



# HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

TIMSS – 2019 BİLGİSAYAR KULLANIMINA İLİŞKİN MADDELERİN ÖLÇME  
DEĞİŞMEZLİĞİ VE DEĞİŞEN MADDE FONKSİYONU AÇISINDAN İNCELENMESİ

Zeynep Neveser KIZILÇİM

Yüksek Lisans Tezi

Ankara, 2024

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

*Daha ileriye... En iyiye...*



Eğitim Bilimleri Ana Bilim Dalı  
Eğitimde Ölçme ve Değerlendirme Programı

TIMSS – 2019 BİLGİSAYAR KULLANIMINA İLİŞKİN MADDELERİN ÖLÇME  
DEĞİŞMEZLİĞİ VE DEĞİŞEN MADDE FONKSİYONU AÇISINDAN İNCELENMESİ

EXAMINATION OF TIMSS- 2019 COMPUTER USE ITEMS IN TERMS OF MEASUREMENT  
INVARIANCE AND DIFFERENTIAL ITEM FUNCTION

Zeynep Neveser KIZILÇİM

Yüksek Lisans Tezi

Ankara, 2024

## Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Zeynep Neveser KIZILÇİM'in hazırladıđı "TIMSS- 2019 Bilgisayar Kullanımına İlişkin Maddelerin Ölçme Deđişmezliđi ve Deđişen Madde Fonksiyonu Açısından İncelenmesi" başlıklı bu çalıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölçme ve Deđerlendirme Ana Bilim Dalında Yüksek Lisans** olarak kabul edilmiřtir.

J¼ri Başkanı	Doç. Dr. K¼bra ATALAY KABASAKAL	İmza
J¼ri Üyesi (Danıřman)	Doç. Dr. Sevda ÇETİN	İmza
J¼ri Üyesi	Dr. Öğr. Üyesi Meral ALKAN	İmza

Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, Öğretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 07 / 06/ 2024 tarihinde uygun gör¼lm¼ř ve Enstit¼ Yönetim Kurulunca tarafından .... / ... / 2024 tarihi itibarıyla kabul edilmiřtir.

Prof. Dr. İsmail Hakkı MİRİCİ  
Eđitim Bilimleri Enstitüsü M¼d¼r¼

## Öz

Bu arařtırmada TIMSS 2019'da bilgisayar kullanımına iliřkin maddelerin dil ve kltre gre lme deęiřmezlięi ve deęiřen madde fonksiyonu yntemleri ile yanlılıęının incelenmesi amalanmıřtır. Arařtırmanın evrenini TIMSS 2019 dngsne 8. sınıf dzeyinde katılan tm lkeler; arařtırmanın rneklemini ise Trkiye, İngiltere ve Katar oluřturmaktadır. İki boyutlu ve 11 madde ile kurulan modelin lme deęiřmezlięi GDFA ile, Deęiřen Madde Fonksiyonu (DMF) analizleri OLR ve Poly-SIBTEST yntemleri ile incelenmiřtir. Analizlerin gerekleřtirilmesinde R Studio, SIBTEST, SPSS yazılımları kullanılmıřtır. Analizler ncesinde gvenirlik iin Cronbach'ın alpha katsayısı hesaplanmıř, geerlik iin aımlayıcı ve doęrulamayı faktr analizleri yapılmıřtır. Yapılan karřılařtırmalarda farklı dil aynı kltrde (ARP-İNG) yapısal ve metrik deęiřmezlięin saęlandığı, farklı dil farklı kltre gre (TR-İNG) yapısal deęiřmezlięin saęlanmış olduęu sonucuna ulařılmıřtır. Yanlılık analizleri incelendięinde farklı dil aynı kltrde (ARP-İNG) OLR yntemine gre drt maddede A dzeyinde tek biimli DMF olduęu, farklı dil aynı kltr (ARP-İNG) karřılařtırmasında Poly-SIBTEST yntemi sonularına gre yedi maddenin C, drt maddenin B dzeyinde ve drt maddenin İngilizce dili lehine, yedi maddenin ise Arapa lehine DMF olduęu grlmřtr. Farklı dil farklı kltr (TR-İNG) karřılařtırmalarında OLR yntemiyle yapılan analizlerde beř madde A dzeyinde ve tek biimli,  madde B dzeyinde tek biimli olmayan, iki madde ise C dzeyinde tek biimli olmayan DMF gstermiřtir. Farklı dil farklı kltr (TR-İNG) karřılařtırmalarında poly -SIBTEST yntemi sonularına bakıldıęında drt maddenin A dzeyinde, 7 maddenin C dzeyinde DMF gsterdięi bulunmuřtur. Bu maddelerden yedisi Trke lehine, drt madde ise İngilizce lehine DMF'li madde vardır. Arařtırmada farklı DMF yntemleri kullanılmıř ancak farklı yntem sonularının tam uyum gstermedięi grlmřtr.

**Anahtar szckler:** TIMSS 2019,eTIMSS, lme deęiřmezlięi, yanlılık, DMF, Poly-sibtest, OLR, oklu grup doęrulamayı faktr analizi

### Abstract

This study aimed to investigate the measurement invariance and bias of computer use-related items in TIMSS 2019 across different languages and cultures using the changing item function method. The analysis included data from Turkiye, England, and Qatar, and utilized MG-CFA and OLR with Poly-SIBTEST methods for measurement invariance testing. The study used R studio, SIBTEST, and SPSS for analysis and calculated Cronbach's alpha coefficient for reliability, as well as conducting exploratory and confirmatory factor analyses for validity. The results showed that structural invariance was achieved for items in the same culture (AR-ENG), and structural and metric invariance was achieved for items in different languages and cultures (TUR-ENG). Upon analysing the bias, it was found that the OLR method revealed four uniform differences in the same culture and language, while seven items were rated at the C level. Four items were rated at the B level, and four items showed a preference for English according to the Poly-SIBTEST method in comparing different languages in the same culture. In the OLR method analysis of different languages and cultures, five items exhibited uniform differences at the A level, while three items displayed non-uniform differences at the B level, and two items showed non-uniform differences at the C level. The Poly-SIBTEST method results for comparing different languages and cultures revealed that four items showed differences at level A, and seven items showed differences at level C. Of these, seven items showed a preference for Turkish, and four items showed a preference for English.

**Keywords:** TIMSS 2019, eTIMSS, measurement invariance, bias, DIF, Poly -sibtest, OLR, multi group confirmatory factor analysis

**6 Şubat 2023 depreminde hayatını kaybeden canım arkadaşım Elisa  
NALBANTOĞLU, ailesi ve diğer tüm hayatını kaybedenlere ...**

## Teşekkür

Tez hazırlama sürecinde bilgisini, desteğini ve anlayışını esirgemeyen, değerli danışmanım Doç. Dr. Sevda ÇETİN' e, tez jürimde yer alan dönütleriyle tezime katkıda bulunan değerli jüri üyelerim Doç. Dr. Kübra ATALAY KABASAKAL 'a ve Dr. Öğr. Üyesi Meral ALKAN' a çok teşekkür ederim. Aynı zamanda tez öneri sürecimde değerli görüş ve önerileriyle katkı veren Prof. Dr. Burcu ATAR'a teşekkürlerimi sunarım.

Ölçme ve Değerlendirmeyi bana sevdiren, lisans eğitimim boyunca gerek ders anlatımı gerek davranışlarıyla örnek aldığım Doç. Dr. Duygu KOÇAK' a her zaman yanımda olduğu için çok teşekkür ederim. Akademi sürecimde ne zaman desteğe ihtiyaç duysam her zaman yanımda olan Öğr. Gör. Dr. Aysel EYERCİ'ye çok teşekkür ederim.

Akademik her sürecimde en önemli motive kaynağım, başarımla her zaman gurur duyduklarını bildiğim babam Yavuz KIZILÇİM, ablam Nergis Gizem KIZILÇİM ve annem Sevgi KIZILÇİM' e çok teşekkür ederim.

Kırşehir'e alışmamı sağlayan ve akademik anlamda her zaman desteğini hissettiğim Anabilim Dalı Başkanım Doç. Dr. Ahmet Salih ŞİMŞEK' e çok teşekkür ederim. Ne zaman olumsuzluğa kapılsam beni motive eden, sabırla beni dinleyen ve her zaman destekleyen Dr. Öğr. Üyesi Ersoy KARABAY'a çok teşekkür ederim. Beni destekleyen Doç. Dr. Yasemin KUZU' ya teşekkür ederim. Tezimin son okumalarını yapan Arş. Gör. Rüveyda Nur GARİP'e çok teşekkür ederim. Fikir ve deneyimlerine ne zaman başvurursam yardımcı olmaya çalışan araştırma görevlisi arkadaşlarıma teşekkür ederim.



## İçindekiler

Kabul ve Onay.....	ii
Öz.....	iii
Teşekkür.....	vi
Tablolar Dizini.....	ix
Şekiller Dizini.....	x
Simgeler ve Kısaltmalar Dizini.....	xi
Bölüm 1 Giriş.....	1
Problem Durumu .....	1
Araştırmanın Amacı ve Önemi .....	6
Araştırma Problemi .....	8
Sayıltılar .....	9
Sınırlılıklar .....	9
Tanımlar.....	9
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	10
Ölçme Değişmezliği .....	10
Değişen Madde Fonksiyonu (DMF) ve Yanlılık .....	22
Uluslararası Sınavlar ve TIMSS .....	31
İlgili Araştırmalar .....	33
Ölçme Değişmezliği ile İlgili Yapılan Ulusal Araştırmalar .....	33
Ölçme Değişmezliği ile İlgili Yapılan Uluslararası Araştırmalar .....	34
Yanlılık ile İlgili Yapılan Ulusal Araştırmalar .....	36
Yanlılık ile İlgili Yapılan Uluslararası Araştırmalar .....	40
İlgili Araştırmaların Genel Özeti.....	41
Bölüm 3 Yöntem.....	43
Araştırmanın Türü .....	43
Araştırmanın Evreni ve Örneklemi .....	43

Veri Toplama Süreci.....	44
Veri Toplama Araçları .....	44
Bölüm 4 Bulgular, Yorumlar ve Tartışma.....	49
Bölüm 5 Sonuç ve Öneriler.....	57
Sonuçlar.....	57
Öneriler .....	59
Kaynaklar .....	61
EK-A: Araştırma Etik Komisyon İzin Muafiyeti Formu.....	72
EK-B: Etik Beyanı .....	73
EK-C: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu .....	74
EK-Ç: Thesis/Dissertation Originality Report.....	75
EK-D: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı.....	76

## Tablolar Dizini

<b>Tablo 1</b> Ölçme Değişmezliği Özet Tablosu .....	19
<b>Tablo 2</b> Uyum İyiliği İndeksleri .....	22
<b>Tablo 3</b> SIBTEST DMF Düzeylerinin Yorumlanması (Roussos & Stout (1996) ...	29
<b>Tablo 4</b> $\Delta R^2$ Değerlerine Göre DMF Düzeyi Ölçütleri .....	31
<b>Tablo 5</b> Araştırma Örnekleminin Frekans ve Yüzdeleri .....	43
<b>Tablo 6</b> Veri Toplama Araçları.....	45
<b>Tablo 7</b> Ükelere göre Cronbach Alfa Güvenirlik Katsayıları .....	46
<b>Tablo 8</b> Bilgisayar Kullanım Sıklığı Boyutu Madde Faktör Yükleri .....	47
<b>Tablo 9</b> Bilgisayar Kullanımı Öz Yeterlik Boyutu Madde Faktör Yükleri .....	47
<b>Tablo 10</b> TIMSS 2019 8. Sınıf Bilgisayar Kullanımına İlişkin Maddelerin DFA Uyum İndeks Değerleri .....	48
<b>Tablo 11</b> Aynı Kültür -Farklı Dile Göre Değişmezlik Aşamalarında Uyum Katsayıları .....	49
<b>Tablo 12</b> Farklı Dil- Farklı kültüre Göre Değişmezlik Aşamalarında Uyum Katsayıları .....	51
<b>Tablo 13</b> Farklı Dil- Aynı Kültürde OLR Analiz Sonuçları .....	52
<b>Tablo 14</b> Farklı Dil- Aynı Kültüre Göre Poly-SIBTEST Analiz Sonuçları .....	53
<b>Tablo 15</b> Farklı Dil- Farklı Kültüre Göre OLR Analiz Sonuçları .....	54
<b>Tablo 16</b> Farklı Dile Farklı Kültüre Göre Poly-SIBTEST Analiz Sonuçları.....	55

**Şekiller Dizini**

<b>Şekil 1</b> <i>Yapısal Değişmezlik Modeli Örneği</i> .....	15
<b>Şekil 2</b> <i>Metrik Değişmezlik Modeli Örneği</i> .....	16
<b>Şekil 3</b> <i>Ölçek Değişmezliği Modeli Örneği</i> .....	17
<b>Şekil 4</b> <i>Katı Değişmezliği Modeli Örneği</i> .....	18
<b>Şekil 5</b> <i>Tek Biçimli DMF (Swaminathan&amp;Rogers, 1990)</i> .....	26
<b>Şekil 6</b> <i>Tek biçimli olmayan DMF (Swaminathan&amp;Rogers, 1990)</i> .....	26

## Simgeler ve Kısaltmalar Dizini

**CFI:** Comparative Fit Index

**ÇGDFA:** Çoklu Grup Doğrulayıcı Faktör Analizi

**DFA:** Doğrulayıcı Faktör Analizi

**DMF:** Değişen Madde Fonksiyonu

**GFI:** Root Mean Square Error of Approximation

**KTK:** Klasik Test Kuramı

**MEB:** Millî Eğitim Bakanlığı

**MTK:** Madde Tepki Kuramı

**OECD:** Ekonomik İş birliği ve Kalkınma Teşkilatı (Organization of Economic Cooperation and Development)

**OLR:** Ordinal Lojistik Regresyon

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

**RMSEA:** The Root Mean Square Error Of Approximation

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

**TLI:** Tucker-Lewis İndeksi

## Bölüm 1

### Giriş

Bu bölümde; problem durumu, araştırmanın amacı ve önemi, araştırma problemi, sayıtlar, sınırlılıklar ve tanımlara yer verilmiştir.

#### Problem Durumu

Teknolojik gelişmelerin hız kazanmasıyla birlikte eğitim sistemlerinde de değişimlerin ve yeniliklerin meydana geldiği görülmektedir. Ülkeler bu gelişmeleri takip ederek ve farklı ülkelerdeki yeniliklerden haberdar olarak kendilerini güncellemektedir. Ülkeler eğitim sistemlerinde iyileştirmeler yaparak farklı disiplinlerde ilerlemeler kaydetmeyi hedeflemektedir. Gerek ekonomik planlamalar gerekse sağlık, hukuk gibi pek çok farklı disiplin eğitim sistemlerindeki gelişmelerden etkilenmektedir. Uluslararası alandaki gelişmelerin takip edilmesi Ekonomik İş birliği ve Kalkınma Örgütü (Organisation for Economic Co-operation and Development-OECD), Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (IEA) gibi kuruluşların yapmış oldukları geniş ölçekli sınavlar ile mümkün olabilmektedir. Yapılan sınavlar ile öğrencilerin başarı durumlarının karşılaştırılmasının yanı sıra öğrenci, veli ve okul idarecileri anketleri aracılığı ile eğitimin paydaşı olan pek çok bireyden dönüt alınmaktadır.

Belirli döngülerle ülkelerin eğitim sistemleri karşılaştırılmaktadır. Bu sınavlardan biri OECD'nin programlarından biri olan Uluslararası Öğrenci Değerlendirme sınavıdır (Programme for International Student Assessment-PISA). PISA, 15 yaşındaki bireylerin okuma becerileri, matematik okuryazarlığı ve fen okuryazarlığındaki bilgi ve becerilerini ölçmeyi, özellikle de bireylerin günlük yaşamda karşılaşmış oldukları zorlukları çözme becerilerini ölçmeyi amaçlar (OECD, 2024). İlk defa 2000 yılında yapılan PISA döngüleri, üç yılda bir tekrar edilmektedir. Araştırmanın her yeni döngüsünde bir alan ağırlıklı olarak

seçilmektedir. Türkiye, PISA araştırmasına 2003 yılından beri katılmaktadır. PISA arařtırmalarına katılan lke sayısı her dngde artış gstermektedir.

Uluslararası Matematik ve Fen Eęilimleri Arařtırması (Trends in International Mathematics and Science Study -TIMSS) ise PISA'dan farklı olarak IEA tarafından yapılan bir arařtırmadır. TIMSS, ęrencilerin uluslararası alanda matematik ve fen bilimleri bařarisını lmek iin yapılmaktadır. TIMSS dngsne drdnc ve sekizinci sınıf dzeyinde olan ęrenciler katılmaktadır. TIMSS dngs, drder yıllık periyotlarla yapılmaktadır. TIMSS arařtırmasının drt yıllık periyotlarla gerekleřtirilmesi bu arařtırmanın boylamsal aıdan da bireylerin deęerlendirilmesine olanak saęlamaktadır. TIMSS uygulamasının ilki 1995 yılında gerekleřtirilmiřtir. Yaklařık 28 yıllık gemiře sahip olan bu arařtırmaya Trkiye ilk olarak 1999 yılında sekizinci sınıf dzeyi ęrenciler ile katılmıřtır (MEB, 2020). TIMSS arařtırması, 2019 yılında bilgisayar tabanlı deęerlendirmeye geiř yapmıřtır. Bilgisayar tabanlı uygulamaya TIMSS dngsne katılan lkelerin neredeyse yarısı geiř yaparken, dngye katılan dięer yarısındaki lkeler kęit-kalem tabanlı uygulamalara devam etmiřtir. TIMSS 2019 uygulamasına drdnc sınıf dzeyinde 58 lke ve 6 eyalet katılmıř; sekizinci sınıf dzeyinde 39 lke, 7 eyalet katılım gstermiřtir (Mullis vd., 2020). Uluslararası alanda yapılan bu arařtırmalardan anlařılacaęı zere belirli periyotlar ile lkelerin eęitim sistemleri karřılařtırılmakta, bu karřılařtırmalarla lkeler kendilerini dięer lkelerle kıyaslayabilme imknı bulabilmektedir.

IEA tarafından uluslararası alanda dzenlenen bir bařka arařtırma Uluslararası Okuma Becerilerinde Geliřim arařtırmasıdır (Progress in International Reading Literacy Study-PIRLS). PIRLS beřer yıllık dnglerle gerekleřtirilmektedir. Bu arařtırma ilkokul drdnc sınıf dzeyindeki bireylerin farklı metin trlerini anlama becerilerini lme, okuma alışkanlıklarını aıęa ıkarmayı amalar. Trkiye PIRLS alıřmalarına 2001 yılında ve 2021 yılında katılmıřtır (MEB, 2021).

Yenilikler ve geliřmeler iřıęında uluslararası sınav dnglerinde ęrencilerin teknoloji kullanımı ile ilgili maddelerin de yer aldıęı grlmektedir. Bylece ęrencilerin

bilgisayar kullanımına ilişkin temel yeterliliklere sahip olup olmadığı ölçülmektedir. PISA 'da 2000 yılından itibaren döngülerde, öğrencilerin farklı ortamlarda bilgisayar kullanımı, bu kaynaklara erişilmesi gibi durumların anlaşılması amacıyla bilgi ve iletişim teknolojilerine aşinalık anketine (ICT- Familiarity Questionnaire) yer verildiği görülmektedir (OECD,2024). TIMSS döngüsünde 2019 yılında eTIMSS uygulamasına geçilmiş, döngüye katılan ülkelerin de neredeyse yarısı bu döngüye bilgisayar tabanlı olarak katılmıştır. Öğrenci anketlerinde de öğrencilerin bilgisayar kullanımına aşinalığı, tablet veya bilgisayarı ne sıklıkla kullandığı, bazı işlevleri kullanabilmedeki öz yeterliği ölçülmeye çalışılmıştır. Bilgisayar tabanlı eTIMSS uygulamalarının gerçekleştirilmesinde teknik aksaklıkların olmaması için mümkün olduğunca çeşitli dijital araçla uyumlu sistem geliştirilmiştir (Fisbein vd., 2018).

Uluslararası alanda yapılan eğitim araştırmalarında uygulanan sınavların aynı yapıyı ölçmesi, yapılan karşılaştırmaların ve değerlendirmelerin daha anlamlı olmasını sağlayacaktır. Çeviriden veya kültürden kaynaklı farklılıklar öğrencilerin maddeleri farklı şekilde anlayabilmesine neden olmaktadır. Öğrencilerin maddeleri aynı şekilde anlaması, uluslararası alanda uygulanan sınavların güvenilirliği ve geçerliğine ilişkin kanıtlar toplanmasını gerekli kılmaktadır. Güvenirlik ve geçerliğe ilişkin kanıtlar toplamak ölçme ve değerlendirmenin etkili bir şekilde yapılmasını sağlar. Etkili değerlendirmeler ile bireyler hakkında daha doğru kararlar alınabilmektedir (Güler 2019; Turgut, 1977).

Ölçme ve değerlendirme yaparken bireylerin, sınavdan aldıkları puan farklılıklarının ölçme aracından kaynaklı olmadığından emin olunmalıdır. Başka bir ifadeyle bireylerin farklı puanlar almasının, bireysel özelliklerden kaynaklanması beklenmektedir. Bireylerin ölçülmek istenen yapıya ilişkin özellikleri maddeler aracılığıyla ölçülür (Baykul, 1999). Geçerlik, Eğitimde ve Psikolojide Test Standartları'nda (2014) puanların yorumlanmasında, test yönergelerinin öneriler ölçüsünde uygulanmasında ve bu uygulamaların değerlendirilmesi ile ilgili olarak tanımlanmıştır. Ölçme aracını geliştiren uzmanlar, testin yapısını önceden belirlemelidir. Değerlendirmelerin geçerli şekilde kullanılması kavramsal çerçevenin netleştirilmesiyle de ilgilidir. Geçerlik tekil bir kavram olup, testin geçerliği



kanıtlarla güçlenir. Yapı geçerliği, testin geçerliğinin doğrulamak için kullanılabilir. Yapı geçerliğinin belirlenmesinde pek çok yöntem başvurulmaktadır. Bunlardan en yaygın olanlardan biri faktör analizidir. Faktör analizi yönteminde birbirleri ile ilişkili pek çok değişken, anlamlı ve daha az sayıda değişken ile açıklanmaya çalışılır (Büyüköztürk, 2002.)

Yapıya ilişkin kanıt toplama sürecinde ölçme aracının değişmezliği de göz ardı edilmemesi gereken konulardan biridir. Birbirinden bağımsız farklı gruplarda karşılaştırmalar yapabilmek için ölçme değişmezliği sağlanmalıdır. Ölçme değişmezliği, aynı ölçme aracının evrenin birbirinden bağımsız örneklemelerinde aynı yapıya dair kanıtlar sunmasıdır (Şekercioğlu, 2018). Kısacası ölçme değişmezliği oluşturulan bir ölçme modelinin farklı gruplarda aynı faktör yapılarına sahip olmasıdır. Ölçme değişmezliğinin sağlandığı durumlarda yapı geçerliğine ilişkin kanıtlar da daha güçlü olacaktır. Ölçme değişmezliğinin test edilmesinde Çoklu Grup Doğrulayıcı Faktör Analizi (ÇGDFA) yöntemi tercih edilmiş olup, bu yöntem aracılığıyla farklı alt gruplar arasındaki ölçme değişmezliğinin aynı anda incelenebildiği görülmüştür (Brown, 2006).

