
GAZİ ÜNİVERSİTESİ
Gazi
1926

PISA 2018 OKUMA BECERİLERİ ALT TESTİNİN DEĞİŞEN MADDE FONKSİYONU AÇISINDAN FARKLI YÖNTEMLERLE İNCELENMESİ

Özge Erdoğan

YÜKSEK LİSANS TEZİ

EĞİTİM BİLİMLERİ ANABİLİM DALI

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EKİM, 2021

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren 24 (yirmidört) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Özge
Soyadı : Erdoğan
Bölümü : Eğitimde Ölçme ve Değerlendirme
İmza :
Teslim tarihi :

TEZİN

Türkçe Adı : PISA 2018 Okuma Becerileri Alt Testinin Değişen Madde Fonksiyonu Açısından Farklı Yöntemlerle İncelenmesi

İngilizce Adı : Examination Of The PISA 2018 Reading Skills Sub-Test With Different Methods In Terms Of Differantial Item Function

ETİK İLKELERE UYGUNLUK BEYANI

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

Yazar Adı Soyadı: Özge Erdoğan

İmza: ……..

JÜRİ ONAY SAYFASI

Özge ERDOĞAN tarafından hazırlanan "PISA 2018 Okuma Becerileri Alt Testinin Değişen Madde Fonksiyonu Açısından Farklı Yöntemlerle İncelenmesi" adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı'nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Hakan Yavuz ATAR …….……
(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi)
Başkan: Doç. Dr**.** Kübra ATALAY KABASAKAL ……..
(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Hacettepe Üniversitesi)
Üye: Doç. Dr. Dilara BAKAN KALAYCIOĞLU ……..
(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi)

Tez Savunma Tarihi: 27/09/2021

Bu tezin Eğitimde Ölçme ve Değerlendirme Anabilim Dalı'nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Yücel GELİŞLİ ……..
Eğitim Bilimleri Enstitüsü Müdürü

Biricik Oğluma

TEŞEKKÜR

Tez çalışmam boyunca bana vaktini ayırıp bilgi ve deneyimleriyle yol gösteren sevgili danışmanım Prof. Dr. Hakan Yavuz Atar'a teşekkürlerimi sunuyorum. Dönütleri ile tezime katkı sağlayan kıymetli jüri üyelerim Doç. Dr. Dilara Bakan Kalaycıoğlu'na ve Kübra Kabasakal Atalay'a teşekkürlerimi sunuyorum.

Tez sürecim boyunca her zaman desteğini hissettiğim bana inanmaktan vazgeçmeyen çok kıymetli hayat arkadaşım Erdi Erdoğan'a ve beni koşulsuz sevgisiyle motive eden biricik oğlum Umut Erdoğan'a teşekkürlerimi sunuyorum.

PISA 2018 OKUMA BECERİLERİ ALT TESTİNİN DEĞİŞEN MADDE FONKSİYONU AÇISINDAN FARKLI YÖNTEMLERLE İNCELENMESİ

(Yüksek Lisans Tezi)

