and application with panel data. *Survey Research Methods*, 12(1), 29–41. <https://doi.org/10.18148/srm/2018.v12i1.7210>
- Sırgancı, G. (2022). Geniş ölçekli değerlendirmelerde çoklu grup karşılaştırmalarının geçerliği üzerine yeni yaklaşımlar. *Eğitim ve Bilim*, 47(211), 1–20. <https://doi.org/10.15390/EB.2022.11045>
- Sideridis, G.D., Tsaousis, I., & Alamri, A.A., 2020. Accounting for differential item functioning using Bayesian approximate measurement invariance. *Educational and Psychological Measurement*, 80(4), 638–664. <https://doi.org/10.1177/0013164419887482>
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639. <https://doi.org/10.1111/1467-9868.00353>
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2014). *OpenBUGS user manual* (Version 3.2.3). MRC Biostatistics Unit. <http://www.mrc-bsu.cam.ac.uk/bugs>
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS user manual* (Version 1.4). MRC Biostatistics Unit. <http://www.mrc-bsu.cam.ac.uk/bugs>
- Steenkamp, J.E.M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research, *Journal of Consumer Research*, 25(1), 78–90. <https://doi.org/10.1086/209528>
- Steiger, J. H., & Lind J. M. (1980). *Statistically based tests for the number of factors*. Paper presented at the Annual Meeting of the Psychometric Society, Iowa City, IA.
- Steiger, J. H. (2016). Notes on the Steiger–Lind (1980) handout. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(6), 777–781. <https://doi.org/10.1080/10705511.2016.1227754>
- Steinmetz, H., Schmidt, P., Tina-Booh, A., Wiczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in the structure of job satisfaction. *European Journal of Psychological Assessment*, 25(1), 1–11. <https://doi.org/10.1027/1015-5759.25.1.1>
- Stevanovic, D., Costanzo, F., Fucà, E., Valeri, G., Vicari, S., Robins, D. L., ... Knez, R. (2021). Measurement invariance of the Childhood Autism Rating Scale (CARS) across six countries. *Autism Research*, 14(12), 2544–2554. <https://doi.org/10.1002/aur.2586>
- Stigler, S. M. (1983). Who discovered Bayes's theorem? *The American Statistician*, 37(4), 290–296. <https://doi.org/10.1080/00031305.1983.10483110>

- Struening, E. L., & Cohen, J. (1963). Factorial invariance and other psychometric characteristics of five opinions about mental illness factors. *Educational and Psychological Measurement*, 23(2), 289-298. <https://doi.org/10.1177/001316446302300206>
- Şen, S. (2023). *Mplus ile yapısal eşitlik modellemesi uygulamaları* (2. basım). Nobel Akademik Yayıncılık.
- Tabachnick, B. G., & Fidell, L. S. 2001 . *Using Multivariate Statistics* (4th ed.). Allyn & Bacon.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA* (Vol. 724). Thomson/Brooks/Cole.
- Terzi, R. (2021). Test geliştirme. S. Şen ve İ. Yıldırım (Ed.ler), *Bilimsel araştırma* (ss. 130-145). Nobel Yayıncılık.
- Turgut, M. F. ve Baykul, Y. (1992). *Ölçekleme teknikleri* (2. basım). ÖSYM Yayınları.
- Turgut, M. F. ve Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme* (4. basım). Pegem Akademi.
- Turgut, M.F. (1987). *Eğitimde ölçme ve değerlendirme metotları* (5. basım). Saydam Matbacılık
- Usta, M.E., Demirtaş, Z. ve Demir, M.Ş. (2013). Yurtdışına gitmiş eğitimcilerin küreselleşmeye ilişkin görüşleri. *Yeni Yüzyıl Üniversitesi Eğitim Fakültesi Dergisi*, 10(1), 227-247. <https://dergipark.org.tr/tr/download/article-file/146239>
- Uyar, Ş. (2021). Factor structure and measurement invariance of the TIMSS 2015 mathematics attitude questionnaire: Exploratory structural equation modelling approach. *International Journal of Assessment Tools in Education*, 8(4), 855–871. <https://doi.org/10.21449/ijate.796862>
- Van de Schoot, R., Lugtig, P., & Hox, J. (2012) A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492. <https://doi.org/10.1080/17405629.2012.686740>
- Van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychologist*, 16(2), 75-84. <https://ehps.net/ehp/index.php/contents/article/view/ehp.v16.i2.p75>
- Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, 85(3), 842–860. <https://doi.org/10.1111/cdev.12169>