Ölçme değişmezliğinin test edilmesi aşamaları farklı şekillerde sınıflandırılmıştır. Bu araştırma kapsamında Meredith'in (1993) önermiş olduğu aşamalara göre ölçme değişmezliği test edilmiştir. Değişmezlik aşamaları şu şekilde sıralanabilir; yapısal değişmezlik, metrik değişmezlik, ölçek değişmezliği ve katı değişmezlik. Ölçme değişmezliği aşamalarından ilki olan yapısal değişmezlikte yapılmış olan ölçümlerin kavramsal olarak farklı alt gruplarda aynı yapıları ölçmesidir. Eğer yapılan ölçümler bağımsız olan farklı gruplarda aynı anlama gelmiyorsa yapısal değişmezlikten söz edilememektedir. Yapısal değişmezlik, diğer ölçme değişmezliği aşamaları için ön koşul olduğundan sağlanmadığı zaman diğer aşamaların da test edilmesi anlamlı olmayacaktır (Vandenberg & Lance, 2000). Metrik değişmezlik test edilmeden önce yapısal değişmezliğin sağlandığı varsayılır. Metrik değişmezlikte yapısal değişmezliğe ek olarak farklı alt gruplarda benzer faktör yüklerinin örüntüsünde yer alan maddelerin de artık eşit olacağı şekilde model kısıtlanmıştır (Vandenberg & Lance, 2000). Ölçek değişmezliği test

edilmeden önce ön koşulu olan metrik değişmezlik sağlanmış olmalıdır. Ölçek değişmezliği, farklı alt gruplarda aynı gizil yapıya sahip bireylerin aynı puanı almasıdır (Milfont & Fischer, 2010). Ölçme değişmezliği modelinin son aşaması olan katı değişmezliği, faktör yapısı, faktör yükleri ve madde sabitlerinin yanında hata varyanslarının da eşit olmasıdır (Milfont & Fischer, 2010). Katı değişmezliği test edilmeden önce tüm ölçme değişmezliği aşamaları sağlanmış olmalıdır. Katı değişmezliği, ölçme değişmezliği aşamaları içinde en fazla sınırlandırılmış olanıdır bu yüzden uygulamada bu değişmezlik aşamasının varsayımlarını sağlamak oldukça güçtür (Kıbrıslıoğlu, 2015).

Yapılan ölçümlerde yanlılığın olmaması, araştırmaların ve değerlendirmelerin geçerli ve güvenilir olmasını sağlamaktadır. Yanlılık, ölçme sürecinde farklı yollar ile karışan sistematik hatalar olarak tanımlanabilir (Dorans & Holland, 1993). Sistematik hataların yönü ve kaynağı bilindiği için önlenmektedir (Hambelton, 2006). Bir ölçme aracının geçerliğine ilişkin kanıt toplanırken aynı yapıya sahip birbirinden bağımsız farklı alt gruplardan birçok istatistiksel yöntemden yararlanılabilir. Bunun yanında ölçme aracının herkes için aynı şeyi ifade ettiğinden emin olunmalıdır. Ölçme aracının geliştirilmesi sürecinde maddelerin yazılması, yapının tanımlanması gibi sürecin bütün aşamalarında, ölçme aracının her birey için aynı şeyi ifade edip etmediği doğrulanmalıdır. Ölçme aracının her birey için aynı yapıyı ölçmemesi istenmeyen başka bir boyutu işaret edebilir (Doğan & Öğretmen, 2005).

Hazırlanan bir ölçme aracı, eğer testi alan tüm bireyleri etkiliyorsa bu durum yanlılık olarak algılamak yanlış olacaktır (Zumbo, 1999). Etki ve yanlılık kavramı birbirinden farklıdır. Ölçme araçlarında etkiye daha sık rastlanabilir çünkü testi alan bireyler ırk, cinsiyet gibi açılardan farklılıklara sahiptir. Ancak maddeler ile ölçülmek istenen özellik yetenek veya nitelikler açısından eşleştirildikten sonra maddeleri farklı yönde çalışıyor ise bu durumda yanlılık olabilir ve yanlılığın belirlenebilmesinde değişen madde fonksiyonu analizleri yapılabilir veya yanlılık gösteren maddeler ile ilgili alan uzmanlarının görüşleri alınabilir. DMF, yanlılığı belirleme yöntemlerinden sadece birisidir (Doğan & Öğretmen, 2005). DMF testi alan bireyler arasındaki beklenilmeyen farklılıklardır. Bağımsız farklı alt gruplarda olan

ve aynı yetenek düzeyinde olan bireylerin başarılarındaki farklılıklar DMF yöntemi ile ele alınabilmektedir. Bu farklılıklar cinsiyet, okul türü, sosyoekonomik düzey, kültürel farklılıklar olabilir (Pae & Park, 2006). DMF'nin tespit edilmesinde pek çok yöntem kullanılmaktadır. Araştırmacılar farklı yöntemler ile analizlerin gerçekleştirilmesini, araştırmaların sonucunun karşılaştırılmasını önermektedir (Ackerman, 1992). Bu araştırmada DMF analizlerinde Poly SIBTEST ve OLR yöntemlerinden yararlanılmıştır.

### **Araştırmanın Amacı ve Önemi**

Uluslararası düzeyde ülkelerin eğitim çıktılarını değerlendirebilmek, karşılaştırabilmek ve dolayısıyla da geniş ölçekli sınavların eşdeğer olduğuna dair yürütülen çalışmalar oldukça önemlidir. Bu araştırmanın amacı TIMSS 2019 öğrenci anketi maddelerine verilen cevapların aynı kültür farklı dil (ARP-İNG), farklı kültür farklı dil (İNG-TÜR) çerçevesinde ölçme değişmezliğinin test edilmesi ve DMF'nin farklı yöntemlerle analiz edilerek karşılaştırılmasıdır. TIMSS öğrenci anketine ilişkin ölçme değişmezliği çalışması yapılması, daha sonra ölçme değişmezliği sonuçları göz önünde bulundurularak yanlılık araştırmalarının yapılması literatür için önemli bulunmuştur. Türkiye'de ölçme değişmezliği ve DMF'nin bir arada ele alındığı çalışmaların sınırlı olduğu görülmektedir. Bu çalışma ile ölçme değişmezliği ve DMF ile ilgili yapılacak yeni araştırmalara katkı sağlanması beklenmektedir. Araştırmada bilgisayar kullanımına ilişkin anket maddelerinin ele alınmasının temel sebepleri, teknolojinin hızla gelişmesi, TIMSS döngüsünde 2019 yılında eTIMSS uygulamasına geçilmesi ve bazı ülkelerin değerlendirme döngüsüne bu şekilde katılmış olmasının, bilgisayar ve tablet desteğinin yaygınlaşmasının değerlendirme sonuçlarını etkilemiş olabileceği düşüncesidir.

Literatür incelendiğinde uluslararası sınavlardan biri olan PISA ile ilgili çalışmaların TIMSS ile ilgili yapılan çalışmalara göre daha fazla ele alınmış olduğu görülmüştür. Yapılan araştırmalarda, araştırmacıların bilişsel testlerle ilgili (okuma beceri testi gibi) maddelerle karşılaştırmalar yaptığı gözlenmiştir. Ancak eğitim sistemlerinin karşılaştırılmasında

Matematik, Fen ile ilgili maddelerin yanı sıra öğrenci anketleri, okul anketleri, yönetici anketleri gibi pek çok anket verilerinin olduğu görülmektedir. Buradan anlaşılacağı üzere, uluslararası düzeydeki sınavlarda daha çok iki kategorili maddelerle çalışıldığı, çoklu puanlanan maddelerle ilgili analizlerin daha az incelendiği görülmüştür. Bu anlamda çalışmada dörtlü likert tipindeki maddelerde ölçme değişmezliği ve DMF analizlerine yer verilmesi önemli bulunmuştur. Aynı zamanda bu çalışmanın örnekleme dahil edilen ülkelerin TIMSS sınavını İngilizce dilinde almış olması ve bu sınavın orijinalinin de İngilizce olmasının yapılan karşılaştırmalarda önemli bulgular sağlayacağı düşünülmüştür.

Öğrencilerin anket maddelerine vermiş oldukları yanıtların ölçme değişmezliği ve DMF analizlerini yapmak, bilişsel yeterlik alanları ile ilgili edinilen bilgilerin değerlendirilmesini daha güçlü kılabilir. Bu sayede mevcut maddelere öğrencilerin vermiş oldukları cevaplar göz önüne alınarak değerlendirilmeler yapılabilir, test puanlarına göre yapılmış çıkarımların isabetliliğine yönelik geri dönütler alınabilir ve yeni ölçme araçlarının geliştirilmesine veya mevcut ölçme aracında revizyonlar yapılması sağlanabilir.

Yapılan TIMSS araştırmaları sonuçları konu ile ilgili çalışan uzmanlara, mevcut müfredatlar ve öğrencilerin durumları hakkında bilgi verebilir. Aynı zamanda yapılan çalışmalarla öğrencilere duyuşsal özellikleri ve öğrenme çevrelerinde yer alan kişilerle ilgili edinilen bilgilerle bütüncül düzenlemeler yapılabilir.

Öğrencilerin eğitim hayatı ile ilgili doğru kararlar verebilmek ve ülkenin eğitim sistemi açısından geleceğe yönelik kararlar verirken uygulanan büyük ölçekli testlerin güvenilirlik ve geçerliğini artırabilmek amacı ile uygulamadaki her sınav için bahsedilen analizlerin gerçekleştirilmesi önemlidir. Aynı zamanda mevcut analizlerin sonucunun doğru bir şekilde değerlendirilmesi gelecekte yapılacak olan sınavların planlanması için de uzmanlara öngörü sağlayabilir (Demirus & Gelbal, 2016). Böylece anket maddelerine verilen yanıtlardan yola çıkılarak yanlı olan veya olabilecek maddelerden arınmış şekilde ölçme araçları geliştirilerek, bireyler hakkında daha isabetli kararlar alınması sağlanabilir.

## **Araştırma Problemi**

TIMSS 2019 öğrenci anketinde yer alan bilgisayar kullanımına ilişkin maddeleri, aynı kültür farklı dil ve farklı kültür farklı dil bakımından ölçme değişmezliği ve DMF göstermekte midir?

## **Alt Problemler**

1a) TIMSS 2019 öğrenci anketinde yer alan bilgisayar kullanımına ilişkin maddeleri, aynı kültür farklı dile göre (ARP- İNG) ölçme değişmezliği göstermekte midir?

1b)TIMSS 2019 öğrenci anketinde yer alan bilgisayar kullanımına ilişkin maddeler, farklı kültür farklı dile göre (TÜR-İNG) ölçme değişmezliği göstermekte midir?

2a) TIMSS 2019 öğrenci anketi yer alan bilgisayar kullanımına ilişkin maddeler, aynı kültür farklı dilde (ARP- İNG) OLR yöntemine göre DMF göstermekte midir?

2b) TIMSS 2019 öğrenci anketi yer alan bilgisayar kullanımına ilişkin maddeler, aynı kültür farklı dilde Poly-SIBTEST (ARP- İNG) yöntemine göre DMF göstermekte midir?

2c) TIMSS 2019 öğrenci anketi yer alan bilgisayar kullanımına ilişkin maddeler, farklı kültür farklı dilde (TÜR-İNG) OLR yöntemine göre DMF göstermekte midir?

2d) TIMSS 2019 öğrenci anketi yer alan bilgisayar kullanımına ilişkin maddeler, farklı kültür farklı dilde (TÜR-İNG) Poly-SIBTEST yöntemine göre DMF göstermekte midir?

2e) TIMSS 2019 öğrenci anketi yer alan bilgisayar kullanımına ilişkin maddeler, aynı kültür farklı dilde (ARP- İNG) OLR ve Poly-SIBTEST yöntemleri ile yapılan analiz sonuçları birbiriyle uyumlu mudur?

2f) TIMSS 2019 öğrenci anketi yer alan bilgisayar kullanımına ilişkin maddeler, farklı kültür farklı dilde (TÜR-İNG) OLR ve Poly-SIBTEST yöntemleri ile yapılan analiz sonuçları birbiriyle uyumlu mudur?

## **Sayıtlılar**

Araştırmada TIMSS 2019 testini alan öğrencilerin anketi duyarlılıkla, içtenlikle ve ciddi bir şekilde yanıtlamış oldukları varsayılmıştır.

## **Sınırlılıklar**

Bu araştırma TIMSS 2019 çalışmasına katılan ülkeler dil ve kültüre göre Türkiye ile karşılaştırılmak istenmiştir ancak bu araştırma Türkiye- İngiltere ve Katar karşılaştırması ile sınırlıdır.

## **Tanımlar**

**Değişen Madde Fonksiyonu:** Farklı gruptaki bireylerden oluşan sınav katılımcılarının maddenin ölçmeyi amaçladığı temel yetenek düzeyleri eşleştikten sonra farklı başarı olasılıklarının oluşmasıdır.

**Madde Yanlılığı:** Test maddesi veya durumunun, testin amacı ile ilgili olmayan bazı özellikler nedeniyle bir gruptaki sınav katılımcılarının maddeyi doğru yanıtlama olasılığının diğer grupta yer alan sınav katılımcılarına göre daha düşük veya yüksek olması durumunda ortaya çıkar.

**Test Yanlılığı:** Bir testi alan gruptaki bireylerin için daha az geçerli olduğu fikrinin resmileşmiş halidir.

**Madde Etkisi:** Farklı gruptan sınav katılımcılarının bir maddeye doğru yanıtlama olasılığı farklı olduğu durumlarda ortaya çıkan etkidir. Madde etkisinde madde tarafından ölçülen temel özellikte gerçekten farklılıklar vardır.

## Bölüm 2

### Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

#### Ölçme Değişmezliği

Yapılacak olan bir ölçme işleminde kişilerin bireysel özelliklerinden dolayı farklı sonuçlar alması beklenmektedir. Ancak bu farklılıkların her zaman bireylerin farklı özellikleri ile açıklanması doğru değildir. Çünkü bireyler arası farklılıklar bazen ölçme aracından kaynaklanabilmektedir (Başusta, 2010). Çeviri ve uyarlama çalışmaları ile farklı gruplar arasındaki eşitsizlik ortadan kaldırılmaya çalışılmaktadır ancak bazı durumlarda maddelerden kaynaklanan ve bireylerin yanıtlarını etkileyebilen durumlar söz konusu olabilmektedir. Bireylerin içinde bulunduğu yaşam koşulları, kültürel özellikleri maddeleri farklı algılamasına sebep olabilir.

Ölçme değişmezliği ile ölçme aracının, ölçmek istediği özelliğin, evren içerisinde birbirinden bağımsız örneklem için aynı yapı geçerliğine dair kanıtlar sunmasıdır (Şekercioğlu, 2018). Başka bir ifadeyle ölçme aracı ile ölçmeyi amaçlanan özellikler, farklı örneklem için aynı faktör yapısına sahip olmalıdır. Gruplar arasında geçerli karşılaştırmalar yapabilmek için kullanılan ölçme aracının ölçme değişmezliği incelenmelidir. Ölçme değişmezliğinin sağlandığı durumlarda yapı geçerliğine ilişkin kanıtlar da daha güçlü olacaktır.

Farklı dili konuşan pek çok ülkede veya farklı kültürlerin olduğu bölgelerin karşılaştırılmasında ölçme aracının uygulanacağı ülkenin diline uyarlanması yapılırken, farklı endüstri alanlarında çalışan grupların karşılaştırılmasında cinsiyete veya etnik farklılıklara göre yapılacak olan karşılaştırmalarda ölçme değişmezliği analizleri kullanılmaktadır (Cheung & Rensvold, 2002).

Ölçme değişmezliğinin sağlanması, bir başka ifadeyle kullanılan ölçme aracının geçerliğine dair kanıtların olması gerekmektedir. Elde edilen geçerlik kanıtları ile testi

hazırlayan uzmanlar ve testi alan bireyler için ölçme aracının aynı anlamı ifade ettiği varsayılmaktadır.

Geçerliğe ilişkin kanıtlardan biri olan yapı geçerliğinin, diğer geçerlik kanıtlarının temelini oluşturduğu 1954 yılında yayınlanmış olan Test Standartları raporunda yer edinmiştir (Cronbach & Meehl, 1955). Özetle bir ölçme aracında başlangıçta kapsam veya ölçüt geçerliği için kanıtlar edinilmiş olsa bile bireylerin testlerdeki performanslarının hesaplanmasında testin yapı geçerliğini de ele almak gereklidir (Kane, 2001). Kline (2016), yapı geçerliğinin diğer geçerlik kanıtlarını içerdiğini, diğer geçerlik türlerinin yapı geçerliği ile ele alınması gerektiğini savunmuştur. Yapı geçerliğine ilişkin kanıt toplanırken sıklıkla faktör analizine başvurulur. Faktör analizi ile birbirleri ile yüksek derecede ilişkili olan çok sayıda değişken, anlamlı ve daha az sayıda değişken ile analizlerin gerçekleşmesine imkân sağlar (Büyüköztürk, 2002). Faktör analizi ile araştırmacılar doğrudan ölçme yapmanın zor olduğu ve ölçülmek istenen birden fazla değişken olduğu durumlarda, birden fazla olan faktörün bir araya getirilerek daha anlamlı olabilecek bir ana değişken olarak ele alınmasını sağlar (Akbay, 2021). Faktör analizinde faktörlerin, değişkenlerin nedeni olduğu düşünülür. Bir başka ifadeyle değişkenlere neden olan yapı, değişkenler üzerinde puanlar üretir. Faktör analizindeki temel soru şu şekilde ifade edilebilir: Değişkenler arasındaki korelasyonları neler üretmiş olabilir (Tabachnick & Fidell, 2013)?

Faktör analizi, yeni bir ölçek geliştirilme süreci veya yeni bir yapı tanımlanmasında açıklayıcı faktör analizi (AFA); mevcut olan bir yapının doğrulanması, doğrulayıcı faktör analizi (DFA) olarak araştırmalarda kullanılmaktadır (Doğan & Aybek, 2021). Açıklayıcı faktör analizi, birbiriyle ilişkili birden fazla değişkenin daha az sayıda faktörle ele alınmasıdır (Doğan & Aybek, 2021). Diğer bir ifadeyle belirli bir değişkenin altında toplanmış olan göstergelerin, mevcut kuramsal yapıya ait olup olmadığı anlaşılmaya çalışılır. Doğrulayıcı faktör analizinde, önceden test edilmiş ve tanımlanmış olan yapının doğruluğu test edilir. Doğrulayıcı Faktör analizi yapı geçerliğine dair güçlü bir kanıt sunmaktadır. Doğrulayıcı faktör analizinin, açıklayıcı faktör analizine göre avantajı, tüm ölçümlerin veya yapıların



eşdeğerliğinin incelenebilmesidir (Brown, 2006). Başka bir ifadeyle ölçme için kullanılan modeller, ölçülen özelliklerin göstergeleri ile ilgili olduğundan kurulan modellerde faktör yükleri, kesişimler ve artık varyanslar incelenerek bilgi edinilir.

Ölçme değişmezliğinde incelenen yapının her grup için aynı anlamı anlama gelip gelmediğini incelemenin yanı sıra bu analiz yöntemi ile evrenin heterojenliğinin de test edilmesi durumu da ortaya çıkmaktadır. Literatür incelendiğinde faktör yapılarının eşitliğinin yapısal değişmezlik, metrik değişmezlik, ölçek değişmezliği ve katı değişmezliği gibi farklı değişmezlik aşamaları ile test edilip incelendiği görülmüştür. Meredith (1993) yaptığı çalışmada ölçme değişmezliğinin bu dört aşamasına değinmektedir. Vandenberg ve Lance (2000) çalışmalarında ölçme modelinin test edilmesinin öncesinde kovaryans matrislerinin test edilmesinin gerekliliğini vurgulamışlardır. Ölçme değişmezliği aşamalarının adlandırılmasını da şu şekilde ele almışlardır; yapısal değişmezlik, metrik değişmezlik, ölçek değişmezliği, özgün faktör değişmezliği. Literatür incelendiğinde ölçme değişmezliğinin test edilme aşamalarının sınıflandırılmasında fikir birliğinin olmadığı görülmektedir. Bu çalışmada Meredith'in 1993 yılındaki çalışmada önerdiği sınıflama doğrultusunda analizler gerçekleştirilmiştir.

Ölçülen psikolojik özelliklerin farklı dil ve kültürel değişkenler fark etmeden gruplar arası karşılaştırmalar yapılmasına izin vermesi, maddelerin aynı yapıyı ölçtüğü durumlarda ölçme değişmezliğinden bahsedilebilir. Bir diğer ifadeyle ölçme değişmezliğinin sağlanması için maddelerin faktör yapılarının ve hata varyanslarının aynı olması gerekmektedir. Ölçme değişmezliği grupların anlamlı olarak karşılaştırılmasında kritik öneme sahiptir (Şekercioğlu, 2018).

### ***Çoklu Grup Doğrulayıcı Faktör Analizi (ÇGDFA) ve Ölçme Değişmezliği***

ÇGDFA, doğrulayıcı faktör analizinin içerisinde yer almakla birlikte, ÇGDFA'nın daha detaylı bir yöntem olduğu söylenebilir. Gruplar arası ölçme değişmezliğinin incelenmesinde, faktör yapıları bakımından incelemelerde ÇGDFA'ya başvurulabilir. Bu yöntem ile farklı tüm gruplar arasında değişmezliğin yönleri incelenebilir (Brown, 2006). Bu

analiz yöntemi ile birden fazla grubun parametrelerinin yapısal anlamla eşitliğinin aynı anda test edilebilir (Şekercioğlu, 2018).

Kline (2016) ÇGDFA'da önemli bir noktanın farklı gruplarda aynı gizil değişkenlerin ölçülüp ölçülmediğinin test edilmesi olduğunu vurgulamıştır. Böylece aynı yapıda olan maddelerin evrenin tüm alt gruplarında kullanılıp kullanılmayacağı araştırılır.

Byrne (2006) ÇGDFA ile ölçme değişmezliği analizlerinin gerçekleştirilmesi sürecinde beş temel soruya cevap aranmasını önermiştir: Ölçme aracında yer alan maddeler farklı gruplarda aynı şekilde çalışıyor mu? Ölçme aracının faktör yapısı ölçülmek istenen özelliği grubun her düzeyinde eşit şekilde ölçüyor mu? Deneysel yapılar her grupta aynı mı? Modelde oluşan yapıların ortalamaları gruplar arasında farklı mı? Ölçme aracının faktör yapısı evrende yer alan bağımsız örneklemelerde eşit mi? Analizler sonucunda yukarıda verilen bu sorulara yanıt alınamaması durumlarında yapı geçerliğinin azalacağı, gruplar arası karşılaştırmasına dayanan çıkarımların doğru olmayacağı söylenebilir.

Herhangi bir grubun yapısına göre geliştirilen ölçme araçları, daha sonra farklı gruplar için de kullanılmaktadır böyle bir durum söz konusu olduğunda da ölçme değişmezliği sorunu ortaya çıkmaktadır (Borsboom, 2006). Ölçme değişmezliğinin test edilmesinde birden fazla hiyerarşik testin uygulanması daha geçerli sonuçlara ulaşılmasını sağlar (Byrne, 2008).

ÇGDFA ile ölçme değişmezliği aşağıda Eşitlik 1 denkleminde yer verilen faktör modeli ile test edilebilir;

$$X_{jk} = \tau_{jk} + \gamma_{jk}W_{JK} + u_{jk} \quad (1)$$

“ $\tau_{jk}$ ” = ifade edilen değişken gözlenen ve örtük yapılar arasındaki katsayı faktörünü,

“j” = ölçülen değişkeni,

“k”= ilgili grubu,

"i" = birey

"W<sub>jk</sub>"= r x 1 desende i sayıdaki birey için ortak faktör yükleri vektör matrisini,

"Y<sub>jk</sub>"= r madde sayısı için, r x 1'lik faktör yükleri matrisini,

"u<sub>jk</sub>"= bağımsız olarak gözlenen değişkenlerin hata vektörünü ifade etmektedir.

Aynı zamanda hataların birbirleriyle ve ortak faktör puanları ile ilişkisinin sıfır olduğu varsayılmaktadır. Bu varsayım ile ilgili kovaryans denklemi ise Eşitlik 2'de şu şekilde ifade edilebilir;

$$cov(X_{ijk}) = \sum_k = \lambda_k \phi_k \lambda_k + \theta_k \quad (2)$$

Eşitlik denkleminde  $\Lambda_k$  satırları Y<sub>jk</sub>'den oluşan p x r desenindeki matrisi (p boyut),

$\Phi_k$  daki varyans ve kovaryansları,

$\theta_k$  ise hataların köşegen matrisi ifade etmektedir.

Başka önemli bir vektör ise Eşitlik 3'te şu şekilde ifade edilebilir;

$$E(X_{ik}) = \mu_k = \tau_k + \Lambda_k K_k \quad (3)$$

Eşitlik 3'teki denklemde  $\tau_k$  ölçme hatalarının vektörünü ifade etmektedir.

Daha önce de bahsedildiği üzere ölçme değişmezliğinin test edilmesi aşamaları farklı araştırmacılar tarafından farklı şekillerde adlandırılmıştır.

ÇGDFA'da oluşturulan model karşılaştırılacak olan gruplar için kabul edilebilir olmalıdır. Yani her grupta tahmin edilen faktör yükleri anlamlı olmalı model uyum indeksleri incelenmelidir.

Ölçme değişmezliğinin ÇGDFA ile test edilmesinde farklı araştırmacı tarafından farklı aşamalarda test edilebildiğine daha önce değinilmiştir. Bu çalışmada Meredith'in

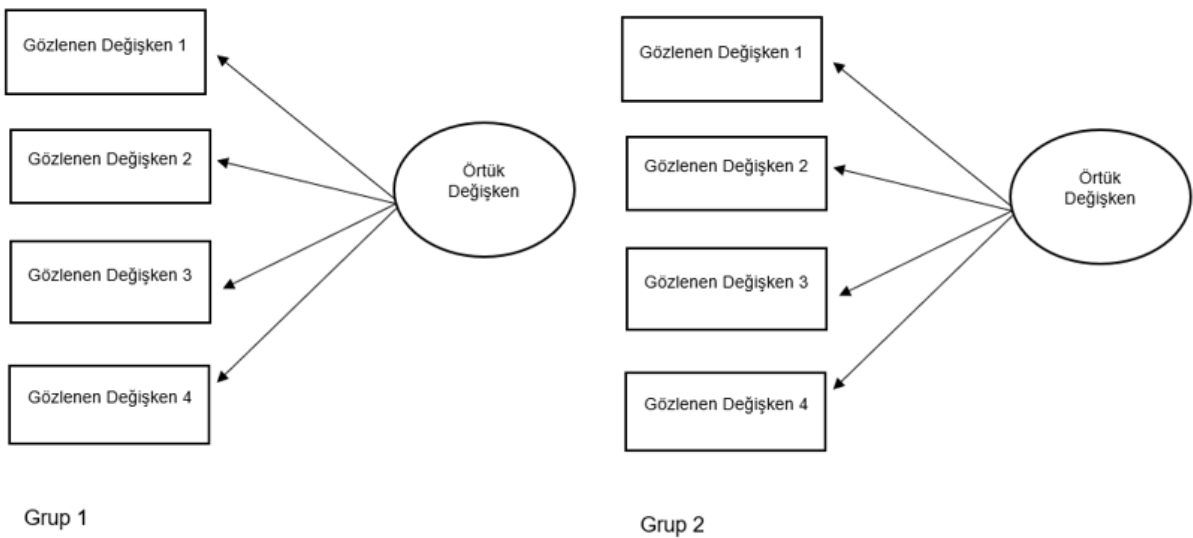
(1993) önermiş olduğu aşamalar test edilmiştir. Buna değişmezlik testleri ise aşağıda açıklanmıştır.

**Yapısal (Şekil, Biçimsel) değişmezlik (Configural invariance).** Eşdeğer gerçek puanlar veya yapılarda ilk varsayım yapısal değişmezliktir. Bu aşamada ölçümlerin kavramsal olarak farklı alt gruplarda aynı yapıların var olması anlamına gelir. Yapılmış olan bir dizi ölçümün bir grup için başka bir anlama gelmesi, diğer bir grupta ise farklı bir anlama sahip olması istenmemektedir. Böyle bir durumun karşılaşılması halinde ölçme değişmezliğinin diğer aşamalarına geçilemeyecektir (Vandenberg & Lance, 2000).

Yapısal değişmezlik, her grupta tüm parametrelerin serbest kestirilmesi durumunda faktör yükleri, faktör korelasyonları ve hata varyanslarının tüm alt gruplarda aynı olmasıdır (Kline, 2016). Yapısal değişmezliğin sağlanmış olması diğer ölçme değişmezliği aşamaları için ön koşul niteliğindedir. Bu koşulun sağlanmamış olması diğer ölçme değişmezliği aşamalarının da sağlanamayacağı anlamını taşır (Kline, 2016). Yapısal değişmezlik modeli örneğine Şekil 1’de yer verilmiştir (Somer vd., 2009).

## Şekil 1

### Yapısal Değişmezlik Modeli Örneği



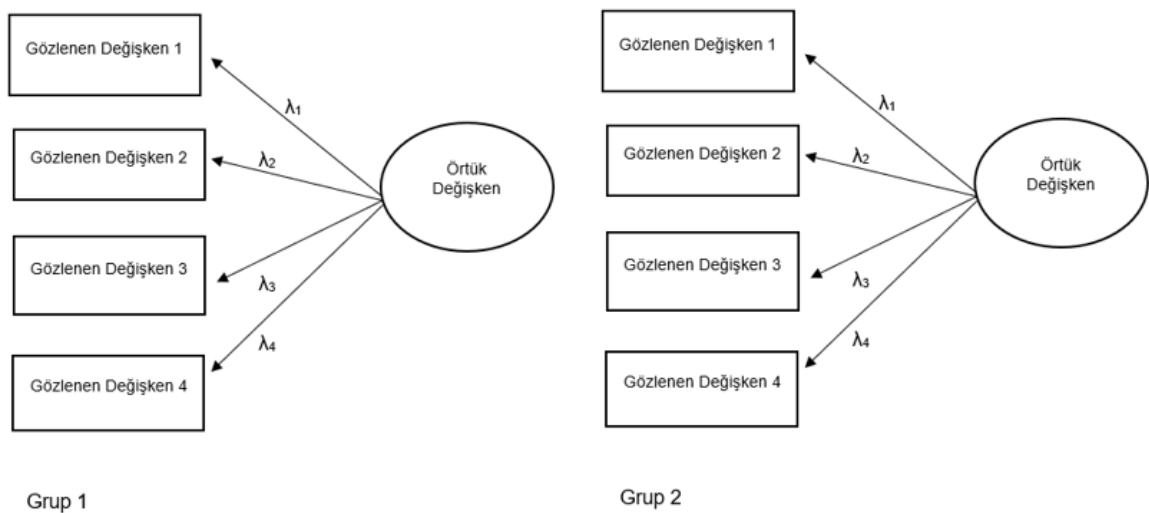
**Metrik Değişmezlik (Model Değişmezliği ya da Zayıf Değişmezlik).** Metrik değişmezlik, yapısal değişmezliğin sağladığını varsayar ve ek olarak standartlaştırılmamış

olan model katsayılarının da eşit olduğunu ileri sürer. Metrik değişmezlik sağlandığında, ölçülmek istenen yapı farklı alt gruplarda benzer şekilde ortaya çıkar (Kline, 2016).

Metrik değişmezlik, yapısal ölçme değişmezliği testinden daha güçlü sonuçlar veren faktöriyel değişmezlik testidir. Bu değişmezlik testi ile bir faktör modeli belirlemenin yanında benzer faktör yüklerinin örüntüsünde yer alan maddeler artık eşit olacak şekilde kısıtlanmıştır (Vandenberg & Lance, 2000). Maddelere farklı alt gruplarında yer alan bireylerin aynı biçimde cevap vermiş olması metrik değişmezliği testini varsayımının karşılandığı anlamını taşımaktadır. Yani farklı ülkelerde yaşayan bireylerin puanlarının karşılaştırılmasının anlamlı olduğu söylenebilir (Steenkamp & Baumgartner, 1998). Bu ölçme değişmezliği testinin sağlanmış olması, farklı alt gruplardaki bireylerin aynı faktör yapısı ve faktör yüklerine sahip olması beklenir (Cheung & Rensvold, 2002). Metrik değişmezliğinin sağlanmamış olduğu durumlarda farklı gruplarda yer alan bireylerin aynı testte bulunan maddeleri gözle görülür derecede farklı yorumlamış olduğu söylenebilir (Byrne, 2013). Örnek metrik değişmezlik modeline Şekil 2'de yer verilmiştir (Somer vd, 2009).

## Şekil 2

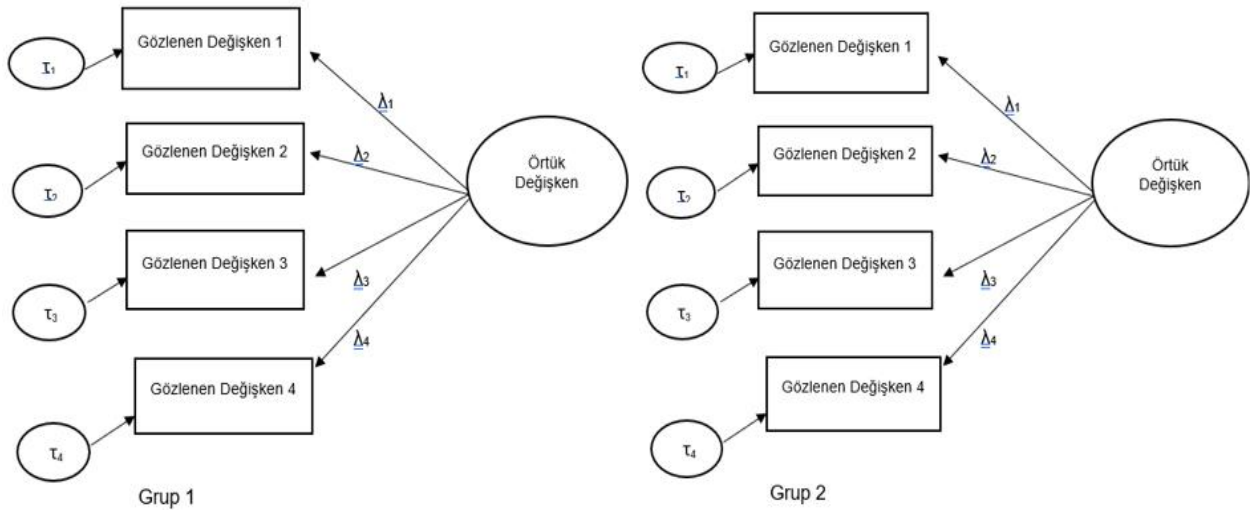
### Metrik Değişmezlik Modeli Örneği



**Ölçek Değişmezliği (Güçlü Değişmezlik ya da Skalar Değişmezlik).** Metrik değişmezlik varsayımlarının sağlanması durumunda ölçek değişmezliği test edilebilir. Ölçek değişmezliği, faktör gruplarının, faktör yüklerinin ve madde sabitlerinin sınırlı tutulması ile test edilebilir ölçek değişmezlik farklı alt gruplarda yer alan bireylerin (gizli) ortalamalarının karşılaştırılabilirliğinin test edilmesinde gereklidir. Bu değişmezliğe göre gözlenen ortalamalar örtük puanlar ile ilgili olmalıdır. Başka bir ifade ile aynı gizil yapıya sahip farklı alt gruplarda yer alan bireylerin aynı puanı almış olması ölçek değişmezliğinin sağlanmış olduğunu gösterir (Milfont & Fischer, 2010). Ölçek değişmezliği varsayımının sağlanması, farklı alt gruplardaki bireylerde, maddelerin gizil yapıları açıklama düzeylerinin benzerlik göstermesidir. Örnek ölçek değişmezliği modeline Şekil 3'te yer verilmiştir (Somer vd., 2009).

### Şekil 3

#### Ölçek Değişmezliği Modeli Örneği

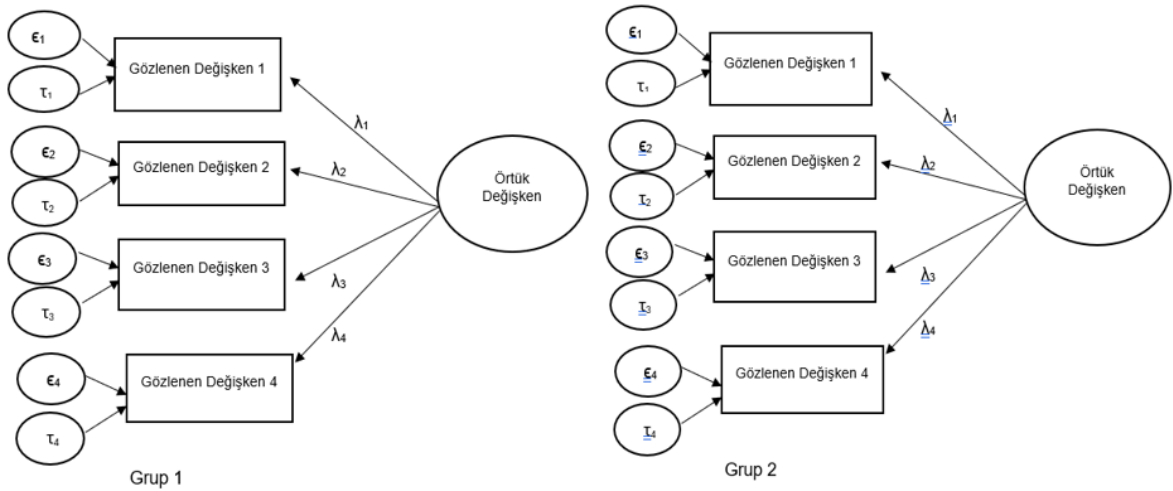


**Katı Değişmezlik (Strict Invariance).** Katı değişmezlik varsayımlarının sağlanması diğer ölçme değişmezliği aşamalarında test edilen varsayımların sağlanmasından daha güçtür. Bu ölçme değişmezliği hata varyansı değişmezliği olarak da adlandırılabilir. Katı değişmezlik, faktör yapısı, faktör yükleri ve madde sabitlerinin eşitliğinin yanında artık varyansların da eşitliğini gerektirir. Farklı alt gruplarda, ölçme aracının maddelerinin aynı

düzeyde ölçüm hatasına sebep olmasıdır başka bir ifade ile gruplar arasında tüm hata varyanslarının eşit olmasıdır (Milfont & Fischer, 2010). Farklı alt gruplarda yer alan aynı faktör seviyesine sahip birey gösterge üzerinde aynı puanı elde eder (Kline, 2016). Katı değişmezlik oldukça sınırlandırılmış varsayımlara sahip olduğu için uygulamada karşılanması oldukça güçtür (Kıbrıslıoğlu, 2015). Ayrıca birçok araştırmacıya göre gizil değişkenlerin ortalamalarının karşılaştırılması amaçlandığında katı değişmezlik varsayımlarının test edilmesinin gerekli olmadığı belirtilmiştir (Schmitt & Kuljanin, 2008). Örnek katı değişmezlik modeline Şekil 4'te yer verilmiştir (Somer vd., 2009).

#### Şekil 4

##### Katı Değişmezlik Modeli Örneği



Ölçme değişmezliğinin test edilmesinde hipotezler belirli bir aşamalılık göstermektedir. Her aşamada oluşturulan model bir önceki ölçme değişmezliği için oluşturulmuş olan modeli de içerecek şekilde oluşturulur. Oluşturulan modelde önceki aşamadaki model ile karşılaştırılır; iki modelin uyumlu sonuçlar vermediği durumlarda bir sonraki ölçme değişmezliği varsayımı test edilmez. Çünkü ölçme değişmezliğindeki varsayımlar arasında aşamalılık söz konusu olup, her ölçme değişmezliği modeli bir sonraki

modelin ön koşuludur (Kıbrıslioğlu, 2015). Ölçme değişmezliği aşamaları Tablo 1’de özetlenmiştir;

**Tablo 1**

*Ölçme Değişmezliği Özet Tablosu*

Değişmezlik Hipotezi	Değişmezlik Varsayımları	Kanıtlanabilir Grup Karşılaştırmaları
Yapısal Değişmezlik	+Madde/ Faktör grupları	-
Metrik Değişmezlik	+Faktör yükleri	• Faktör varyans ve kovaryansları
Ölçek Değişmezlik	+Madde sabitleri	• Faktör ve gözlenen değişken ortalamaları
Katı Değişmezlik	+Madde artık varyansları	• Gözlenen varyans ve kovaryanslar

Kaynak: Gregorich, S. E. (2006).

Tablo 1 ‘de Ölçme değişmezliği aşamaları özetlenmiştir. Tablo incelendiğinde değişmezlik hipotezlerinin yanı sıra ilgili farklı alt grupların belirli kısıtlamalar yapıldığında karşılaştırmalarına da yer verildiği görülmektedir. Gözlenen veya faktör ortalamaları, varyansları ve kovaryanslarının gruplar arasında karşılaştırılır ancak yapılan karşılaştırmaların amacı ölçme aracının yapısının, ortalamaları ve varyasyonlarındaki grup farklılıklarında sağlanıp sağlanmadığının test edilmesidir (Gregorich,2006).

***Uyum İndeksleri***

Oluşturulmuş olan modelin, araştırmada kullanılan veri ile ne kadar uyumlu olduğunu karşılaştırmak için standart değerlere ihtiyaç duyulmuştur (Erşan, 2016). Uyum istatistikleri model test istatistikleri ve yaklaşık uyum istatistikleri olarak iki temel bulunmaktadır (Kline, 2016). Model istatistiklerinde modelde karşılaştırılmak istenen kovaryans matrisi ile analizler sonucu elde edilen örneklemin kovaryans matrisi ile ne kadar uyumlu olduğu test edilir (Kıbrıslioğlu, 2015). Model istatistiklerinde anlamlı sonuçlara ulaşılması model veri uyumunun ulaşılmadığını gösterir. En temel model test istatistiğinin ki-kare istatistiği olduğu bilinmektedir. Ki kare istatistiğinde örneklemin büyüklüğünün etkisini en aza indirgeyebilmek için normlaştırılmış ki kare istatistiği kullanılır.



Normlaştırılmış ki kare istatistiğinde ulaşılan sonuç değeri küçüldükçe model uyumunun artmış olduğu çıkarımı yapılabilir (West vd., 2012).

Yaklaşık uyum indekslerinde ise model test istatistiklerindeki gibi uyumu var- yok şeklinde ikili bir sonuç yerine model veri uyumunu gösteren sürekli ölçümler bulunmaktadır. Model-veri uyumunda elde edilen yüksek sonuçlar model veri uyumunun da yüksek olduğunu gösterir (Kıbrıslıoğlu, 2015). Literatür incelendiğinde uyum iyiliğinin değerlendirilmesinde kullanılan birden çok uyum iyiliği indeksi bulunmaktadır. Araştırmalarda sıkça yer verilen uyum indekslerinden bazıları normlaştırılmış ki-kare, CFI, SRMR, RMSEA ve TLI şeklinde sıralanabilir. Literatürde CFI ve RMSEA uyum indekslerinin sıklıkla tercih edildiği görülmektedir (Tabachnick & Fidel, 2013). Araştırmacılar SRMR model uyum indeksinin de raporlanması gerektiğini vurgulamışlardır (Hu & Bentler, 1999).

**CFI.** Karşılaştırmalı uyum indeksi olan CFI değeri, araştırmacıların kurmuş oldukları mevcut modelden bağımsız bir şekilde belirlenen bir taban modeline göre model uyumunun ölçülmesidir. CFI indeksi uyum iyiliği indeksidir (Kline, 2016). CFI uyum indeksi küçük örneklem için iyi sonuçlar verebilir. CFI indeksi sonuçlarına göre .95 ve üzeri değerler model veri uyumu açısından kabul edilebilir değerlerdir (Hu & Bentler, 1999). Test edilmek istenen ana model ile kurulan modelin ne derece uyumlu olduğunun test edilmesinde CFI uyum indeksinden yararlanılabilir (Tabachnick & Fidel, 2013). CFI değeri, ki kare serbestlik derecesine eşit veya bu değerden daha küçükse CFI değeri 1 olarak kabul edilir. Ancak diğer durumlarda CFI değerinin hesaplanmasında kullanılan denklem Eşitlik 4 'te verilmiştir.

$$CFI = 1 - \frac{X_M^2 - df_M}{X_B^2 - df_B} \quad (4)$$

**RMSEA.** Yaklaşık hataların ortalama karekökü (RMSEA) indeksinin uyum kötülüğü indeksi olduğu söylenebilir. RMSEA merkezi olmayan ki kareyi izler. Merkezizlik parametresi (noncentrality parameter) modelin ne derece kötü uyum gösterdiğine bağlı olarak değişir. RMSEA denkleminde Eşitlik 5'te yer verilmiştir.

$$RMSEA = \sqrt{\frac{x^2 - d_f}{d_f (N-1)}} \quad (5)$$

RMSEA indeksinin alabileceği en küçük değer 0'dır. Teorik olarak üst sınırının olmadığı bilinmektedir. RMSEA değerinin küçük çıkması model uyumunun arttığını göstermektedir. Asgari düzeyde kabul edilebilir RMSEA değerinin  $\leq .05$  olması beklenir. Elde edilen katsayı eğer .10 değerinin üzerinde olduğunda modelin veri ile uyumsuz olduğu yorumu yapılabilir. RMSEA değeri örneklemin büyüklüğünden olumsuz yönde etkilenen bir indeks olduğu söylenebilir. Diğer bir ifade ile RMSEA indeksi küçük örneklerde beklenenden daha az kestirim yapabilir (West vd., 2012).

**SRMR.** SRMR uyum indeksinde gözlenen ve kestirilen kovaryans matrisleri temel alınarak hesaplamalar yapılır. Elde edilen kovaryans matrisleri korelasyon matrislerine dönüştürülür. SRMR değeri korelasyon artıklarının mutlak değerlerinin ortalamasının ölçüsünü vermektedir (Kline, 2016). Model veri uyumunun sağlandığı çıkarımının yapılabilmesi için asgari düzeyde kabul edilebilir SRMR değeri .08 'den küçük değerlerdir. Ancak elde edilen SRMR uyum değerinin .05'ten küçük olması mükemmel model veri uyumunun sağlanmış olduğunu göstermektedir (Hu & Bentler, 1999).

**GFI.** GFI yaklaşık uyum indeksinin 0 ile 1 arasında değer alır, 1 en iyi uyumun gösterir (Kline,2016). GFI uyum indeksinin genel formülü Eşitlik 6'da ifade edilmiştir.

$$GFI = 1 - \frac{C_{res}}{C_{tot}} \quad (6)$$

Eşitlik 6'da yer verilen formül incelendiğinde örneklemdaki kovaryans matrisinin artık değişkenlik ( $C_{res}$ ), ve toplam değişkenliğin ( $C_{tot}$ ) kullanılarak uyum iyiliği hesaplamalarının yapıldığı görülmüştür. Bu uyum indeksi hesaplama yönteminin en büyük sınırlılığı örneklem büyüklüğünden etkilenmesidir (Kline, 2016).