- Van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4*, 1–15. <https://doi.org/10.3389/fpsyg.2013.00770>
- Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*, Sage.
- Van de Vijver, F. J. (1998). Towards a theory of bias and equivalence. *ZUMA Nachrichten Spezial*, *3*, 41-65. [https://www.ssoar.info/ssoar/bitstream/handle/document/49731/ssoar-1998-vijver-Towards\\_a\\_theory\\_of\\_bias.pdf](https://www.ssoar.info/ssoar/bitstream/handle/document/49731/ssoar-1998-vijver-Towards_a_theory_of_bias.pdf)
- Van de Vijver, F. J., & Leung, K. (2021). *Methods and data analysis for cross-cultural research* (Vol. 116). Cambridge University Press. <https://doi.org/10.1017/9781107415188>
- 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. <https://doi.org/10.1177/109442810031002>
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>
- Wang, S. (2017). *Bayesian approximate measurement invariance approach* (Unpublished doctoral dissertation). University of Cincinnati.
- Wheaton, B., Muthen, B., Alwin, D. F., & Summers, G. (1977). Assessing reliability and stability in panel models. *Sociological Methodology*, *8*(1), 84-136.
- Whitaker, B. G., & McKinney, J. L. (2007). Assessing the measurement invariance of latent job satisfaction ratings across survey administration modes for respondent subgroups: A MIMIC modeling approach. *Behavior Research Methods*, *39*(3), 502-509. <https://doi.org/10.3758/BF03193029>
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). American Psychological Association.
- Winter, S. D., & Depaoli, S. (2020). An illustration of Bayesian approximate measurement invariance with longitudinal data and a small sample size. *International Journal of Behavioral Development*, *44*(4), 371-382. <https://doi.org/10.1177/0165025419881172>

- Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, 12(3), 1–26. <https://doi.org/10.7275/tsq9-7h78>
- Wurster, S. (2022). Measurement invariance of non-cognitive measures in TIMSS across countries and across time: An application and comparison of multigroup confirmatory factor analysis, Bayesian approximate measurement invariance and alignment optimization approach. *Studies in Educational Evaluation*, 73, Article 101143. <https://doi.org/10.1016/j.stueduc.2022.101143>
- Yiğiter, M. S. (2023). Matematik duyuşsal özellik faktörlerinin cinsiyete göre ölçme deęişmezlięinin incelenmesi: TIMSS 2019 Türkiye örneęi. *Anadolu Üniversitesi Eğitim Fakültesi Dergisi*, 7(4), 859–882. <https://doi.org/10.34056/aujef.1198134>
- Yiğiter, M. S. (2024). Matematik içsel motivasyonunun ülkeler arası ölçümü: On dört ülke arasında MG-CFA ve hizalama yöntemi ile ölçme deęişmezlięinin incelenmesi. *Kuramsal Eğitim Bilim Dergisi*, 17(1), 1–27. <https://doi.org/10.30831/akukeg.1207350>
- Yin, L., & Fishbein, B. (2020). Creating and interpreting the TIMSS 2019 context questionnaire scales. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (pp. 16.1-16.331). Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/methods/chapter-16.html>

## ÖZGEÇMİŞ

### KİŞİSEL BİLGİLER

**İsim Soyisim** : ABDULLAH YARADILMIŞ  
**Doğum Tarihi** : 1989-06-01  
**Doğum Yeri** : GÜROYMAK  
**Telefon** : 5422131308  
**Posta** : ayaradlms@gmail.com