**TLI.**Tucker-Lewis indeksi kısaca TLI olarak da adlandırılan uyum indeksi, kareler toplamının serbestlik derecesine oranlanarak elde edilen uyum fonksiyonu F 'i temel alır (Kıbrıslioğlu,2015). Hipotezi kurulan modelin ortalama değeri ile taban modelin ortalama değeri karşılaştırılır (West vd., 2012). TLI uyum iyiliği indeksidir, bu indeks 0 ile 1 arasında değerler almaktadır. Örneklem büyüklüğünden etkilenmektedir. Özellikle küçük örneklerde daha küçük model veri uyum indeksi değeri elde edilebileceği söylenebilir. Bu durumda modelin veri ile kötü uyum gösterdiği çıkarımı yapılır (Tabachnick & Fidel, 2013).

**Tablo 2**

*Uyum İyiliği İndeksleri*

İndeks	İyi Uyum	Kabul Edilebilir Uyum
$\chi^2$	$p > 0.05$	$p > 0.05$
$\chi^2/sd$	$0 \leq \chi^2/sd \leq 3$	$3 \leq \chi^2/sd \leq 5$
RMSEA	$0 \leq RMSEA \leq 0.05$	$0.05 \leq RMSEA \leq 0.08$
GFI	$0.95 \leq GFI \leq 1.00$	$0.90 \leq GFI < 0.95$
AGFI	$0.95 \leq AGFI \leq 1.00$	$0.90 \leq AGFI < 0.95$
CFI	$0.95 \leq CFI \leq 1.00$	$0.90 \leq CFI < 0.95$
TLI	$0.95 \leq TLI \leq 1.00$	$0.90 \leq TLI < 0.95$

Kaynak: Hu ve Bentler, 1999

Tablo 2'de modelin uyumluluğunun belirlenmesinde sıkça kullanılan iyim uyum ve kabul edilebilir uyum iyiliği indekslerine yer verilmiştir.

**Değişen Madde Fonksiyonu (DMF) ve Yanlılık**

Ulusal ve uluslararası sınav uygulamalarında gruplarda yanlılık olup olmadığının incelenmesi ile ilgili araştırmalar geçmişten günümüze devam etmektedir. Yapılan sınav uygulamalarında kullanılan ölçme aracında yanlılık bulunması testin geçerliğini tehdit edici bir unsurdur. Testin geçerliği testten yapılan çıkarımlar ile mümkün olmaktadır. Aynı

zamanda bir test objektif gibi görünse bile çıkarımların geçerliği ile testin yanlılığı değerlendirilmelidir (Zumbo,1999). Farklı alt gruplarda testi alan bireylerin testten aldığı puanlar incelenmeli ve alınan puanların kaynakları incelenmeli ve ölçme aracından kaynaklanıp kaynaklanmadığı kontrol edilmelidir (Atalay, 2010). Özetle ölçme aracının yanlılık içermesi geçerliği olumsuz etkilemekte, sistematik hatanın ortaya çıkmasına neden olmaktadır (Camili & Shephard, 1994). Yanlılık, ölçüm anormalleri olarak tanımlanabilir ve bu adaletsizliğin hangi grup için avantaj hangi grup için dezavantaj sağladığı araştırılması gereken bir konudur. Bireyler ile ilgili karar alma sürecinde yapılan tahminlerde tesadüfi hatalar vardır ancak yanlılık, sistematik bir fark anlamına gelmektedir (Camili, 2006). Öğrencilerle ilgili önemli karar verilmesi gereken durumlarda kullanılan ölçme aracının yansız olması daha geçerli değerlendirmelerin yapılabilmesinde önemlidir. Testin adil olması eşit çıktılardan ziyade grup farklılıklarının test kapsamının dışındaki faktör yapılarından kaynaklanıp kaynaklanmadığı anlamına gelmektedir (Camili, 2006). Ancak her durum yanlılık olarak değerlendirilmemelidir. Eğer testi alan bütün öğrencileri etkileyen bir madde varsa bu durum yanlılıktan ziyade maddeden kaynaklı bir durumdan bahsedilebilir (Zumbo, 1999).

Zumbo (1999), madde yanlılığını incelemek için iki yaklaşım bulunduğunu ifade etmiştir. Bu yaklaşımlardan biri olan yargısal yaklaşımda potansiyel olarak yanlı olarak görülen maddelerin bir veya daha fazla uzman tarafından incelenmesidir. Yargısal yaklaşım ile maddeleri incelendikten sonra yanlı olduğu düşünülen potansiyel maddelerin ikinci yaklaşımla yani istatistiksel tekniklerle test edilmesi gerekmektedir. Madde yanlılığının araştırılmasında dış veya iç ilişkileri incelenmektedir. Maddenin dış ilişkileri çalışmalarında daha çok kriter- ölçek ilişkileri araştırılmaktadır. Puanlar ve ölçütler arasındaki ilişkilerde çeşitli açılardan farklılıklar bulunduğu madde yanlılığına ilişkin dışsal kanıtlar olduğu düşünülebilir. İçsel ilişkilerde ise maddelerin birbirleri arasındaki ilişkilerde kanıt aranmasıdır. DMF madde içi ilişkilerle ilgilidir.

DMF, ölçme aracının standart bir değerlendirme sağlayıp sağlamadığı yani test adaletsizliğinin olup olmadığını araştırmak için kullanılan yöntemlerin başında gelmektedir (French & Miller, 1993). DMF, yanlılığın belirlenmesinde gereklidir ancak madde yanlılığını belirlemede tam anlamıyla yeterli değildir (Clauser & Mazor, 1998).

DMF, farklı iki veya daha fazla grubun üyelerinin başarı durumları açısından farklı olup olmadığı ortaya konmasıdır. DMF kullanılmadan önce bu gruplardaki katılımcıların yetenek düzeyleri bakımından eşleştirilmesi gerekmektedir. (Zumbo,1999).

Scheuneman (1979), DMF'nin daha iyi anlaşılabilmesi için bazı kavramlar tanımlamıştır. Aynı zamanda test maddelerinin yansız sayılabilmesi için homojen alt testlerde aynı puan almış kişilerin oranının da aynı olması gerekliliğini vurgulamıştır.

Lord (1980), madde tepki kuramı kapsamında DMF'yi şu şekilde tanımlamıştır: Her bir test maddesi gruptaki her bireyde tam olarak madde tepki fonksiyonu göstermiş olsaydı, aynı yetenek düzeyinde olan bireyler aynı test maddesini doğru cevaplama konusunda aynı şansa sahip olurlardı. Ancak bir test maddesi farklı gruptaki bireylerde farklı madde tepki fonksiyonuna sebep oluyorsa maddenin yanlı olduğu söylenebilir.

Thissen vd., (1988), Schmitt ve Dorans (1993) DMF üzerinde tartışmalar gerçekleştirirken yapılan tanımlara eklemeler yapmışlardır. Farklı örneklerde yer alan bireylerin yeteneklerini ölçmek için kullanılan testlerin yanlı olması, testin geçerliğini olumsuz yönde etkilediğini dile getirmişlerdir. Aynı test maddesi farklı grupları karşılaştırmada kullanıldığı zaman geçerliği azaltarak, testten elde edilen puanların testin amacı dışında özellikleri gösteriyor olabileceğini ifade etmişlerdir.

DMF'nin belirlenebilmesi için istatistiksel yöntemlerden yararlanır. Kullanılan yöntemler aracılığı ile farklı örneklerdeki bireylerin, teste yer alan her bir maddenin anlamının aynı olup olmadığı test edilir. DMF analizleri sonuçlarının doğru şekilde anlamlandırılmasıyla yapı ve yordama geçerliği kanıtları güçlendirilebilir (Clauser & Mazor, 1998).

Etki ve DMF arasındaki farkı anlamak önemlidir. Etki, iki grup arasında performans farklılıklarını ifade eder. Test maddelerinde etkiye rastlanabilir çünkü testi alan bireyler ırk, cinsiyet gibi açılardan farklılıklara sahip olabilir. Ancak maddenin ölçtüğü özellik yetenek veya nitelikler açısından farklı gruplarda eşleştirmeler yapıldıktan sonra maddelerin farklı yönde işlev gösteriyorsa DMF'den yararlanılabilir. DMF etkiden farklıdır; çünkü testin ölçtüğü özellikler testi alan bireyler arasında beklenilmeyen bir farklılığa sebep olur.

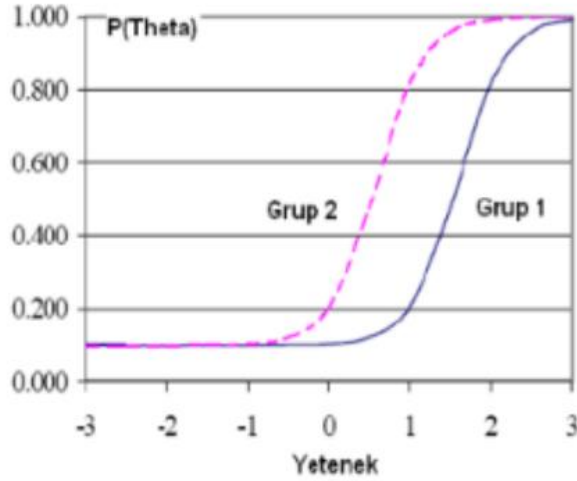
Zumbo (1999), DMF kullanılacağı zaman beş noktaya dikkat edilmesini önermektedir. Bunlardan ilki farklı kültürleri sahip pek çok alt grubun var olduğunun farkında olunarak kişisel ve ahlaki konularda net olunmalı, gruplar arasında karşılaştırmalar yapılırken belirli standartların (cinsiyet, ırk, alt kültür veya dil gibi) alınmasıdır. İkinci durumda DMF'li maddeleri incelerken sadece istatistiksel olarak değerlendirmemelidir, incelenen durumun madde etkisinden kaynaklanıp kaynaklanmadığı incelenmelidir. Üçüncü bir noktada sadece bir grubun lehine DMF'li madde olduğu zamanda mı ölçme aracının inceleneceği konusu uzmanlar tarafından netleştirilmesidir. Dördüncü hususta DMF analizlerinin gerçekleştirilme zamanlamasına doğru karar verilmesidir (Analizlerin pilot uygulama aşamasında mı yoksa puanlar raporlanmadan önce mi yapılacağına kararı verilmelidir). Son olarak bir maddenin DMF gösterdiğine karar verdikten sonra maddenin ne zaman ve hangi koşullarda testten çıkarılacağına karar verilmesidir.

### ***Tek Biçimli (TB) ve Tek Biçimli Olmayan (TBO) DMF***

DMF'nin belirlenmesinde başlıca iki tane istatistiksel analiz yönteminden yararlanılır. Bunlardan biri tek biçimli DMF; Diğeri ise tek biçimli olmayan DMF'dir.

### Şekil 5

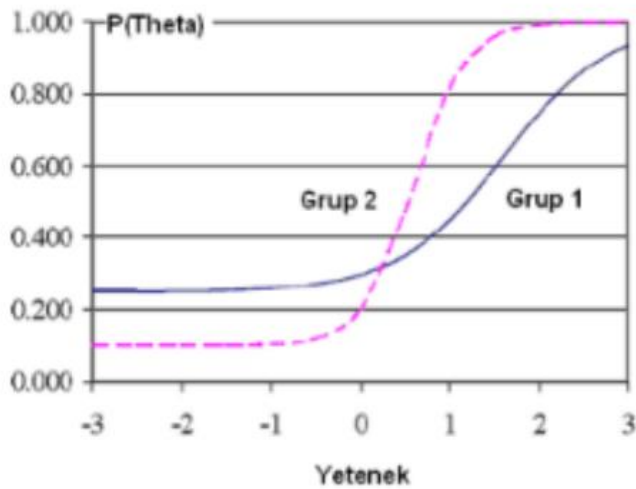
*Tek Biçimli DMF (Swaminathan & Rogers, 1990)*



Şekil 5'te önemli ölçüde tek biçimli DMF gösteren madde örneği yer almaktadır. Madde, tüm yetenek düzeylerinde odak ve referans grupları için yalnızca gruplardan birine avantaj sağlamasıdır. Diğer bir ifadeyle tüm yetenek düzeylerinde bulunan bireylerin, maddelere doğru yanıt verme olasılığının bir grup için daha yüksek olmasıdır (Swaminathan & Rogers, 1990).

### Şekil 6

*Tek biçimli olmayan DMF (Swaminathan & Rogers, 1990)*



Şekil 6, tek biçimli olmayan DMF'ye sahip bir maddedir. Maddeye doğru yanıt verme olasılığının tüm yetenek düzeyleri için farklı olmasıdır. Yani madde bir yetenek düzeyine kadar gruplardan biri daha avantajlıyken, diğer yetenek düzeyine geçildiğinde diğer grubun daha avantajlı olmasıdır (Swaminathan & Rogers,1990). Yani madde performansı ile yetenek düzeyi arasında bir etkileşimin olduğu, DMF' nin yetenek ölçeğinde yön değiştirdiği söylenebilir (French & Miller, 1996).

DMF analizleri, farklı grupların madde karakteristik eğrilerini karşılaştırarak değerlendirme yapılmasına imkân sağlar. Araştırmacılar aynı maddelere, değerlendirmek istediği her durum için ayrı madde karakteristik eğrisi çizebilir. Eğer farklı gruplarda madde karakteristik eğrileri birbirine çok yakın çıkıyorsa maddenin değişen fonksiyonu göstermediği, ancak madde karakteristik eğrileri birbirinden anlamlı derecede farklı çıktıysa DMF'nin olabileceği söylenebilir.

### ***DMF Belirleme Yöntemleri***

Farklı grupların, maddelere vermiş oldukları cevapların sorgulanarak yanlışlık olup olmadığı sorusuna cevap aramak amacıyla farklı birçok DMF yöntemi geliştirilmiştir. DMF'nin belirlenmesinde tek bir yöntemden elde edilen bulguların ele alınması yanlış çıkarımların yapılmasına sebebiyet verebilmektedir. Araştırmacılar bu yüzden DMF analizinin gerçekleştirilmesinde farklı yöntemlerden yararlanılmasının, bulguların yorumlanmasında kullanılması gerektiğini düşünmektedir (Ackerman,1992).

Puanlama şekillerine göre farklı DMF belirleme yöntemleri bulunmaktadır. Çok kategorili puanlan maddelerin DMF düzeylerinin belirlenmesi, iki kategorili maddelerde DMF belirlenmesine göre daha karmaşıktır. Ancak çok kategorili DMF'yi belirleme yöntemlerinin, iki kategorili maddelerde DMF belirleme yöntemlerinin uzantıları olduğu söylenebilir (Gür, 2019). Çok kategorili maddelerde DMF'nin belirlenmesinde gruplar ile madde puanları arasındaki etkileşime ek olarak madde puan düzeyinin de etkileşiminin olduğu görülmektedir, bu nedenle çok kategorili maddelerde DMF'nin belirlenmesinde kullanılacak olan yöntemler tüm bu olasılıkları değerlendirebilecek kapsamda ele alınmalıdır (French &



Miller,1996). İki kategorili puanlanan maddeler için SIBTEST, Mantel- Haenszel gibi yöntemler kullanılırken; çok kategorili maddeler için Poly-SIBTEST, OLR, Genelleştirilmiş Mantel- Haenszel gibi yöntemlerden yararlanıldığı görülmektedir. Bu çalışmada likert tipi maddeler kullanıldığından DMF belirleme yöntemlerinden Poly-SIBTEST ve OLR yöntemi kullanılacaktır. Bu yöntemler aşağıda açıklanmıştır.

### ***Çok kategorili verilerde DMF***

***Poly- SIBTEST.***SIBTEST, çok boyutlu madde tepki kuramı temel alınarak madde yanlılığını tespit edebilmek için geliştirilen (Shealy,1989; Shealy & Stout, 1991b) istatistiksel bir yöntemdir. Bu yöntem ile madde yanlılığını tespit edebilmenin yanı sıra DMF veya test yanlılığı da tespit edilebilir. SIBTEST 'in (Shealy,1989; Shealy & Stout, 1991b) geliştirilmesi çok boyutlu madde tepki kuramı, test yanlılığını modeline (Shealy,1989; Shealy & Stout, 1991a) dayanmaktadır. Bu yöntem, birçok DMF belirleme yönteminin aksine aynı anda birden fazla maddede DMF'yi tespit edilebilir (Clauser& Mazor,1998).

SIBTEST yöntemi parametrik olmayan ve madde tepki kuramına dayalı bir yöntemdir. Madde yanlılığını tespit eder, 1. tip hatadan fazla etkilenmemektedir. Bu yöntem ile ele alınan önemli bir diğer nokta ise çok boyutlu test yanlılığını modellemesi sebebiyle gerçek test yanlılığı ile sınav performansı sonucu oluşan ve etki olarak adlandırılan yanlı olmayan farklılıkları ayırt edebilmesidir.

Shealy ve Stout (1991) yanlılığın kavramsallaştırılmasında iki belirgin özellikten bahsetmektedir. Bunlardan ilki bireysel olarak yanlılığa sebep olabilecek birkaç ögenin bir araya gelerek test düzeyinde tutarlı ve büyük bir yanlılığa sebebiyet vermesidir. Aynı zamanda özellikle bir madde az miktarda yanlılık gösteriyor olsa bile bu doğru olabilir. Diğer bir özellikte çok boyutlu modellemeden bahsetmektedirler (Shealy & Stout,1991a). Bahsi geçen model ile psikolojik etkenlerin yanlılık oluşturabileceğini savunurlar. Hedef yetenek olarak belirlenen yetenek ile test performansını etkileyen diğer yetenekler arasında ayırım olduğunu vurgulanmaktadır. Böyle bir durumda farklı gruplarda farklı miktarlarda yanlılık

ortaya çıkabilir. Çok boyutluluk ile test bilgeliği, dil gibi bağlamsal yetenekler gerektiren durumlar yanlılığa sebep olabilir.

Çoklu puanlanan maddelerde, DMF'nin olmaması incelenen gruplarda maddelerden alınan puanın örtük değişken ile regresyonun aynı olmasını gerektirir. İkili puanlanan maddeler için kullanılan SIBTEST yönteminde nispeten değişiklikler yapılarak çoklu puanlanan maddelerdeki DMF 'de tespit edilebilir (Chang vd., 1996). Çoklu puanlanan maddelerde DMF'nin tanımlanmasında Eşitlik 6'da yer alan formül kullanılabilir;

$$B_0 = E_R[YI\theta] - E_F[YI\theta] \quad (7)$$

Formülde yer alan  $E_R$  referans grubu temsil ederken  $E_F$  odak grubu temsil etmektedir. Formülde eşitliğin sağlanması durumunda DMF'nin bulunmadığı yorumu yapılır.

SIBTEST analizleri sonucunda  $\beta$  katsayısı hesaplanır. Bu katsayının yorumlanabilmesi için Roussos ve Stout'un (1996) kategorize ettikleri DMF oranları ve değer aralığı tablosundan yararlanılabilir.

**Tablo 3**

*SIBTEST DMF Düzeylerinin Yorumlanması (Roussos & Stout (1996))*

DMF oranı	Değer Aralığı	Açıklama
A	$ \beta  < 0,059$	Yok veya ihmal edilebilir düzeyde
B	$0,059 \leq  \beta  < 0,088$	Orta düzeyde
C	$ \beta  \geq 0,088$	Yüksek düzeyde

SIBTEST yöntemi ile analizler gerçekleştirildikten sonra elde edilen  $\beta$  katsayısı Tablo 3'teki düzeylere göre yorumlanabilir. Elde edilen katsayının pozitif olması odak grubun aleyhine bir yanlılığı gösterebilirken; bu katsayının negatif çıkmış olması durumunda referans grubun aleyhine bir DMF'nin olduğunu göstermektedir (Abbott, 2007).

**Ordinal Lojistik Regresyon.** Lojistik regresyon yöntemi, TB ve TBO DMF'nin tespit edilmesinde kullanılan güçlü yöntemlerden biridir (Swaminathan & Rogers,1990). Ordinal lojistik regresyon (OLR) yöntemi hem iki kategorili maddelerde hem de likert tipi maddelerde DMF'nin olup olmadığının test edilmesinde kullanılan bir yöntemdir (Zumbo,1999).Tek biçimli veya tek biçimli olmayan DMF 'nin belirlenmesinde kullanılabilen yöntem bu açıdan da araştırmacılara avantaj sağlamaktadır. Çok kategorili puanlanan maddelerde DMF'nin belirlenmesi iki kategorili maddelerde DMF'nin belirlenmesine göre daha karmaşıktır (French ve Miller, 1996). OLR yöntemi ile DMF'nin test edilmesinde değişkenler hiyerarşik olarak modele dahil olurlar. OLR yöntemine eşitlik 8'de yer verilmiştir.

$$y = b_0 + b_1TOP + b_2 GRUP + b_3 TOP * GRUP + \varepsilon_i \quad (8)$$

Eşitlik 8 incelendiğinde ilk olarak toplam puan modele dahil olduğu, sonraki adımda grup değişkeni (cinsiyet gibi) modele dahil edildiği ve en son etkileşim terimi (Toplam puan x cinsiyet gibi) modele dahil edilmiş olduğu; yani söz konusu maddelerden alınan toplam puanın grup değişkenine göre değişim gösterip göstermediği incelenir (Zumbo,1999). Her bir adımda ki kare istatistiğinden yararlanılarak DMF test edilir. TB DMF 'nin belirlenmesinde ilk ve ikinci adımda R<sup>2</sup> arasındaki değişimler incelenebilir. TB DMF'nin tespit edilmesinde R<sup>2</sup> ile etki büyüklüğü de hesaplanır. Etki büyüklüğü ile aynı zamanda hangi grubun diğerine göre DMF gösterdiğinin belirlenmesinde kullanılabilir (Gelin & Zumbo, 2003).

Etki büyüklüğünün belirlenmesinde araştırmacılar farklı önermelerde bulunmuştur. Cohen'nin (1992) küçük, orta ve büyük etki büyüklüğü tanımlamaları da göz önüne alınarak Zumbo ve Thomas (1996),  $\Delta R^2$ 'nin kullanılmasıyla DMF düzeyinin belirlenmesini önermişlerdir. Crane vd. 2006, TBO DMF'nin belirlenmesinde Bonferroni tekniğinin kullanılmasını önermişlerdir. Bu teknik,  $\alpha$  katsayısının madde sayısına oranı ( $\alpha/n$ ) ile hesaplanmaktadır. Bu araştırmada iki farklı alt boyut olduğu için iki farklı değer hesaplanmıştır.

**Tablo 4** *$\Delta R2$  Değerlerine Göre DMF Düzeyi Ölçütleri*

DMF Düzeyi	Yorumlama Ölçütü	DMF Durumu
A düzey	$\Delta R2 < 0.13$	DMF yoktur veya ihmal edilebilir
B düzey	$0.13 \leq \Delta R2 < 0.26$	Orta düzeyde DMF
C düzey	$\Delta R2 \geq 0.26$	Yüksek düzeyde DMF

**Uluslararası Sınavlar ve TIMSS**

Geçmişten günümüze kadar ülkeler arası eğitim sistemlerinin değerlendirilmesi için farklı sınavlar kullanılmıştır. Matematik ve fen alanlarında 4. ve 8. Sınıf düzeyindeki öğrencilerin başarılarını uluslararası anlamda izlemek amacıyla IEA tarafından dörder yıllık döngülerle yapılan TIMSS' de bu sınavlardan biridir. TIMSS döngüleri, Boston Kolejindeki Uluslararası Çalışma merkezi tarafından yürütülmektedir. 1995'ten bugüne dünyanın farklı yerlerindeki ülkelerin matematik ve fen alanlarındaki performanslarını değerlendirerek, ülkelerin eğitim politikalarında düzenlemeler yapmasına imkân sağlamıştır (Mullis vd., 2020). Dördüncü sınıf öğrencilerinin performanslarının değerlendirilmesi ve eğitim sistemlerinde gerekli düzenlemeler yapılması sonrasında sekizinci sınıf döngüsünde bu düzenlemelerin sonuçlarının izlenmesi TIMSS sınavı ile sağlanabilir. Bu döngülerin ülkeler ölçeğinde uygulanması, sonuçların raporlanması, sonuçların ülkelerin kendi bağlamında değerlendirilmesi Ulusal Araştırma Koordinatörleri tarafından yapılmaktadır. Bu bağlamda TIMSS döngüleri her ülkenin matematik ve fen eğitimleri hakkında bilgi verir, bu alanda çalışan araştırmalar için de zengin bir veri kaynağı sunar (Mullis vd., 2020).

IEA ve OECD tarafından düzenlenen geniş ölçekli sınavlar aracılığı ile farklı dili konuşan, farklı kültüre sahip ülkelerdeki öğrencilerin çeşitli alanlardaki performansları, öğrencilerin temel yaşam becerileri ve ülkelerin eğitim sistemlerinin etkililiği incelenmeye çalışılmaktadır. Bu sınavların farklı alt gruplar için aynı şeyi ifade edip etmediği farklı analiz yöntemleriyle ve uzman görüşleriyle belirlenmeye çalışılmaktadır. Bu belirlemeler ve

değerlendirmeler sonucunda ülkeler arası farklılıklar, eğitim sistemleri ve müfredatlarda yapıcı yönde değişikliklerin oluşturulmasına katkı sağlayabilmektedir. Doğru geri bildirimler ve çıkarımlar aracılığıyla öğrencilerin ihtiyaçlarını da ön plana alan eğitim sistemleri oluşturulabilmektedir.

Türkiye ülkeler arası yapılan bu sınavlara katılım sağlamaktadır. Türkiye, 1999 yılında TIMSS; 2001 yılında PIRLS ve 2003 yılına gelindiğinde PISA sınavına katılmıştır. Uluslararası anlamda yapılan bu sınavlara farklı yaş gruplarındaki bireyler katılmaktadır. Aynı zamanda bu sınavların her biri farklı bir amaca hizmet ederek farklı alanlarda bireylerin performansını ölçmektedir. Bu sayede ülkeler arası bağlamda Türkiye'nin nerede yer aldığı, diğer ülkelerdeki bireylerin performansları ile kıyaslanarak belirlenebilmektedir. Bu sayede eğitim sistemlerindeki gelişmeler takip edilmiş olup güncellemeler yapılabilmektedir.

TIMSS in 2019 döngüsü ile eTIMSS olarak dijital ortamda da değerlendirmeler yürütülmeye başlanmıştır. TIMSS 2019'a katılım sağlayan ülkelerin yaklaşık yarısı 2019 döngüsünde eTIMSS sistemiyle sürece katılım sağlamıştır. Matematik ve fen alanlarında kapsamı genişletilmesine olanak sağlayan eTIMSS, yenilikçi problem çözme, sorgulama ve PSI olarak bilinen uygulamalara olanak sağlar. PSI gerçek dünyanın laboratuvar ortamında simüle edilmesine olanak tanıyarak öğrencilerin matematik ve fen alanlarında süreç becerilerin geliştirmesini destekler. Yani PSI öğrencilerden beklenen görevlerin daha çekici hale getirilmesine katkı sağlar. Ayrıca, PSI ile öğrencilerin problem çözme veya sorgulama yollarını dijital olarak izleme fırsatı edinilir. Öğrencilerin sorulara verdiği yanıtlar insan yerine bilgisayarın puanlaması sağlanarak değerlendirmede pratiklik sağlanmıştır (Mullis vd., 2020).

Dünyanın farklı yerlerinde eğitim sistemleri ve eğitim ortamlarında iyileştirmeler ve gelişmeler için öğretim teknolojilerine yatırımlar yapılmaktadır. TIMSS 2019 döngüsüyle öğrencilerin ev veya okul ortamlarında teknoloji kullanımları hakkında bilgi edinilebilir. Kısacası öğrencilerin bir kısmı dijital değerlendirmelerde daha deneyimli olacakken bir kısmı ise bu konuda daha az deneyimli olacaktır.

TIMSS, IEA tarafından dört yıllık periyotlar halinde gerçekleştirilen, ulusal çapta dördüncü ve sekizinci sınıfta eğitimine devam eden öğrencilerin Matematik ve Fen başarılarının değerlendirilmesini amaçlar (MEB, 2020). TIMSS araştırması ile öğrencilerin bilişsel anlamda performanslarının yanında öğrenci, öğretmen, veli, öğretim programları, eğitim sistemleri hakkında bilgiler edinilir. Bu sayede diğer ülkelerin başarılarını artırma konusunda bilgi edinilerek eğitim planlamaları yapılabilmektedir.

## **İlgili Araştırmalar**

### **Ölçme Değişmezliği ile İlgili Yapılan Ulusal Araştırmalar**

Kıbrıslıoğlu (2015) yüksek lisans tezinde PISA 2012 verisinde matematik öğrenme alt boyutu anketleri ile oluşturulan matematik öğrenme modelin ülkeler ve cinsiyetler arasında değişmezliğini incelemiştir. Araştırmanın örneklemini Türkiye, Çin-Şangay ve Endonezya'dan oluşmaktadır. Başlangıçta açımlayıcı faktör analizi kullanılarak 55 madde ve 9 faktörden oluşan öğrenme algısı doğrulanmış devamında ise modelin ülke ve cinsiyet grupları açısından ölçme değişmezliğinin sağlanıp sağlanmadığı test edilmiştir. Ölçme değişmezliği analizi çok gruplu doğrulayıcı faktör analizi yöntemi kullanılarak gerçekleştirilmiş olup model aşamalar aracılığıyla ele alınmıştır. Araştırmada referans grup olarak cinsiyet için kızlar referans alınırken; Ülkeler arası karşılaştırmalar için Türkiye referans alınmıştır. Araştırma sonucunda matematik öğrenme algısı modelinin ülkeler arasında sadece şekil değişmezliği sağladığı görülmüştür. Cinsiyet karşılaştırmaları için kurulan modellerde ise ölçme değişmezliğinin her aşamada sağlanmış olduğu gözlenmiştir.

Ölçüoğlu ve Çetin (2016), TIMSS 2011 verisinde sekizinci sınıf öğrencilerinin matematik başarısını etkileyen değişkenleri bölgelere göre incelemiştir. Öncelikle çalışmada AFA yapılmış olup maddelerin hangi boyutta toplandığı incelenmiştir. Yapılan açımlayıcı faktör analizinden sonra modele alınan değişkenler ile Matematik başarısının yordanması incelenmiştir. İlişkinin incelenmesi için Yapısal Eşitlik Modeli kurulmuştur. Modelin tüm bölgeler için geçerli olduğuna karar verdikten sonra bölgeler arası

karşılaştırmalar yapabilmek için ÇGDFA ile ölçme değişmezliği test edilmiştir. Analiz sonuçlarına göre alt gruplarda yapısal ve metrik değişmezliğin olduğu bulunmuş olup tam eşdeğerliğin sağlanmadığı sonucuna ulaşılmıştır.

Polat (2019) yüksek lisans tezinde TIMSS 2015 8. Sınıf verilerinde matematik ve fen duyuşsal özellik modelinin cinsiyet ve bölge değişkenleri açısından ölçme değişmezliğini incelemiştir. Araştırmanın örneklemini Türkiye, Singapur ve Suudi Arabistan ülkeleri oluşturmaktadır. AFA ve DFA ile model doğrulandıktan sonra ölçme değişmezliği ÇGDFA ile incelenmiştir. Analiz sonuçları ışığında göre kurulan modelin kültür ile bölge değişkenleri açısından ölçek değişmezliğini, cinsiyete göre ise katı değişmezliğinin sağlandığı görülmüştür. Dile göre yapılan ölçme değişmezliği sonuçlarına göre kısmi homojen model kabul edilmiştir. Madde düzeyinde dil ve cinsiyete göre yapılan analizlerde matematik ve feni sevme boyutunda ölçme değişmezliğinin sağlandığı görülmüştür.

Samast (2024) yüksek lisans tezinde PISA 2018 verisinde Bilgi ve İletişim Teknolojisi ölçeğinin özyeterlik ve sosyalleşme alt ölçeklerinin ölçme değişmezliği incelenmiştir. Araştırmanın örneklemini Türkiye, Almanya ve Amerika Birleşik Devletleri oluşturmaktadır. Türkiye, Almanya ve ABD veri setlerindeki öğrencilerin fen testi başarısını ölçen makul değerlerden ilkinde (PV1SCIE) göre alt ve üst gruplar (%27'lik alt ve üst gruplar) oluşturulmuştur. Araştırmanın sonuçlarına göre BİT ölçeği düşük ve yüksek fen başarısı gösteren öğrencilere göre benzer faktör yapısına sahip olduğu görülmüştür.

### **Ölçme Değişmezliği ile İlgili Yapılan Uluslararası Araştırmalar**

Marsh ve arkadaşlarının (2006) yapmış olduğu bir çalışmada PISA 2000 döngüsündeki veriler kullanılarak kültürel açıdan ölçme değişmezliği incelenmiştir. Araştırmanın örneklemini PISA 2000 döngüsüne katılmış olan 25 ülkeden 4000 kişiden oluşmaktadır. Çalışmada ölçme değişmezliğini test etmek için ÇGDFA yöntemi kullanılmıştır. OECD çalışmalarında ülkeler arası karşılaştırmalar yapılırken öğrencilerin başarılarının yanında maddelerin çevrilen kültürle ne kadar uyum gösterdiği, farklı

ülkelerdeki öğretim programı ile ne kadar uyumlu olduğu konularının önemine değinilmiştir. Araştırmanın sonucunda uyum iyiliği indeksleri incelenmiş olup, ele alınan ölçeğin kültürlerarası açıdan ölçme değişmezliğinin sağlanmış olduğu görülmüştür.

Asil ve Brown (2016) çalışmalarında PISA okuduğunu anlama testini ulus, kültür ve dil değişkenine göre ölçme değişmezliğini incelemiştir. Araştırmanın örneklemi PISA 2009 'a girmiş olan Yeni Zelanda, Kanada ve ABD' den oluşmaktadır. Araştırmanın odak noktasını PISA Okuduğunu Anlama testinin İngilizce konuşan referans olarak alınan grubun performansı dil, kültür ve ekonomik gelişmişliğe göre değişmezliğin sağlanıp sağlanmaması oluşturmuştur. Ölçme değişmezliği ÇGDFA ile test edilmiştir. Analizler sonucunda ölçek değişmezlik aşamasına kadar ölçme değişmezliğinin sağlandığı görülmüştür.

He vd., (2019) yapmış oldukları çalışmada TIMSS ve PISA uygulamalarında bilişsel olmayan yapıların kültürler arası karşılaştırmışlardır. Çalışmaya PISA 2015 ve TIMSS 2015 döngülerine katılan 29 kültür dahil edilmiştir. 29 kültürde bulunan öğrencilerin araçsal motivasyonu maddeleri, 21 kültürde yer alan bilimden zevk alma ile ilgili maddeler ve 28 kültürde de ortak yer alan aidiyet duygusu ile ilgili maddeler kullanılmıştır. Ölçme değişmezliği, çoklu grup doğrulayıcı faktör analizi yöntemiyle gerçekleştirilmiştir. Araştırmanın sonucunda araçsal motivasyon, bilimden zevk alma ve aidiyet duygusu ile başarı arasında pozitif yönde ilişki olduğu sonucuna ulaşılmıştır.

Liou ve Lin (2021) yapmış oldukları çalışmada TIMSS 2019 verilerini kullanarak fen'e karşı motivasyonel inançları ülke ve cinsiyet değişkenlerine göre ölçme değişmezliğini incelemiştir. Araştırmanın örneklemini Tayvan, Avustralya ve Amerika Birleşik Devletleri oluşturmaktadır. Ölçme değişmezliğinin araştırılmasında çoklu grup doğrulayıcı faktör analizi kullanılmışlardır. Araştırmanın modeli üç boyutlu 25 maddeden oluşmaktadır. Araştırmanın sonuçlarına göre ülkelere göre ölçme değişmezliği üç ülkede metrik değişmezliğin sağlanmış olduğu görülmüştür. Cinsiyet değişkenine göre yapılan karşılaştırmalarda ülkelerin ölçek değişmezlik aşamasına kadar ölçme değişmezliğinin



sağlandığı görülmüştür. Araştırmanın sonuçları fen'e karşı motivasyonel inançların cinsiyet değişkeni açısından karşılaştırılabilir olduğunu göstermiştir.

Senden vd., (2023) çalışmalarında TIMSS 2019 verisinde öğretmenlerin davranış yönetimi ile ilgili altı madde ve öğretimsel netlik ile ilgili yedi madde ile öğrencilerin algıları, ülkelere göre ölçme değişmezliği incelenmişlerdir. Araştırmanın modeli 2 faktörlü 13 madde ile kurulmuştur. Araştırmanın örneklemini TIMSS 2019 döngüsünde katılmış olan 38 ülke oluşturmuştur. Bulgular ışığında yapılan karşılaştırmalarda metrik değişmezliğin sağlandığı görülmüştür. Araştırmanın sonucunda ülkeler arasında sınırlı karşılaştırmalar yapılabileceği görülmüştür.

### **Yanılılık ile İlgili Yapılan Ulusal Araştırmalar**

Atalay (2010) yüksek lisans tezinde PISA 2006 öğrenci anketindeki fen bilimlerine karşı tutum maddelerini cinsiyet ve kültür değişkenlerine göre DMF gösterip göstermediğini incelemiştir. Araştırmanın örneklemini bu döngüye katılmış olan öğrencilerden ABD ve Türkiye ülkeleri oluşturmaktadır. Analizler sonucunda Türkiye verisinde cinsiyete göre yapılan analizlerde OLR yöntemi ile 12 maddede, poly-SIBTEST yöntemi ile 13 maddede DMF gözlenmiştir. ABD verisinde cinsiyete göre yapılan analizlerde ise hem OLR hem de poly-SIBTEST yöntemi sonucunda 15 maddede DMF gözlenmiştir. Her iki yöntemde ortak ve önemli düzeyde DMF gösteren madde sayısı Türkiye'de 7, ABD' de 13'tür. İki ülke arasında kültüre göre her iki yöntemde yapılan analizler sonucunda 30 maddede önemli düzeyde DMF görülmüştür. Kültüre göre yapılan analizler sonuçları incelendiğinde her iki yöntemde de DMF gösteren 31 madde olduğu sonucuna ulaşılmıştır. Ayrıca her iki yöntemde de DMF'li alan maddelerin 21 tanesinde aynı düzeyde DMF olduğu görülmüştür.

Asil ve Gelbal (2012), PISA Öğrenci Anketinin Kültürler Arası Eş Değerliğinde Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 öğrenci anketinin kültürler ve diller arası eşdeğerliğini incelemişlerdir. Araştırmanın örneklemini Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye oluşturmaktadır. Araştırmanın ilk

aşamasında Doğrulayıcı Faktör Analizi yapılmış sonrasında ÇGDFA'dan yararlanılarak ölçme değişmezliği test edilmiştir. ÇGDFA sonucunda yanlılık gösteren maddeler DMF analizleri ile araştırılmıştır. Analiz sonuçlarına göre dil ve kültür açısından DMF gösteren maddelerin olduğu tespit edilmiştir. Bu maddeler ülkelerarasında çeviri kaynaklı, kültürel açıdan kaynaklanan farklılıklardan olduğu ortaya koymuştur. Avustralya- Türkiye karşılaştırılmasında uzman görüşlerine başvurulmuş, uzmanlar DMF'li olan maddelerin tamamında çeviri veya uyarlamadan kaynaklı farklılıkların olabileceği ile ilgili görüşlerini bildirmişlerdir.

Gök vd., (2014) makalelerinde PISA 2009 verileri ile çalışmıştır. Araştırmalarında 11 maddeden oluşan tutum ölçeği kullanmışlardır. Araştırmanın örneklemini Türkiye, Rusya, Yeni Zelanda ve Avustralya oluşturmaktadır. Bu ülkelerin seçilmesinde ülkelerin başarı sıralaması, dil ve kültürel özellikler önemli olmuştur. Başlangıçta AFA ile faktör yapısı test edilmiş sonrasında DMF analizleri için poly-SIBTEST, OLR ve madde tepki kuramı olabirlik yöntemi kullanılmıştır. Farklı dil ve farklı kültüre sahip ülkelerde DMF'li madde sayısının 10, farklı kültür ve aynı dilde DMF'li madde sayısının 9, aynı kültür ve aynı dilde DMF 'li madde sayısının ise 4 olduğu sonucuna ulaşılmıştır. Bu araştırmanın sonuçlarına göre kültür farklılıklar nedeniyle DMF'li madde sayısının daha fazla olduğu söylenebilir.

Uzun ve Gelbal (2017), PISA 2006 verisinde Fen Başarı Testinin Madde Yanlılığının Kültür ve Dil Açısından incelenmişlerdir. Örneklem olarak Türkiye, Avustralya, İngiltere ve Kanada seçilmiş, bu ülkelerin başarı testine ilişkin veriler kullanılarak DMF analizleri gerçekleştirilmiştir. Çalışmanın temel amacında klasik test kuramı ve madde tepki kuramı açısından DMF belirlemede kullanılan yöntemlerin benzer ve farklı yönleri ortaya konmaya çalışılmıştır. Bu çalışmanın diğer amacında ise DMF gösteren maddelerin sebeplerinin ortaya konmasıdır. Bu çalışmada KTK tekniklerinden biri olan Mantel- Haenszel tekniği ve Lojistik Regresyon tekniği; MTK'ya göre madde karakteristik eğrileri incelenerek alan indeksleri tekniğine başvurulmuş ve DMF incelenmeye çalışılmıştır. Yapılan incelemeler sonucunda KTK'ya dayalı tekniklerin kendi aralarında uyumlu olduğu; aynı şekilde MTK'ya

dayalı tekniklerinde kendi aralarında uyumlu sonuçlar verdiği görülmüştür. Yanlı olduğu düşünülen maddeler 15 uzman tarafından incelenmiştir. Uzmanlardan alınan görüşler ışığında maddelerin yanlılık içermesi çeviri, müfredatlarda farklılıkların olması, uyarlamada hatalar, metninin çeviri sırasında ağır bir dile dönüşmesi, sosyoekonomik düzeylerin gruplar arasında farklı olması gibi nedenlerden kaynaklanabileceği düşünülmüştür. Alt gruplar arasında farklılıkların artması ile DMF gösteren maddelerin de artmış olduğu görülmüştür.

Gür (2019), yüksek lisans tezinde PISA 2015 verisini kullanmış olup, fen bilimlerine karşı tutum maddelerinin dil ve kültüre göre DMF gösterip göstermediğini incelemiştir. DMF analizlerinde genelleştirilmiş Mantel Haenszel, poly-SIBTEST ve OLR yöntemlerini kullanmıştır. Araştırmanın örneklemini Türkiye, İrlanda, Amerika Birleşik Devletleri ve İngiltere'deki öğrenciler oluşturmaktadır. Bu çalışma için dokuz tutum maddesi seçilmiş ve analizler gerçekleştirilmiş, İngiltere ile ABD'nin karşılaştırmasında GMH ve OLR yöntemlerine göre yedi maddede ve poly-SIBTEST yöntemine göre altı maddede DMF olduğu görülmüştür. Yine yapılan analizler sonucunda İngiltere ile Türkiye karşılaştırılmış olup GMH yöntemiyle sekiz maddede, OLR yöntemiyle dokuz maddede ve poly-SIBTEST yöntemiyle yedi maddede DMF olduğu görülmüştür. Uygulanan her üç yöntemin analiz sonuçlarına göre İngiltere-İrlanda ile İngiltere-ABD karşılaştırmalarında beş maddede, İngiltere-Türkiye karşılaştırmasında ise yedi maddede DMF içeren tutum maddeleri olduğu sonucuna ulaşılmıştır. Araştırmada DMF belirlemek için kullanılan yöntemler arasında, DMF düzeyinin belirlenmesinde farklılıklar olduğu gözlenmiştir. Aynı zamanda OLR ve poly-SIBTEST yöntemi ile DMF analizi gerçekleştirildiğinde, DMF içeren madde türleri ve DMF gösteren maddelerin sağlamış olduğu avantajların gruplar arasında uyumlu olduğu görülmüştür.

Tiryaki (2019) yüksek lisans tezinde PISA 2015 uygulamasında öğrenci anketlerini DMF ve ölçme değişmezliği açısından incelemiştir. Araştırmada, Fen bilimlerine yönelik tutum anketlerinde 37 madde, 6 alt test kullanılmıştır. Ölçme değişmezliği ÇGDFA yöntemi ile incelenmiş; DMF ise madde tepki kuramı olabilirlik oranı yöntemi aracılığıyla ile

incelenmiştir. Araştırmanın örneklemini Türkiye ve ABD’de ülkeleri oluşturmaktadır. Ölçme değişmezliği analizleri sonucu incelendiğinde Fen öğrenmekten zevk alma, Fen öğreniminde araçsal güdülenme alt testlerinde değişmezlik aşamaları sağlanmaktadır. Geniş fen konularına ilgi, Fen’e yönelik öz yeterlik, Fen bilimlerine yönelik epistemik inançlar ve Bilim aktiviteleri alt testlerinde sadece metrik değişmezlik aşamasının sağlandığı görülmüştür. DMF açısından analizler cinsiyet bağlamında incelendiğinde Türkiye örnekleminde 30 madde, ABD örnekleminde 20 maddede ihmal edilmeyecek düzeyde DMF tespit edilmiştir. Kültür bağlamında yapılan DMF analizlerine ABD ve Türkiye arasında 34 maddede ihmal edilmeyecek düzeyde DMF tespit edilmiştir. Araştırmada ölçme değişmezliği ve DMF sonuçlarının birbirleriyle tutarlı olduğu görülmüştür.

Gören vd., (2024) makalelerinde PISA 2018’de okuduğunu anlama ve hatırlama stratejileri anketinde madde yanlılığını incelemek önce ölçme değişmezliği sonrasında DMF analizleri gerçekleştirmişlerdir. Araştırmada ABD, Avustralya, Kore, Japonya, Türkiye ve Yeni Zelanda örneklemleri kullanılmıştır. Yapılan ikili karşılaştırmalar sonucunda karşılaştırmaların ölçek değişmezliği aşamasına kadar sağladığı görülmüş bunun yanında Avustralya- ABD, Avustralya -Yeni Zelanda ülkeleri karşılaştırmalarında katı değişmezliğin de sağlandığı görülmüştür. DMF analizleri Poly-SIBTEST ile incelenmiş, Avustralya -Yeni Zelanda karşılaştırmalarında üç madde C düzeyinde yüksek düzeyde yanlı; Avustralya – ABD karşılaştırmalarında iki madde B yani orta düzeyde, iki madde de C seviyesinde yanlı yani yüksek düzeyde DMF’li ; Japonya -Kore karşılaştırmalarında beş madde C, bir madde A seviyesinde ; son olarak Türkiye- ABD karşılaştırmalarında ise beş madde C , bir madde de B seviyesinde yanlılık olduğu görülmüştür. Türkiye- ABD ikili karşılaştırmaları sonucunda 1,2 ve 3. maddeleri ABD’ye avantaj sağlarken; 4,5 ve 6. maddelerinde Türkiye örnekleminin avantajlı olduğu görülmüştür. Bu maddelerin yanlı olup olmadığına karar verilmesinde uzman görüşlerine başvurulmuştur. Türkçe Eğitimi uzmanlarından %70’inin dört maddenin yanlı olduğu görüşünde olduğu görülmüştür. Ölçme ve Değerlendirme uzmanlarının görüşlerine göre ise yanlılık olmadığı görülmüştür.

## Yanlılık ile İlgili Yapılan Uluslararası Araştırmalar

Huang (2010) doktora tezinde PISA 2006 uygulamasında dil, öğretim programı ve kültürel anlamda DMF'yi incelemiştir. Araştırmanın örneklemini PISA 2006 döngüne katılmış olan ülkelerden ABD, Kanada, Hong Kong ve Çin oluşturmaktadır. DMF analizlerinde lojistik regresyon kullanılmıştır. Ayrıca hem tek boyutlu hem de çok boyutlu çok terimli logit model katsayıları (RCML) DMF'nin belirlenmesinde kullanılmıştır. Ayrıca bu araştırmada içerik analizi aracılığı ile DMF kaynakları açıklanmaya çalışılmıştır. Araştırmanın sonucunda DMF'li maddelerin en az Kanada ve ABD arasında olduğu; en fazla ise ABD ile Çin arasında olduğu görülmüştür. ABD ve Çin arasında öğretim programı açısından da fark olduğu ortaya konmuştur. DMF'nin sadece küçük bir kısmının dilsel farklılıklarından kaynaklandığı sonucuna ulaşılmıştır.

Liu, 2019'daki doktora tezinde İngilizce dilini öğrenmeye devam eden veya etmeyen bireylerin matematik başarılarında açıklanamayan potansiyel kaynakları açıklamaya çalışmıştır. Araştırmada olası DMF kaynakları üç yöntem ile incelenmeye çalışılmıştır. Bunlar; Mantel-Haenszel, Rasch Analizi ve hiyerarşik genelleştirilmiş doğrusal modeldir. Yapılan araştırmalar sonucunda 10 maddede DMF olabileceği görülmüştür. Çalışmada PISA 2012 döngüsünde ABD örneklemini kullanılmıştır. Araştırmada, PISA 2012 döngüsünde ABD'den katılmış olan öğrencilerin ikili kodlanmış olan 76 maddeye vermiş olduğu cevaplar kullanılmıştır. Araştırma sonucunda İngilizce öğrenen veya öğrenemeyen bireyler arasında DMF olduğu belirlenmiştir. 14 madde için İngilizce öğrenemeyen bireyler, İngilizce öğrenen bireylere göre maddelerde zorluk yaşadıkları, bu durumun DMF'ye sebep olduğu sonucuna ulaşılmıştır. Okul türü ve okulun eğitim kaynaklarının DMF'ye sebep olabileceği çıkarımı yapılmıştır.

Sandilands vd., (2013), yaptıkları çalışmada büyük ölçekli sınavlarda farklı yaklaşımlarla DMF'nin olası kaynaklarını incelemiştir. Araştırmada PIRLS 2001 yılında katılan ülkelerden ABD, İngiltere ve Kolombiya'nın verilerini kullanarak bilişsel yükleri test etmişlerdir. Çalışmada İngiltere- Kolombiya, İngiltere-ABD, ABD- Kolombiya ülkelerinin

karşılaştırmasında DMF analizleri yapılmıştır. Aynı zamanda teori ve geçmiş zamanda DMF ile ilgili araştırmalar dikkate alınarak varsayılan maddeler gruplandırılarak karşılaştırmalar yapılmıştır. Yapılan karşılaştırmalarda ABD- Kolombiya karşılaştırmasında 16 maddede; İngiltere- Kolombiya karşılaştırmasında 17 maddede; İngiltere- ABD karşılaştırmasında, 10 maddede DMF olduğu görülmüştür.

Oliveri vd., (2012) makalelerinde PISA 2003 verisinde problem çözme ölçeğinin Kanada'daki İngilizce ve Fransızca versiyonlarına göre 10 madde ile karşılaştırmalar yapmışlardır. Yapılan karşılaştırmalarda üç farklı DMF yöntemi kullanılmıştır. Madde düzeyindeki DMF analiz bulguları benzerlik göstermiştir ancak etki büyüklükleri farklılıklar gösterdiği sonucuna ulaşılmıştır. DMF'li maddelerin bazıları İngilizce dili lehine sonuçlanırken bazılarında Fransızca lehine sonuçlar gösterdiği görülmüştür. DMF belirlemede üç yöntem (NIRT, L-H ve OLR) kullanılmıştır. Analizler sonucunda iki madde Fransızca lehine tek biçimli DMF, bir madde İngilizce lehine tek biçimli olmayan DMF göstermiştir.

Bialo ve Li (2022) yaptıkları çalışmada PISA 2015 verisinde akademik motivasyon ile ilgili maddeleri cinsiyet ve etnik köken değişkenine göre DMF açısından incelemişlerdir. DMF analizleri için OLR yöntemi kullanılmıştır. Araştırmanın örneklemini PISA 2015 döngüsüne katılan ABD.'deki öğrenciler oluşturmaktadır. Yapılan analizler sonucunda cinsiyet değişkenine göre üç maddenin DMF olduğu görülmüştür. Aynı zamanda etnik kökene göre de Afrikan Amerikalı öğrencilerin Asyalı öğrencilere göre yapılan ikili karşılaştırmalarda iki maddenin DMF 'li olduğu görülmüştür.

### **İlgili Araştırmaların Genel Özeti**

Araştırmalar incelendiğinde, uluslararası sınavlarda dilsel ve kültürel farklılıklardan dolayı yanlış maddelerin ortaya çıktığı görülmüştür. Yapılan araştırmalar incelendiğinde uluslararası ölçekte gerçekleştirilen sınavlardaki bilişsel testlerde ülkelere göre karşılaştırmalar yapılmış (Gönen,2021; Bağdu Söyler,2020) ancak ölçme değişmezliğinin

sağlanmadığı sonucuna ulaşmışlardır. Duyuşsal özellikler ile ilgili kurulan modellerde kültüre göre ölçme değişmezliğinin sağlanmadığı görülmüştür (Tekin, 2019; Polat, 2019). Araştırmacılar dil ve kültürel açıdan ölçme değişmezliğini test etmenin yanında cinsiyet değişkenine göre de ölçme değişmezliğini incelemişleridir (Kıbrıslıoğlu, 2015; Polat 2019). Araştırmacılar cinsiyete göre yapmış oldukları ölçme değişmezliği sonuçlarına göre cinsiyete göre ölçme değişmezliğinin sağlandığı sonucuna ulaşmışlardır. Ölçme değişmezliği çalışmalarında genel olarak ÇGDFA yöntemi kullanıldığı görülmüştür. Araştırmacılar en çok uluslararası sınavlardan biri olan PISA verisi ile çalışmışlardır.

Yanlılık çalışmalarında DMF yöntemlerine sıklıkla başvurulmuş aynı zamanda uzman görüşleri alınmıştır. Ancak araştırmacıların yapmış oldukları çalışmalara göre ölçme değişmezliği çalışmaları bulgularında olduğu gibi dil ve kültüre göre yanlılık olduğu analizler ile ortaya konmuştur (Olveri vd., 2012; Sandilas vd., 2013; Gök vd.,2014; Gür 2019).

### Bölüm 3

#### Yöntem

Araştırmanın bu bölümünde araştırmanın türü, araştırmanın evreni ve örnekleme, veri toplama süreci, verilerin analizi ve araştırmanın sayıltıları incelenecektir.

#### Araştırmanın Türü

Araştırmada TIMSS 2019 verisinde bilgisayar kullanımına ilişkin maddelerin ölçme değişmezliği ve DMF incelenmiştir. Yapılan araştırmanın türü betimsel araştırmadır. Betimsel araştırma ile evren hakkında örneklemlerden edinilen bilgilerle çıkarımlar yapılabilir (Fraenkel & Wallen, 2006).

#### Araştırmanın Evreni ve Örnekleme

Bu araştırmanın evreni TIMSS 2019'a katılmış olan öğrencilerden oluşmaktadır. Araştırmanın örnekleme ise TIMSS 2019'a katılmış olan Türkiye, İngiltere ve Katar'dan oluşmaktadır. Araştırmada bu ülkelerin seçilmesinin amacı, farklı dil farklı kültür (TÜR-İNG), farklı dil aynı kültür (ARP-İNG) karşılaştırmalarını yapabilmektir.

#### Tablo 5

##### *Araştırma Örnekleminin Frekans ve Yüzdeleri*

Ülkeler	Frekans	Yüzde
Türkiye	4077	36.44
İngiltere	3365	30.07
Katar (Testi Arapça alan)	1891	16.90
Katar (Testi İngilizce alan)	1854	16.57
Toplam	11187	100,0

Tablo 5 'te ülkeler bazında örneklem sayılarına yer verilmiştir. Katar örnekleme aynı kültür farklı dil açısından inceleneceğinden örneklem sayıları da ayrı ayrı ele alınmıştır.



## Veri Toplama Süreci

Türkiye’de TIMSS araştırması Millî Eğitim Bakanlığı (MEB) bünyesindeki Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü koordinasyonluğunca yürütülmektedir. TIMSS gibi uluslararası öğrenci başarılarını belirli periyotlarla takip eden araştırmalar, yapılan uygulamaların daha pratik hale getirilmesinde yenilikçi çözümlerin geliştirilmesini amaçlamaktadır (MEB, 2020). Bu amaçla TIMSS sınavı 2019 uygulamasında bilgisayar tabanlı değerlendirmeye geçiş yapılmıştır. TIMSS 2019 uygulamasına katılmış ülkelerden yarısı bilgisayar tabanlı değerlendirmeye geçmiş ancak uygulamaya katılan ülkelerin diğer yarısı kâğıt-kalem tabanlı değerlendirmeler yapmaya devam etmiştir (Mullis vd., 2020).

TIMSS döngülerinde başarı puanları ile ilgili bilgi edinmenin yanı sıra uygulamaya katılmış olan öğrenciler, öğretmenler, veliler ve okul idarecilerinden de veri toplanmaktadır. Kısacası öğrencilerin başarısını etkileyebilecek değişkenlere ilişkin bilgiler toplanarak uluslararası karşılaştırmalar yapılabilmektedir. Bu araştırmada öğrenci anketlerinde yer alan e TIMSS soruları kullanılmıştır.

## Veri Toplama Araçları

Bu çalışmada Türkiye, İngiltere ve Katar ülkelerinden 8. sınıf TIMSS 2019 döngüsüne katılmış olan öğrencilerin bilgisayar kullanımına ilişkin maddelere vermiş oldukları cevaplar kullanılmıştır. Araştırmanın verilerine TIMSS’in resmî sitesi <https://www.iea.nl/data-tools> üzerinden ulaşılmıştır. Kısacası bu araştırmanın veri toplama tekniği kayıttan elde edilmiş olan verilerdir. Araştırmada kullanılan maddelere Tablo 6 ‘da yer verilmiştir.

**Tablo 6***Veri Toplama Araçları*

MADDELER	
BSBE02A	Okulda rapor veya sunum gibi konuların üzerinde çalışırken ne sıklıkla tablet veya bilgisayarı kullanırsınız?
BSBE02B	Matematik ödevlerinde ne sıklıkla tablet veya bilgisayarı kullanırsınız?
BSBE02C	Fen bilimleri ödevlerinde ne sıklıkla tablet veya bilgisayarı kullanırsınız?
BSBE02D	Teste veya sınava girerken ne sıklıkla tablet veya bilgisayarı kullanırsınız?
BSBE04A	WiFi teriminin anlamını ne kadar iyi bilirsiniz?
BSBE04C	Anlık mesajlaşma teriminin anlamını ne kadar iyi bilirsiniz?
BSBE04D	Kopyala ve yapıştır teriminin anlamını ne kadar iyi bilirsiniz?
BSBE04E	E tablolar terimini ne kadar iyi bilirsiniz?
BSBE04F	Simge terimini ne kadar iyi bilirsiniz?
BSBE04G	Sürükle ve bırak terimlerini ne kadar iyi bilirsiniz?
BSBE04H	Kaydırma terimini ne kadar iyi bilirsiniz?

Bu araştırmada dörtlü likert tipi 11 madde ile model kurulmuştur. Bilgisayar kullanımına ilişkin "BSBE03" kodu ile başlayan maddelerin ülkeler arası AFA sonuç karşılaştırmalarında tek boyutluluğun sağlanmaması nedeniyle bu maddeler analizlere dahil edilmemiştir. Bilgisayar kullanımına ilişkin "BSBE01" kodu ile başlayan maddelerin ise 1-0 puanlan maddeler olduğu görülmüş ve bu maddeler analizlere dahil edilmemiştir. Kullanılan maddeler anket maddeleri olduğundan aynı kodlu olan maddelerin tek bir boyutta ele alınıp alınamayacağını incelemek için AFA ve DFA yapılmıştır. Aynı zamanda araştırmada kullanılacak olan ülke verilerinde güvenilirlik için alfa güvenilirlik katsayısı hesaplanmıştır.

**Verilerin Analizi**

Analizlere başlamadan önce bazı varsayımlar incelenmiştir. Bu varsayımlar kayıp değerlerin incelenmesi, uç değerlerin incelenmesi, normallik-doğrusallık, eş varyanslılık ve

çoklu bağlantılılığın kontrol edilmesidir (Çokluk, Şekercioğlu ve Büyüköztürk, 2012). Herhangi bir değişkeninde kayıp veri olması ve bu durumun örneklemin %5'inden az olması durumunda verilerde liste bazlı silme yöntemi önerilmektedir (Tabachnick ve Fidell, 2013). Ancak bu araştırmanın verileri incelendiğinde kayıp veri oranının %5'ten fazla olduğu görülmüştür. Veri kaybını önlemek için kayıp veri ataması "Expectation-Maximization algoritması (EM)" yöntemine göre kayıp veri atanmıştır.

Tek değişkenli uç değerler için her bir maddeye verilen yanıtlar z standart puanına çevrilmiş ve  $\pm 3$  değerleri dışındaki veriler çıkarılmıştır. Daha sonra çok değişkenli uç değerleri belirlemek amacıyla Mahalonobis uzaklığı hesaplanmış ve 0.01'den küçük değerler veri setinden çıkarılmıştır. Analize Türkiye örnekleminde 875, İngiltere örnekleminde 351 kişi, Katar örnekleminde 141 veri silinmiştir.

Verilerin analiz edilmesinde öncelikle veri rastgele 500 olacak şekilde bölünmüş ve açımlayıcı faktör analizi yapılmıştır. Açımlayıcı ve doğrulayıcı faktör analizi için R studio uygulaması kullanılmıştır. Ölçme değişmezliği ve DMF analizleri için SPSS 22.0 ve R studio ve SIBTEST yazılımları kullanılmıştır.

## Tablo 7

### *Ülkelere göre Cronbach Alfa Güvenirlik Katsayıları*

Ülkeler	İngiltere	Türkiye	Katar
Bilgisayar Kullanım Sıklığı Boyutu	0.84	0.66	0.78
Bilgisayar Kullanım Öz -yeterlik Boyutu	0.78	0.78	0.76

Tablo 7'de eTIMSS sorularının Alfa güvenirlik katsayısı sonuçlarına yer verilmiştir. Ülkelerin güvenirlik katsayılarının 0.66 ile 0.84 arasında değiştiği görülmektedir. Araştırmanın AFA sonuçları incelendiğinde İngiltere'nin için AFA sonuçları incelendiğinde KMO değeri 0.798, Bartlett  $\chi^2$  değeri 7438.218 ve p değeri anlamlı çıkmıştır. Türkiye verisinin AFA sonuçlarına bakıldığında KMO değeri 0.808, Bartlett  $\chi^2$  değeri 13070.391, p değeri ise anlamlı çıkmıştır. Son olarak Katar verisinin KMO değeri 0.780, Bartlett  $\chi^2$  değeri

11251.176 ve p değeri anlamlı çıkmıştır. Bu sonuçlar göz önüne alınarak DFA analizlerine geçilmiştir.

**Tablo 8**

*Bilgisayar Kullanım Sıklığı Boyutu Madde Faktör Yükleri*

Ülke	BSBE02A	BSBE02B	BSBE02C	BSBE02D
İngiltere	0.55	0.78	0.81	0.60
Türkiye	0.70	0.82	0.83	0.71
Katar	0.58	0.78	0.77	0.60

Tablo 8 'de eTIMSS sorularının bilgisayar kullanım sıklığı boyutuna ait açımlayıcı faktör analizi sonuçlarına yer verilmiştir. Faktör değerlerinin 0.55 ile 0.83 arasında değiştiği görülmüştür.

**Tablo 9**

*Bilgisayar Kullanımı Öz Yeterlik Boyutu Madde Faktör Yükleri*

Ülke	BSBE04A	BSBE04C	BSBE04D	BSBE04E	BSBE04F	BSBE04G	BSBE04H
İngiltere	0.57	0.65	0.69	0.65	0.70	0.75	0.66
Türkiye	0.58	0.61	0.67	0.64	0.69	0.60	0.62
Katar	0.39	0.59	0.53	0.58	0.69	0.69	0.64

Tablo 9 'da eTIMSS sorularının bilgisayar kullanımına ilişkin öz yeterlik boyutuna ait açımlayıcı faktör analizi sonuçlarına yer verilmiştir. Bulgular incelendiğinde madde faktör yüklerinin 0.39 ile 0.69 arasında değiştiği görülmüştür. AFA sonucunda kurulan modelin sıklık ve öz yeterlik boyutlarında yer alan 11 maddenin ülkeler için tek boyutlu çıktığı görülmüştür.

Verinin kategorik olması sebebiyle DFA, DWLS kestirim yöntemi ile yapılmıştır. Ölçme değişmezliği gibi yapısal eşitlik modellemelerinde DWLS yöntemi kullanılabilir. DWLS yöntemi ULS yöntemine göre daha detaylı tahmin sonuçları ile kestirim yapabilir (Forero vd. ,2009).

**Tablo 10**

*TIMSS 2019 8. Sınıf Bilgisayar Kullanımına İlişkin Maddelerin DFA Uyum İndeks*

*Değerleri*

Ülke	$\chi^2$	$\chi^2 / sd$	CFI	RMSEA	AGFI	TLI
İngiltere	452.297*	8.533	0.958	0.050	0.979	0.948
Türkiye	607.696*	1.146	0.963	0.053	0.978	0.953
Katar	696.205*	1.313	0.949	0.058	0.976	0.937

p<0,05

Bilgisayar kullanımına ilişkin maddelerle kurulan modelde ülkeler bazında DFA sonuçları uyum iyiliği indekslerine Tablo 12’de yer verilmiştir. Analizler sonucu hesaplanan uyum iyiliği indeksleri incelendiğinde İngiltere için CFI, AGFI, RMSEA uyum iyiliği indekslerinde iyi uyum (CFI>.95; AGFI>.95; RMSEA<.5); TLI kabul edilebilir bir uyum (TLI >.90) düzeyinde olduğu görülmüştür.  $\chi^2$  değerinin anlamlı (p <0.05) olduğu görülmüştür ancak  $\chi^2 / sd$  değerinin 5 ‘ten büyük olduğu görülmektedir. Türkiye için uyum indekslerine bakıldığında CFI, AGFI, RMSEA ve TLI uyum iyiliği indekslerinde iyi uyum (CFI>.95; AGFI>.95; RMSEA<.5, TLI>.95) olduğu görülmektedir.  $\chi^2$  değerinin anlamlı (p <0.05) olduğu ve  $\chi^2 / sd$  değerinin iyi uyum gösterdiği görülmektedir. Katar verisinde AGFI ve RMSEA uyum iyiliği indekslerinin iyi (AGFI>.95; RMSEA<.5); CFI ve TLI indekslerinin de kabul edilebilir (CFI>.90; TLI>.90) aralıkta yer aldığı söylenebilir. Yine  $\chi^2$  değerinin anlamlı (p <0.05) olduğu ve  $\chi^2 / sd$  değerinin iyi uyum gösterdiği sonucuna ulaşılmıştır. Bu durumda kurulan modelin uyum iyiliği katsayıları incelendiğinde İngiltere, Türkiye ve Katar için 11 maddeli ve iki faktörlü yapısının doğrulandığı söylenebilir.

## Bölüm 4

### Bulgular, Yorumlar ve Tartışma

Bu bölümde araştırmanın amaçları doğrultusunda ulaşılan analiz sonuçlarına ve yorumlarına yer verilmiştir. Analiz sonuçları doğrultusunda oluşturulan modelin kültür ve dil bağlamında ölçme değişmezliği; ÇGDFA yöntemi ile ve DMF, Poly-SIBTEST VE OLR yöntemleri kullanılarak incelenmiştir.

#### Alt Problem 1a'ya İlişkin Bulgular ve Yorumlar

TIMSS 2019 öğrenci anketinde yer alan bilgisayar kullanımına ilişkin maddeleri, aynı kültür farklı dile göre ölçme değişmezliği göstermekte midir?

Tablo 13'de TIMSS 2019 bilgisayar kullanımına ilişkin maddelerin aynı kültür farklı dile göre ölçme değişmezliği sonuçları verilmiştir.

#### Tablo 11

*Aynı Kültür -Farklı Dile (ARP-İNG) Göre Değişmezlik Aşamalarında Uyum Katsayıları*

Değişmezlik Türleri	$\chi^2$	$\chi^2 / sd$	RMSEA	CFI	$\Delta CFI$	GFI	TLI
Yapısal Değişmezlik	750.560	1.10	0.064	0.94	...	0.98	0.92
Metrik Değişmezlik	842.732	1.68	0.064	0.93	0.001	0.97	0.92
Ölçek Değişmezliği	1294.309	1.24	0.077	0.89	0.04	0.96	0.88

Sonuçlar incelendiğinde Katar için kurulan modeli Arapça ve İngilizce dillerine göre yapısal değişmezlik uyum iyiliklerinin kabul edilebilir olduğu (CFI>.90; GFI>.95; RMSEA <.08 ; TLI>.90 ) görülmektedir. GFI uyum iyiliğinin iyi uyum; CFI, TLI ve RMSEA'nın ise kabul edilebilir düzeyde olduğu sonucuna ulaşılmıştır.  $\chi^2$  uyum indeksinin örneklem büyüklüğünün artmasıyla birlikte daha büyük sonuçlar verdiği görülmüştür (Kline,2016). Bu durum  $\chi^2$  test istatistiğinin sınırlılıklarından biridir. Modelin anlamlı olması ancak  $\chi^2$  test istatistiğinin kabul edilebilir aralıkta olmaması modelin reddedilmesine sebep olabilir (Çelik & Yılmaz,2013). Bu sonuçlar göz önüne alındığında uyum indeksi değerlerinin en az

kabul edilebilir düzeyde olması Katar'da maddelerin farklı dilde (İngilizce ve Arapça) alan 8.sınıf öğrencileri için ölçme aracının iki grup için aynı yapıda olduğu yani dile göre yapısal değişmezliğin sağlandığı söylenebilir. Metrik değişmezlik için ön koşul olan yapısal değişmezliğin sağlandığı söylenerek bir sonraki değişmezlik aşaması olan metrik değişmezliğin test edilmesine geçilmiştir.

Metrik değişmezliğin uyum iyiliği indeksleri incelendiğinde kabul edilebilir (CFI>.90; GFI>.95; RMSEA <.08; TLI>.90) düzeyde olduğu görülmüştür. CFI, TLI ve RMSEA'nın kabul edilebilir; GFI uyum iyiliği katsayısının iyi uyum gösterdiği gözlenmiştir. Bu sonuçlar göz önüne alındığında ölçme aracının aynı yapıda olmasının yanı sıra madde faktör yüklerinin de farklı dil de sorulara cevap veren öğrenciler için aynı olduğunu göstermektedir. Ki kare istatistiği büyük veri setlerinde yanlış sonuçlar verme olasılığı bulunması sebebiyle iki ölçme değişmezliği aşamalarının karşılaştırılmasında CFI fark değeri hesaplanmıştır (Cheung & Rensvold,2002). Hesaplanan  $\Delta CFI=0,008$  değerinin uyum için kabul edilen sınırlar olan  $\pm 0,01$  aralığında olduğu görülmüştür. Bu sonuç, 8.sınıf öğrencilerinin bilgisayar kullanımına ilişkin maddelerin dile göre metrik değişmezliği sağladığını göstermektedir. Metrik değişmezliğin sağlanmasının ardından ölçme değişmezliği aşamalarının bir sonraki aşaması olan ölçek değişmezliği test edilmiştir.

Ölçek değişmezliği aşamasındaki uyum iyilik indeks değerlerinin RMSEA ve GFI katsayıları dışındaki uyum iyiliği indekslerinin kabul edilebilir (CFI>.90; GFI>.95; RMSEA <.08 ; TLI>.90 ) düzeyde olmadığı görülmüştür. Aynı zamanda metrik değişmezliği ve ölçek değişmezliği arasındaki CFI fark değerinin de kabul edilebilir aralıkta ( $\pm 0,01$ ) yer almadığı görülmektedir. Ölçek değişmezliğinin sağlanmamış olması 8. Sınıf öğrencilerinin TIMSS 2019'daki bilgisayar kullanımına ilişkin maddelerin, madde sabitinin benzer olmadığını göstermektedir. Bu durum teze konu edinilen maddelerin dil açısından yanlılık bulunabileceğini gösterir. Elde edilen bulgularda dile göre değişmezliğin sağlanmamış olması Asil ve Brown'nun (2016) çalışmalarının bulgu sonuçları ile benzer olduğu görülmüştür.

## Alt Problem 2'ye İlişkin Bulgular ve Yorumlar

TIMSS 2019 öğrenci anketinde yer alan bilgisayar kullanımına ilişkin maddeler farklı dil farklı kültüre göre ölçme değişmezliği göstermekte midir?

Tablo 14'de TIMSS 2019 bilgisayar kullanımına ilişkin maddelerin farklı kültür farklı dile göre ölçme değişmezliği sonuçları verilmiştir.

**Tablo 12**

*Farklı Dil- Farklı Kültüre (İNG- TUR) Göre Değişmezlik Aşamalarında Uyum Katsayıları*

Değişmezlik Türleri	$\chi^2$	$\chi^2 / sd$	RMSEA	CFI	$\Delta CFI$	GFI	TLI
Yapısal Değişmezlik	1635.784	1.90	0.073	0.92	...	0.99	0.90
Metrik Değişmezlik	2276.424	2.39	0.083	0.89	0.03	0.98	0.87

Kurulan modelin ölçme değişmezliği incelendiğinde ilk aşamada yapısal değişmezliğe ilişkin uyum iyiliği katsayıları incelenmiştir. Uyum iyiliği düzeylerine bakıldığında kabul edilebilir (CFI>.90; GFI>.95; RMSEA <.08; TLI>.90) düzeyde olduğu görülmektedir.

Uyum iyiliği indeksi sonuçlarının en az kabul edilebilir düzeyde olması İngiltere ve Türkiye bağlamında maddelerin farklı dile (İngilizce ve Türkçe) alan 8.sınıf öğrencileri için ölçme aracının aynı yapıyı ölçtüğü yani dile göre yapısal değişmezliğin sağlandığı söylenebilir. Yapısal değişmezliğin sağlanmış olması metrik değişmezliğinin test edilmesi için ön koşulu sağladığı için metrik değişmezliğin test edilmesine geçilmiştir.

Metrik değişmezlikteki uyum iyiliği katsayılarının kabul edilebilir (CFI>.90; GFI>.95; RMSEA <.08 ; TLI>.90 ) düzeyde olmadığı görülmüştür. Uyum iyiliği indekslerinden yalnızca GFI indeksinin iyi uyum gösterdiği sonucuna ulaşılmıştır. Uyum iyiliği katsayılarının kabul edilebilir düzeyde olmamasının yanı sıra CFI fark değeri de anlamlı ( $\pm 0,01$ ) değildir. Bu durum madde faktör yüklerinin kültüre göre benzer olmadığı, madde yanlılığının olabileceği yorumu yapılabilir.

Özdemir (2023), TIMSS 2019 4. sınıf akran zorbalığı ölçeğinin Rusya, Tayvan ve Türkiye için cinsiyete göre katı değişmezlik sağlanırken 8. sınıf akran zorbalığı ölçeğinin



sadece yapısal değişmezlik modeli sağlanmış olduğu sonucuna ulaşmıştır. Kültüre göre yapılan ölçme değişmezliği karşılaştırmalarında her iki sınıf düzeyi için de sadece yapısal değişmezliğin sağlandığı görülmüştür. Polat (2019), TIMSS 2015 verilerinde matematik ve fen ile ilgili duyuşsal özelliklerle kurmuş olduğu modelde kültür ve bölge değişkenine göre ölçme değişmezliğini incelemiş ve bulguların sonuçlarına göre ölçek değişmezliği aşamasına kadar ölçme değişmezliğinin sağlandığı sonucuna ulaşılmıştır. Araştırma için kullanılan maddeler farklı olmasına rağmen yukarıda bahsedilen her iki çalışmanın da TIMSS döngüsüne ait olduğu görülmektedir. Bu durumda çalışmadan elde edilen sonuç ile alanyazın birlikte ele alındığında ölçme aracından öğrencilerin aynı şeyleri anlamadığı söylenebilir (Liou & Lin,2021; Senden vd., 2023). Yani TIMSS 2019 verisinde farklı anket maddeleri ile yapılan ölçme değişmezliği çalışmalarında farklı dil ve farklı kültüre göre metrik değişmezliğin sağlandığı görülmüştür.

Uluslararası sınavlarda ölçme değişmezliği ile ilgili yapılan çalışmalarda ülkelerarası karşılaştırmaların yapılmış olup yapısal değişmezliğin sağlanmış olduğu görülmüştür (Ceyhan, 2019; Gönen, 2021). Bulguların sonuçları uluslararası olarak uygulanan sınavlarda farklı değişkenler açısından ölçme değişmezliğinin sağlanmadığını ve yanlı maddelerin bulunabileceğini göstermektedir.

2a) TIMSS 2019 öğrenci anketi yer alan bilgisayar kullanımına ilişkin maddeler, aynı kültür farklı dilde OLR yöntemine göre DMF göstermekte midir?

**Tablo 13**

*Farklı Dil- Aynı Kültürde (ARP- İNG) OLR Analiz Sonuçları*

Boyut	Madde	$\chi^2_{12}$	$\chi^2_{13}$	$\chi^2_{23}$	$\beta_1$	Mc Fadden $R^2_{12}$	Mc Fadden $R^2_{13}$	Mc Fadden $R^2_{23}$	DMF Düzeyi	DMF Türü
Sıklık	1	0.968	0.000	0.000	0.000	0.000	0.001	0.001	A	TB
	2	0.033	0.000	0.000	0.023	0.000	0.010	0.009	A	TB
	3	0.100	0.000	0.000	0.017	0.000	0.009	0.009	A	TB
	4	0.255	0.000	0.000	0.019	0.000	0.008	0.008	A	TB
Öz yeterlik	5	0.575	0.325	0.164	0.007	0.000	0.000	0.000	-	-
	6	0.923	0.811	0.522	0.000	0.000	0.000	0.000	-	-

7	0.222	0.132	0.109	0.010	0.000	0.000	0.000	-	-
8	0.540	0.825	0.928	0.003	0.000	0.000	0.000	-	-
9	0.006	0.016	0.369	0.015	0.000	0.000	0.000	-	-
1	0.179	0.333	0.528	0.009	0.000	0.000	0.000	-	-
0									
1	0.125	0.309	0.989	0.012	0.000	0.000	0.000	-	-
1									

Tablo 15 'de iki farklı boyutta yer alan her madde için DMF düzeyleri incelenmiştir. DMF sonuçları McFadden Model3-Model1 R<sup>2</sup> 'ye göre değerlendirilmiştir. Sonuçlar incelendiğinde sıklık boyutunda yer alan maddelerin A düzeyinde DMF'li olduğu görülmüştür. Bu düzeyde DMF'li maddeler için DMF yoktur veya ihmal edilebilir düzeyde olduğu söylenebilir. Aynı zamanda Bonferroni tekniği (İkinci boyut için 0.012 olarak hesaplanmıştır) kullanılarak hesaplanan katsayı ile McFadden Model3-Model2 farkı karşılaştırıldığında TB DMF'nin olduğu görülmüştür.

2b) TIMSS 2019 öğrenci anketi yer alan bilgisayar kullanımına ilişkin maddeler, aynı kültür farklı dilde Poly-SIBTEST yöntemine göre DMF göstermekte midir?

#### Tablo 14

##### *Farklı Dil- Aynı Kültüre Göre (ARP -İNG) Poly-SIBTEST Analiz Sonuçları*

Boyut	Madde	$\beta_{Estimate}$	SE	p	DMF Düzeyi	Dil
Sıklık	1	0.093	0.030	0.002	C	İngilizce
	2	0.242	0.033	0.000	C	İngilizce
	3	0.210	0.031	0.000	C	İngilizce
	4	0.121	0.030	0.000	C	İngilizce
Öz yeterlik	5	-0.064	0.018	0.000	B	Arapça
	6	-0.067	0.026	0.009	B	Arapça
	7	-0.115	0.021	0.000	C	Arapça
	8	-0.074	0.028	0.008	B	Arapça
	9	-0.059	0.028	0.034	B	Arapça
	10	-0.131	0.026	0.000	C	Arapça
	11	-0.125	0.025	0.000	C	Arapça

Referans grup; Testi İngilizce alan, Odak grup; Testi Arapça alan

Tablo 16'da Poly-SIBTEST yöntemine göre farklı dil aynı kültürdeki bireylerin DMF düzeylerine yer verilmiştir. Analiz sonuçları incelendiğinde dört madde A düzeyinde, dört madde B düzeyinde, üç madde C düzeyinde DMF gösterdiği sonucuna ulaşılmıştır. 1,2,3

ve 4. Maddeler İngilizce lehine DMF gösterdiği 5, 6, 7, 8, 9 ve 10. maddeler Arapça lehine DMF gösterdiği görülmüştür.

2c) TIMSS 2019 öğrenci anketinde yer alan bilgisayar kullanımına ilişkin maddeler, farklı kültür farklı dilde OLR yöntemine göre DMF göstermekte midir?

**Tablo 15**

*Farklı Dil- Farklı Kültüre (İNG- TUR) Göre OLR Analiz Sonuçları*

Boyut	Madde	$\chi^2_{12}$	$\chi^2_{13}$	$\chi^2_{23}$	$\beta_1$	Mc Fadden $R^2_{12}$	MC Fadden $R^2_{13}$	Mc Fadden $R^2_{23}$	DMF Türü	DMF Düzeyi
Sıklık	1	0.000	0.000	0.000	0.023	0.002	0.004	0.001	TB	A
	2	0.940	0.000	0.00	0.000	0.000	0.001	0.001	TB	A
	3	0.000	0.000	0.000	0.027	0.003	0.004	0.001	TB	A
	4	0.019	0.000	0.000	0.013	0.000	0.004	0.004	TB	A
Öz yeterlik	5	0.000	0.000	0.017	0.055	0.019	0.019	0.000	TBO	B
	6	0.000	0.00	0.027	0.003	0.002	0.002	0.000	TB	A
	7	0.000	0.000	0.004	0.110	0.041	0.042	0.000	TBO	C
	8	0.000	0.000	0.150	0.010	0.017	0.017	0.000	TBO	B
	9	0.000	0.000	0.003	0.051	0.036	0.037	0.000	TBO	C
	10	0.000	0.000	0.613	0.077	0.047	0.047	0.000	TBO	C
	11	0.000	0.000	0.000	0.042	0.016	0.020	0.004	TBO	B

2d) TIMSS 2019 öğrenci anketi yer alan bilgisayar kullanımına ilişkin maddeler, farklı kültür farklı dilde Poly-SIBTEST yöntemine göre DMF göstermekte midir?

Tablo 17’de iki farklı boyutta yer alan her madde için DMF düzeyleri incelenmiştir. DMF sonuçları McFadden Model3-Model1  $R^2$  ‘ye göre değerlendirilmiştir. Sonuçlar incelendiğinde öz yeterlik boyutunda yer alan maddelerin birinin A düzeyinde, üç tanesinin B düzeyinde ve üçünün de C düzeyde DMF’li olduğu görülmüştür. Bu durumda bir maddenin ihmal edilebilir düzeyde DMF’li olduğu ancak diğer altı maddenin orta düzey veya yüksek düzeyde DMF içerdiği söylenebilir. DMF’li maddelerin türüne karar verebilmek için Bonferroni tekniği (Sıklık boyutu için 0.012; öz yeterlik boyut için 0.007 olarak hesaplanmıştır) kullanılmıştır. Hesaplanan katsayı ile McFadden Model3-Model2 farkı

karşılaştırıldığında altı maddenin TBO DMF'li; beş maddeninse TB DMF'nin olduğu görülmüştür.

**Tablo 16**

*Farklı Dile Farklı Kültüre (TUR-İNG) Göre Poly-SIBTEST Analiz Sonuçları*

Boyut	Madde	$\beta_{Estimate}$	SE	p	DMF Düzeyi	Dil
Sıklık	1	-0.028	0.023	0.229	A	Türkçe
	2	0.233	0.024	0.000	C	İngilizce
	3	0.028	0.022	0.201	A	İngilizce
Öz yeterlik	4	-0.032	0.022	0.134	A	Türkçe
	5	0.221	0.015	0.000	C	İngilizce
	6	-0.041	0.019	0.031	A	Türkçe
	7	0.254	0.014	0.000	C	İngilizce
	8	-0.243	0.020	0.000	C	Türkçe
	9	-0.368	0.020	0.000	C	Türkçe
	10	-0.264	0.015	0.000	C	Türkçe
	11	-0.113	0.013	0.000	C	Türkçe

Referans grup; Testi İngilizce alan, Odak grup; Testi Türkçe alan

Tablo 18'de Poly-SIBTEST yöntemine göre farklı dil farklı kültürdeki bireylerin DMF düzeylerine yer verilmiştir. Analiz sonuçlarına göre 1,3,4 ve 6. maddelerin A düzeyinde, 2,5,7,8,9,10 ve 11. maddelerin C düzeyinde DMF'li olduğu görülmüştür. Aynı zamanda 1,4,6,8,9,10 ve 11. maddelerinin testi Türkçe alanların lehine, 2,3,5 ve 7. maddelerin testi İngilizce alanların lehine çalışmış olduğu görülmüştür.

2e) TIMSS 2019 öğrenci anketinde yer alan bilgisayar kullanımına ilişkin maddeler, aynı kültür farklı dile göre OLR ve Poly-SIBTEST yöntemleri sonucu elde edilen bulgular birbirleriyle uyumlu mudur?

OLR ve Poly-SIBTEST yöntemlerine göre yapılan analiz sonuçları aynı kültür aynı dile göre farklılık göstermiştir. Ancak sıklık boyutuna ilişkin maddeler her iki analiz yöntemine göre düzeyleri farklı olmasına rağmen DMF'li olduğu sonucuna ulaşılmıştır.

2f) TIMSS 2019 öğrenci anketinde yer alan bilgisayar kullanımına ilişkin maddeler, farklı kültür farklı dile göre OLR ve Poly-SIBTEST yöntemleri sonucu elde edilen bulgular birbirleriyle uyumlu mudur?

Sonuçlar incelendiğinde her maddenin her iki yöntem analiz sonuçlarına göre DMF'li olduğu görülmüştür. Elde edilen sonuçlar, Oliveri ve arkadaşlarının (2012) yapmış olduğu çalışma ile karşılaştırıldığında farklı DMF yöntemlerinin DMF'li maddeleri benzer şekilde tespit ettiği ancak tespit edilen maddelerin etki büyüklüklerinin farklı olduğu sonucu ile uyumludur. Bazı maddelerde ihmal edilebilir düzeyde DMF bulunurken bazı maddelerde yüksek düzeyde DMF olduğu görülmüştür. Her iki yöntemde de aynı düzeyde DMF olan maddeler olduğu görülmüştür. TIMSS 2019 bilgisayar kullanımına ilişkin maddelerden dokuz ve onuncu maddelerin her iki yöntem sonucuna göre C düzeyinde DMF'li olduğu görülmüştür. C düzeyi, yüksek düzeyde DMF anlamına geldiğinden bu iki maddenin ölçme aracından çıkarılması önerilebilir. Üç, dört ve altıncı maddelerinde her iki yöntem ile analiz sonuçlarına göre A düzeyinde DMF olduğu görülmüştür. A düzeyinde DMF'li maddeler DMF yoktur veya ihmal edilebilir anlamına geldiği için bu maddeler için alan uzmanlarının görüşlerine başvurulabilir.

Uluslararası sınavlarla ilgili yanlılık araştırmalarında hem DMF analiz yöntemleri hem de bu maddeler ile ilgili alınan uzman görüşlerine göre çeviri ve kültüre göre yanlılığın olabileceği görülmüştür (Asil ve Gelbal,2012; Uzun ve Gelbal,2017). Uluslararası sınavlarda dilsel farklılıkların DMF'nin olası kaynaklarından biri olduğu sonucuna ulaşılmıştır (Huang ,2010; Oliveri vd., 2012; Sandilads vd., 2013 ; Liu ,2019).

## Bölüm 5

### Sonuç ve Öneriler

#### Sonuçlar

Bu araştırmada IEA tarafından yürütülen TIMSS döngülerinden 2019 verisinde bilgisayar kullanımına ilişkin 11 madde ve 2 boyuttan oluşan model kurulmuş olup, kurulan modelde ölçme değişmezliği ve DMF incelenmiştir. Araştırmanın modeli aynı kültür farklı dil, farklı dil farklı kültüre göre incelenmiştir. Ölçme değişmezliği için ÇGDFA kullanılmış olup ölçme değişmezliğinin sağlanmadığı sonucunda DMF'li maddelerin belirlenmesinde OLR ve Poly-SIBTEST yöntemleri kullanılarak yanlılık olan maddeler incelenmiştir. Bu araştırmanın sonuçları özetlenecek olursa;

#### *Alt Problem 1'e İlişkin Sonuçlar*

Alt Problem 1a için kurulan modelin aynı kültür farklı dile göre yani Katar'da ölçme aracını hem Arapça hem de İngilizce alan öğrenciler için ölçme değişmezliğinin sağlanıp sağlanmadığı incelenmiştir. Bulgular ışığında modelin yapısal ve metrik değişmezlik aşamalarını sağladığı ancak ölçek ve katı değişmezliğini sağlamadığı görülmüştür. Bu durumda madde faktör gruplarının ve madde yüklerinin farklı dile göre benzer şekilde dağıldığı söylenebilir. Ancak madde sabitlerinin ve hata varyanslarının ülkelere göre farklı olduğu görülmüştür. Bu problem durumuna ilişkin sonuçlardan anlaşılacağı üzere aynı kültüre sahip olan ancak ölçme aracını farklı dile göre yanıtlayan öğrenciler hakkında yapılacak olan değerlendirmeler öncesinde yanlılık çalışmaları yapılmalı, yanlı olan maddeler ile ilgili düzenlemeler yapıldıktan sonra karar verilmelidir.

Alt problem 1b için kurulan model farklı dil farklı kültüre göre ölçme değişmezliği incelenmiştir. Bu problem durumunda Türkiye ve İngiltere örneklemi karşılaştırılmıştır. Analiz sonuçlarına göre kurulan modelde ölçme değişmezliği yalnızca yapısal aşamasını sağlamıştır. Ulaşılan bu sonuca göre ülkeler arasında madde faktör gruplarının benzer

şekilde dağılım gösterdiği söylenebilir. Ancak madde yükleri, madde sabitleri ve hata varyanslarının farklı olduğu görülmüştür.

### ***Alt problem 2'ye İlişkin Sonuçlar***

Alt problem 2a' da kurulan modeldeki maddelerin aynı kültür farklı dile göre (Testi İngilizce ve Arapça alan bireylerle yapılan karşılaştırmalar) OLR yöntemine göre DMF'li olan maddeler incelenmiştir. Analiz sonucuna göre sıklık boyutunda yer alan dört maddede A düzeyinde yanlılığın olduğu görülmüştür. Bu durumda DMF'li olan maddelerin yok veya ihmal edilebilir düzeyde olduğu söylenebilir. Bu maddelerde yanlılığın olup olmadığına dair çıkarımlarda bulunmak için uzman görüşlerine başvurulabilir.

Alt problem 2b'de kurulan modeldeki maddelerin aynı kültür farklı dile göre (Testi İngilizce ve Arapça alan bireylerle yapılan karşılaştırmalar) poly-SIBTEST yöntemine göre DMF'li olan maddeler incelenmiştir. Analiz sonuçlarına göre beş madde C düzeyinde yani önemli düzeyde DMF içerdiği; dört madde B düzeyinde DMF'li olduğu görülmüştür. Referans grup olarak testi İngilizce alanlar seçilmiş olup sıklık boyutunda yer alan dört maddenin dördünde de C düzeyinde referans grup lehine DMF olduğu görülmüştür. Öz yeterlik boyutunda yer alan yedi maddenin dördünün B, üçünün C düzeyinde DMF'li olduğu ve bu maddelerin odak grup olan testi Arapça alan grubun lehine DMF gösterdiği sonucuna ulaşılmıştır. Bu yöntemin sonuçlarına göre bu maddeler önemli düzeyde DMF içermektedir.

Modelin OLR ve Poly-SIBTEST sonuçlarının aynı kültür farklı dile göre DMF sonuçlarının farklı olduğu görülmüştür. Her iki yöntemin tek ortak noktası sıklık boyutunda yer alan dört maddenin DMF'li olduğu görülmüştür.

Alt problem 2c'ye göre kurulan model farklı dil farklı kültüre göre (Türkiye ve İngiltere örneklemi) OLR yöntemiyle DMF'li maddeler incelenmiştir. Araştırmanın sonucunda beş madde A düzeyinde, dört madde B düzeyinde, üç maddenin C düzeyinde DMF'li olduğu görülmüştür. Elde edilen sonuçlarla Poly-SIBTEST yöntemine göre sonuçlar karşılaştırılmalı, tutarlılıklar incelenmelidir.

Alt problem 2d'de model farklı dil farklı kültüre göre Poly-SIBTEST yöntemine göre DMF'li olan maddeler incelenmiştir. Sonuçlara bakıldığında 11 madde de DMF olduğu görülmüştür. Dört madde A düzeyinde; yedi madde C düzeyinde DMF'li olarak bulunmuştur.

OLR ve poly-SIBTEST yöntemlerine göre kurulan modelde farklı dil farklı kültür açısından tutarlı sonuçları olan maddelerin olduğu görülmüştür. Beş maddeden dokuz ve onuncu maddelerin C düzeyinde; üç, dört ve altıncı maddelerin A düzeyinde DMF'li olduğu her iki yöntemle yapılan analizlerde aynıdır.

Araştırmanın sonucunda dil ve kültür değişkenine göre yapılan karşılaştırmalarda ölçme değişmezliğinin sağlanamadığı ve yanlış maddelerin olduğu görülmüştür.

## **Öneriler**

### ***Araştırmacılara Yönelik Öneriler***

Literatürde konu ile ilgili uluslararası sınavlardan PISA ile gerçekleştirilen araştırma sayısının fazla olduğu görülmüştür. Gelecekteki araştırmacılar diğer uluslararası sınav verilerinde araştırmalar yapabilir.

Bilişsel testlerde ölçme değişmezliği ve yanlılık çalışmalarının duyuşsal özelliklerle ilgili çalışmalara göre daha çok olduğu görülmüştür. Araştırmacılar duyuşsal özelliklerle ilgili çalışmalar yapabilir.

Bu araştırmada DMF'yi belirleme yöntemlerinden OLR ve Poly-SIBTEST kullanılmıştır. Gelecekteki araştırmacılar diğer farklı yöntemlerle analiz yapabilir, sonuçların uyumluluğu incelenebilir.

Yanlılık bu araştırmada istatistiksel yöntemler kullanılarak belirlenmeye çalışılmıştır. Gelecek araştırmacılar, istatistiksel yöntemler ile DMF belirlemenin yanında uzman görüşlerinden faydalanabilir. Bu durum maddeler ile ilgili daha gerçekçi çıkarımların yapılmasında faydalı olabilir.



***Araştırmaya Yönelik Öneriler***

Araştırma sonuçlarına göre modelin ölçme değişmezliğini sağlamadığı ve yanlı maddelerin olduğu görülmüştür. Gelecekteki diğer döngülerde ölçme değişmezliği ve DMF çalışmalarının yapılması, yanlılığın farklı değişkenler açısından incelenmesine olanak sağlar ve bu durum sonuçların değerlendirilmesinde olumlu katkı sağlayabilir.

Uyarılma ve çeviriden kaynaklı olarak yanlı maddelerin ortaya çıktığı göz önüne alındığında konu ile ilgili çalışan araştırmacılar, ülkedeki bireylerin kültür ve dil açısından oluşabilecek yanlılık durumlarına karşı çalışmalarını sürdürebilir.

### Kaynaklar

- Abbott, M. L. (2007). A confirmatory approach to differential item functioning on an ESL reading assessment. *Language Testing*, 24(1), 7-36. <https://doi.org/10.1177/0265532207071510>
- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of educational measurement*, 29(1), 67-91. <https://doi.org/10.1111/j.1745-3984.1992.tb00368.x>
- Asil, M. & Gelbal, S. (2012). Cross-cultural Equivalence of the PISA Student Questionnaire. *Education and Science*, 37(166), 236-249.
- Asil, M., & Brown, G. T. L. (2016). Comparing OECD PISA Reading in English to Other Languages: Identifying Potential Sources of Non-Invariance. *International Journal of Testing*, 16(1), 71–93. <https://doi.org/10.1080/15305058.2015.1064431>
- Atalay, K. (2010). *PISA 2006 öğrenci anketinde yer alan tutum maddelerinin değişen madde fonksiyonu açısından incelenmesi*. (Yüksek Lisans Tezi) Hacettepe Üniversitesi, Ankara. <https://tez.yok.gov.tr>.
- Bağdu Söyler, P. (2020). *PISA 2015 Okuma Becerileri Testinin Ana Dili Değişkenine Göre Ölçme Değişmezliğinin İncelenmesi*. (Yüksek Lisans Tezi). Ege Üniversitesi Eğitim Bilimleri Enstitüsü. İzmir. <https://tez.yok.gov.tr>.
- Başusta, B., N. (2010). Ölçme eşdeğerliği. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 58-64.
- Bialo, J. A., & Li, H. (2022). Fairness and Comparability in Achievement Motivation Items: A Differential Item Functioning Analysis. *Journal of Psychoeducational Assessment*, 40(6), 722-743. <https://doi.org/10.1177/07342829221090113>
- Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, 44(11, Suppl 3), S176–S181. <https://doi.org/10.1097/01.mlr.0000245143.08679.cc>

- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, N. J.: Guilford Press.
- Büyüköztürk, Ş. (2002). Faktör Analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamada Eğitim Yönetimi*, 32(32), 470–483. Retrieved from <http://dergipark.gov.tr/kuey/issue/10365/126871>
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2020). *Bilimsel araştırma yöntemleri* (29. Baskı). Ankara: Pegem Akademi.
- Byrne, B. M. (2006). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, application and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Byrne, B. M. (2013). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Routledge. <https://doi.org/10.4324/9781410600219>
- Byrne, B.M. (2008) Testing for Multigroup Equivalence of a Measuring Instrument: A Walk through the Process. *Psicothema*, 20, 872-882.
- Camilli G.. Shepard L. A. (1994). *Methods for Identifying Biased Test Items*. Volume 4. California: SAGE Publications. Inc.
- Camilli, G. (2006). *Test fairness*. Educational Measurement, 4, 221-256.
- Çelik, H. & Yılmaz, V. (2013). *LISREL 9 .1 İle Yapısal Eşitlik Modellemesi*, Ankara, Anı Yayıncılık
- Ceyhan, E. (2019). *PISA 2012 okuma becerileri ölçeğinin, uygulama dili doğrultusunda belirlenen ülkeler arasında ölçme değişmezliğinin incelenmesi*. (Yüksek Lisans Tezi). Akdeniz Üniversitesi. Antalya. <https://tez.yok.gov.tr>.

- Chang, H. H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement*, 33(3), 333-353.
- Cheung, G. W., and Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255.
- Clauser, B. E., & Mazor, K. M. (1998). *Using statistical procedures to identify differential item functioning test items*. *Issues and Practice*, 17(1), 31–44.
- Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları* (2. Baskı). Ankara: Pegem Akademi.
- Concurrent Internal Bias in Ability Tests. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.
- Construction versus choice in cognitive measurement (pp. 135-165). Hillsdale, NJ:
- Crane, Paul K., Gibbons, L. E., Jolley, L., & Belle, G. Van. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIF detect and difwithpar. *Medical Care*, 44, 115–123.  
<https://doi.org/10.1097/01.mlr.0000245183.28384.ed>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Demir, M.C. (2020). *TIMSS 2015 fen duyuşsal özelliklerin cinsiyet ve bölgelere göre incelenmesi*. (Yüksek Lisans Tezi). Hacettepe Üniversitesi. Ankara.  
<https://tez.yok.gov.tr>.
- Demirus, K. B., Gelbal, S. (2016). Ortak maddelerin değişen madde fonksiyonu gösterip göstermemesi durumunda test eşitlemeye etkisinin farklı yöntemlerle 104 incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 7(1), 182-201.

- Doğan C. ve Aybek, E. (2023). *\_RSP: 'shiny' applications for statistical and psychometric analysis\_*. r package version 0.1, .
- Doğan Gül, Ç. (2022). *TIMSS-2015 araştırmasının dil ve cinsiyete göre ölçme değişmezliğinin gizil sınıf analizi ile incelenmesi*. (Doktora Tezi). Ankara Üniversitesi. Ankara. <https://tez.yok.gov.tr>
- Doğan, N. & Öğretmen, T. (2005). Test ve Madde Yanlılığı. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 5 (1), Retrieved from <https://dergipark.org.tr/tr/pub/aibuefd/issue/1511/18330>.
- Dorans, N. J., & Holland, P. W. (1993). *DIF detection and description: Mantel-Haenszel and standardization*. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Lawrence Erlbaum Associates, Inc.
- Dorans, N. J., & Schmitt, A. P. (1993). *Constructed response and differential item* Erlbaum.
- Erşan, Ö. (2016). *TIMSS 2011 Sekizinci Sınıf Öğrencilerinin Matematik Başarılarını Etkileyen Faktörlerin Çok Düzeyli Yapısal Eşitlik Modeliyle İncelenmesi*. (Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Fishbein, B., Martin, M.O., Mullis, I.V.S. et al. The TIMSS 2019 Item Equivalence Study: examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assess Educ* 6, 11 (2018). <https://doi.org/10.1186/s40536-018-0064-z>
- Forero, G., C., Maydeu-Olivares, A. & Gallardo-Pujol, D. (2009). *Factor analysis with ordinal indicators: A monte carlo study comparing DWLS and ULS estimation*. *Structural Equation Modeling*, 16, 625-641.
- Fraenkel, J. R., ve Wallen, N.E. (2006). *How to design and evaluate research in education* (6th ed.). New York: McGraw-Hill.

- French, A. W., & Miller, T. R. (1996). Logistic regression and its use in detecting differential item functioning in polytomous items. *Journal of Educational Measurement*, 33(3), 315–332. <https://doi.org/10.1111/j.1745-3984.1996.tb00495.x>
- functioning: A programmatic perspective*. In R. E. Bennett & W. C. Ward (Eds.),
- Gelin, M. N., & Zumbo, B. D. (2003). Differential item functioning results may change depending on how an item is scored: An illustration with the center for epidemiologic studies depression scale. *Educational and Psychological Measurement*, 63(1), 65–74. <https://doi.org/10.1177/0013164402239317>
- Gök, B., Kabasakal, K. A., & Kelecioğlu, H. (2014). Analysis of attitude items in PISA 2009 Student Questionnaire in terms of differential item functioning based on culture. *Journal of Measurement and Evaluation in Education and Psychology*, 5(1), 72-87. <https://doi.org/10.21031/epod.64124>
- Gönen, E. (2021). *PISA 2018 okuma becerileri testinin ülkelere ve cinsiyete göre ölçme değişmezliğinin incelenmesi*. (Yüksek Lisans Tezi). Hacettepe Üniversitesi. Ankara. <https://tez.yok.gov.tr>
- Gören, S., Sayın, A., & Gelbal, S. (2024). An Analysis of Item Bias in the PISA 2018 Reading Understanding and Memorising Strategies Questionnaire, *Kastamonu Education Journal*, 32(2), 345-356. doi: 10.24106/kefdergi.xxxx
- 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), 78-94. <https://doi.org/10.1097/01.mlr.0000245454.12228.8f>
- Güler, N. (2019). *Eğitimde ölçme ve değerlendirme* (13. Baskı). Pegem Akademi
- Güngör, M. (2019). *Fen motivasyonu ve özyeterliliği modeli'nin ölçme değişmezliğinin incelenmesi: PISA 2015 türkiye örneği*. (Yüksek Lisans Tezi). Hacettepe Üniversitesi. Ankara. <https://tez.yok.gov.tr>

- Gür, E. (2019). *An investigation of the PISA 2015 in terms of differential item functioning based on culture*. (Thesis No. 582227). [Master's thesis, Hacettepe University]. Council of Higher Education Thesis Center. <https://tez.yok.gov.tr>
- Hambleton, R. K. (2006). Good Practices for Identifying Differential Item Functioning. *Medical Care*, 44(11, Suppl 3), S182–S188. <https://doi.org/10.1097/01.mlr.0000245443.86671.c4>
- He, J., Barrera-Pedemonte, F., Buchholz, J. (2019) Cross-cultural comparability of noncognitive constructs in TIMSS and PISA, *Assessment in Education: Principles, Policy & Practice*, 26:4, 369-385, DOI: 10.1080/0969594X.2018.1469467
- Hu, L. T., & 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>
- Huang, X. (2010). *Differential Item Functioning: The Consequence of Language, Curriculum, or Culture?* (Doktora Tezi). University of California, Berkeley.
- item functioning and differential test functioning. *Language Testing*, 23(4),475-496.
- Jonson, J. L., & Plake, B. S. (1998). A Historical Comparison of Validity Standards and Validity Practices. *Educational and Psychological Measurement*, 58(5), 736-753. <https://doi.org/10.1177/0013164498058005002>.
- Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement*, 38(4), 319–342. <http://www.jstor.org/stable/1435453>
- Kıbrıslıoğlu, N. (2015). *The investigation of measurement invariance PISA 2012 mathematics learning model according to culture and gender: Turkey - China (Shanghai) - Indonesia* (Tez No. 394817). [Master's thesis, Hacettepe University]. Council of Higher Education Thesis Center. <https://tez.yok.gov.tr>.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford.

- Liou, P.Y., Lin, J.J.H. (2021) Comparisons of Science Motivational Beliefs of Adolescents in Taiwan, Australia, and the United States: Assessing the Measurement Invariance Across Countries and Genders. *Front. Psychol.* 12:674902. doi: 10.3389/fpsyg.2021.674902
- Liu, R. (2019). *DIF Among English Language Learners on a Large-Scale*
- Lord, M.D. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates.
- Maria Elena Oliveri, Brent F. Olson, Kadriye Ercikan & Bruno D. Zumbo (2012) Methodologies for Investigating Item- and Test-Level Measurement Equivalence in International Large-Scale Assessments, *International Journal of Testing*, 12:3, 203-223, DOI: 10.1080/15305058.2011.617475
- Marsh, H., Hau, K., Artelt, C., Baumert, J., Peschar, J. (2006). OECD's Brief SelfReport Measure of Educational Psychology's Most Useful Affective Constructs: Cross-Cultural, Psychometric Comparisons Across 25 Countries. *International Journal Of Testing*, 311-360.
- Mathematics Assessment*. (Doktora Tezi). University of Kentucky, ABD.
- MEB. (2021). PIRLS nedir? <https://pirls.meb.gov.tr/www/pirlsnedir/icerik/2>
- Meredith, W. (1993). *Measurement invariance, factor analysis and factorial invariance*. *Psychometrika*, 58(4), 525– 543. <https://doi.org/10.1007/BF02294825>
- Milfont, T., Fischer, R. (2010). Testing measurement invariance across groups: applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111–130. <https://doi.org/10.21500/20112084.857>
- Millî Eğitim Bakanlığı [MEB] (2020). TIMSS 2019 Türkiye Ön Raporu. Eğitim Analiz ve Değerlendirme Raporları Serisi.



Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://Mlandpirls.bc.edu/timss2019/international-results/>.

OECD. (2024). PISA. <https://www.oecd.org/pisa/>

Ölçüoğlu, R. ve Çetin, S. (2016). TIMSS 2011 sekizinci sınıf öğrencilerinin matematik başarısını etkileyen değişkenlerin bölgelere göre incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 7(1), 202-220. [10.21031/epod.34424](https://doi.org/10.21031/epod.34424)

Özdemir, E. (2023). *TIMSS 2019 araştırması akran zorbalığı ölçeklerinin cinsiyete ve kültüre göre ölçme değişmezliğinin incelenmesi*. (Yüksek Lisans Tezi). Ankara Üniversitesi. Ankara. <https://tez.yok.gov.tr>.

Pae, T. I. & Park, G. P. (2006). Examining the relationship between differential

Polat, M. (2019). *TIMSS-2015 Matematik ve Fen Duyuşsal Özellik Modellerinin Kültürlere, Cinsiyete ve Bölgelere Göre Ölçme Değişmezliğinin İncelenmesi*. (Yüksek Lisans Tezi). Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü. Ankara. <https://tez.yok.gov.tr>.

Roussos, L. A., Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel- Haenszel type I error Performance. *Journal of Educational Measurement*, 33(2), 215- 230.

Samast, N. (2024). *PISA 2018 Bilgi İletişim Teknolojileri ölçeğinin ölçme değişmezliğinin incelenmesi*. (Yüksek Lisans Tezi). Ege Üniversitesi. İzmir. <https://tez.yok.gov.tr>.

Sandilands, D., Maria Elena Oliveri, M.E., Zumbo, B.D., Ercikan, K., (2013) Investigating Sources of Differential Item Functioning in International Large-Scale Assessments Using a Confirmatory Approach, *International Journal of Testing*, 13:2, 152-174, DOI: 10.1080/15305058.2012.690140

- Scheuneman, J. (1979). A Method of Assessing Bias in Test Items. *Journal of Educational Measurement*, 16(3), 143–152. <http://www.jstor.org/stable/1433816>
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18(4), 210–222. <https://doi.org/10.1016/j.hrmr.2008.03.003>
- Şekercioğlu, G. (2018). Measurement invariance: Concept and implementation. *International Online Journal of Education and Teaching (IOJET)*, 5(3). 609-634.
- Senden, B., Teig, N., & Nilsen, T. (2023). Studying the comparability of student perceptions of teaching quality across 38 countries. *International Journal of Educational Research Open*, 5, 100309.
- Shealy, R., & Stout, W. (1991a). An item response theory model for test bias. Hillsdale, NJ: Erlbaum.
- Shealy, R.T. (1989). An Item Response Theory-Based Statistical Procedure for Detecting
- Shealy, R.T., & Stout, W.F. (1991b). A model based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika*, 54, 159-194.
- Somer, O., Korkmaz, M., Dural, S. ve Can, S. (2009). Ölçme eşdeğerliğinin yapısal eşitlik modellemesi ve madde cevap kuramı kapsamında incelenmesi. *Türk Psikoloji Dergisi*, 24(64), 61-75
- Steenkamp, J.-B. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78- 90.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361–370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed). Pearson.

Thissen, D., Steinberg, L., Wainer, H. (1988). Use of item response theory in the study of group differences in tracelines. In Wainer, H., Braun, H., (Eds.), *Test validity* (pp.147-169). Hillsdale, NJ: Lawrence Erlbaum Associates.

Thomas, D. R., & Zumbo, B. D. (1996). Using a measure of variable importance to investigate the standardization of discriminant coefficients. *Journal of Educational and Behavioral Statistics*, 21(2), 110–130. <https://doi.org/10.2307/1165213>

Tiryaki, F. (2019). *PISA 2015 öğrenci tutum anketlerinin değişen madde fonksiyonu ve ölçme değişmezliğinin incelenmesi*. (Yüksek Lisans Tezi). Ankara Üniversitesi Eğitim Bilimleri Enstitüsü. Ankara. <https://tez.yok.gov.tr>.

Turgut, M.F. (1977). *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara: Nüve Matbaası

Uzun, N. B., & Gelbal, S. (2017). An investigation of item bias in pisa science test in terms of the language and culture. *Kastamonu Eğitim Dergisi*, 25(6), 2427-2446.

Vanderberg, 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(4), 4-70. <https://doi.org/10.1177/109442810031002>

West, S. G., Taylor, A. B., & Wu, W. (2012). *Model fit and model selection in structural equation modeling*. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling*, (s. 209-231). New York, US: The Guilford Press.

Yayınları.

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (Ordinal) Item Scores*. Ottawa ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zumbo, B. D. (2007). Three generations of DIF analyses: considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223–233. <https://doi.org/10.1080/15434300701375832>

## EK-A: Araştırma Etik Komisyon İzin Muafiyeti Formu



Hacettepe Üniversitesi  
Eğitim Bilimleri Enstitüsü  
Tez Çalışması/Araştırma Etik Komisyon İzin Muafiyeti Formu

F46

22 / 04 / 2024

Hacettepe Üniversitesi  
Eğitim Bilimleri Enstitüsü  
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına

<b>Tez/Araştırma Başlığı</b>	Bilgisayar kullanımına ilişkin özelliklerin kültüre ve dile göre ölçme değişmezliğinin ve yanlılığının incelenme
------------------------------	--

Yukarıda başlığı/konusu verilen tez/araştırma çalışmam,

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir.
4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırmalar niteliğinde değildir.
5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.

Çalışmada kullanacağım veriler:

- (X) Kamusal erişime açık (buraya yazınız): Tezde kullanılacak olan TIMSS 2019 verisi kamusal erişime açıktır.  
( ) Özel izin ve onaya tabi (buraya yazınız): .....  
( ) Üretilmiş veri (buraya yazınız): .....  
( ) Diğer (buraya yazınız): .....

Yükseköğretim Kurumları Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Zeynep Neveser KIZILÇİM

## Araştırmacı Bilgileri

<b>Adı Soyadı</b>	Zeynep Neveser KIZILÇİM		
<b>Öğrenci ise No</b>	N21135550		
<b>Ana Bilim Dalı</b>	Eğitim Bilimleri		
<b>Programı</b>	Eğitimde Ölçme ve Değerlendirme		
<b>Çalışma Türü</b>	<input checked="" type="checkbox"/> Tez	<input type="checkbox"/> Tezden Üretilen Yayın	<input type="checkbox"/> Araştırma Makalesi
<b>Statüsü</b>	<input checked="" type="checkbox"/> Yüksek Lisans	<input type="checkbox"/> Doktora	<input type="checkbox"/> Bütünleşik Dr. <input type="checkbox"/> Diğer

## Danışman Görüşü ve Onayı\*

Tezde kullanılacak olan TIMSS 2019 verisi kamusal erişime açık olduğundan etik kurul iznine gerek yoktur.

Doç. Dr. Sevda ÇETİN  
(İmza)  
(Danışmanın Ünvanı, Adı ve Soyadı)

**EK-B: Etik Beyanı**

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- \* tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- \* görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- \* başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- \* atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- \* kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- \* bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

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

(İmza)

Ad SOYADI

**EK-C: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu**

27/06/2024

HACETTEPE ÜNİVERSİTESİ  
Eğitim Bilimleri Enstitüsü  
Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı Başkanlığına,

Tez Başlığı: TIMSS-2019 Bilgisayar Kullanımına İlişkin Maddelerin Ölçme Değişmezliği ve Değişen Madde Fonksiyonu Açısından İncelenmesi

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
26/06 /2024	90	125129	07/06 /2024	%14	2408945046

Uygulanan filtreler:

- Kaynaklar hariç
- Alıntılar dâhil
- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esaslarını inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

**Ad Soyadı:** Zeynep Neveser KIZILÇİM

**Öğrenci No.:** N21135550

**Ana Bilim Dalı:** Eğitim Bilimleri Bölümü

İmza

**Programı:** Eğitimde Ölçme ve Değerlendirme

**Statüsü:**  Y.Lisans  Doktora  Bütünleşik Dr.

**DANIŞMAN ONAYI**

UYGUNDUR.

Doç. Dr. Sevda ÇETİN

## EK-Ç: Thesis/Dissertation Originality Report

27/06/2024

HACETTEPE UNIVERSITY  
Graduate School of Educational Sciences  
To The Department of Educational Measurement And Evaluation,

Thesis Title: Examination of Timss -2019 Computer Use Items in Terms of Measurement Invariance and Changing Item Function

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
26/06 /2024	90	125129	07/06 /2024	%14	2408945046

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

**Name Lastname:** Zeynep Neveser KIZILÇİM  
**Student No.:** N21135550  
**Department:** Educational Science  
**Program:** Educational Measurement And Evaluation  
**Status:**  Masters  Ph.D.  Integrated Ph.D.

Signature

### ADVISOR APPROVAL

APPROVED  
Assoc. Prof. Sevda ÇETİN



## EK-D: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. <sup>(1)</sup>
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir. <sup>(2)</sup>
- Tezimle ilgili gizlilik kararı verilmiştir. <sup>(3)</sup>

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

(imza)

Öğrencinin Adı SOYADI

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezini erişime açılmasının ertelenmesine karar verebilir.
  - (2) Madde 6.2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internette paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanın önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
  - (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir\*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
- Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir
- \*Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