Özge Erdoğan
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
Ekim, 2021

ÖZ

Bu çalışmanın amacı, PISA 2018 uygulamasında okuma becerileri alt testinde yer alan maddelerin farklı kültür ve dil değişkenine göre değişen madde fonksiyonu gösterip göstermediğini incelemektir. Bu amaç kapsamında, PISA 2018 okuma becerileri alt testinde yer alan madde paketlerinden Temel bölüm için 1.paket (Core RC1), 1.Aşama için 1.paket (Stage 1 - R11H), 2.Aşama için 1.paket (Stage 2 - R21H) seçilmiştir. Çalışmanın veri setini ise Türkiye, Fransa ve Birleşik Krallık ülkelerinden bu madde paketlerini yanıtlayan 536 katılımcının verdikleri yanıtlar oluşturmaktadır. DMF analizlerine geçilmeden önce veri seti, kayıp veri (missing value) ve uç değerler (outliers) bakımından incelenmiştir. Ayrıca grup dağılımının genel özelliklerini belirlemek için betimsel istatistik kullanılmıştır. Ayrı ayrı tüm alt gruplara yönelik madde sayısı, öğrenci sayısı, ortalama, standart sapma, çarpıklık, basıklık, en yüksek puan, en düşük puan ve Cronbach Alpha değerleri hesaplanmıştır. Ele alınan maddelerin DMF analizlerini gerçekleştirmek için Mantel-Haenszel (MH), Lojistik Regresyon (LR) ve SIBTEST (Simultaneous Item Bias Test) yöntemlerinden faydalanılmıştır. Seçilen madde paketleri içerisinden sadece iki kategorili olarak puanlanabilen 35 madde DMF analizlerine dâhil edilmiştir. Çalışma grubunda yer alan ülkeler; Türkiye – Birleşik Krallık, Türkiye – Fransa ve Fransa – Birleşik Krallık şeklinde ikişerli gruplarda analiz edilmiştir. Gerçekleştirilen analizler sonucunda; Türkiye - Birleşik Krallık grubunda MH yöntemiyle 14 madde, LR yöntemiyle 18 madde, SIBTEST yöntemiyle 11 madde DMF'li olarak tespit edilmiştir. Türkiye -

Fransa grubunda MH yöntemiyle altı madde, LR yöntemiyle sekiz madde, SIBTEST yöntemiyle beş madde DMF'li olarak tespit edilmiştir. Fransa - Birleşik Krallık grubunda MH yöntemiyle dokuz madde, LR yöntemiyle 13 madde, SIBTEST yöntemiyle sekiz madde DMF'li olarak tespit edilmiştir. Son olarak kullanılan DMF belirleme yöntemlerinin DMF gösterdikleri maddeler bakımından sonuçlarının birbiriyle ne derece uyum içinde olduğu incelenmiştir. Genel olarak bakıldığında yöntemlerin DMF'li olarak tespit ettiği madde sayıları farklı olsa da DMF'li olarak belirledikleri maddelerin çoğunlukla birbiriyle benzer olduğu tespit edilmiştir. Hesaplanan Spearman Brown sıra farkları korelasyonu katsayısına bakıldığında en yüksek uyum MH – LR yöntemleri arasında görülürken SIBTEST ile en düşük uyumun görüldüğü sonucuna ulaşılmıştır.

Anahtar Kelimeler : PISA, Mantel - Haenszel, Lojistik Regresyon, SIBTEST.
Sayfa Adedi : 114
Danışman : Prof.Dr. Hakan Yavuz Atar

EXAMINATION OF THE PISA 2018 READING SKILLS SUB-TEST WITH DIFFERENT METHODS IN TERMS OF DIFFERANTIAL ITEM FUNCTION

(M.S. Thesis)

Özge Erdoğan

GAZİ UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

October, 2021

ABSTRACT

This study aims to examine whether the items in the reading skills subtest in the PISA 2018 implementation show differential item functions according to different culture and language variables. Within the scope of this purpose, from the item packages in the PISA 2018 reading skills subtest, the 1st package for the Basic section (Core RC1), the 1st package for the 1st Stage (Stage 1 - R11H), and the 1st package for the 2nd stage (Stage 2 - R21H) were selected. The data set of the study consists of the answers given by 536 participants from Turkey, France, and the United Kingdom who answered these substance packages. Before proceeding to the DMF analysis, the data set was examined in terms of missing values and outliers. In addition, descriptive statistics were used to determine the general characteristics of the group distribution. The number of items, number of students, mean, standard deviation, skewness, kurtosis, highest score, lowest score, and Cronbach Alpha values were calculated for all subgroups separately. Mantel-Haenszel (MH), Logistic Regression (LR), and SIBTEST (Simultaneous Item Bias Test) methods were used to perform DIF analyzes of the items under consideration. Among the selected item packages, only 35 items that could be scored in two categories were included in the DIF analysis. Countries in the study group; analyzed in pairs: Turkey – UK, Turkey – France, and France – UK. As a result of the analyzes carried out; in the Turkey - United Kingdom group, 14 items with the MH method, 18 items with the LR method, and 11 items with the SIBTEST method were identified as having DIF. In the Turkey - France group, six items

with the MH method, eight items with the LR method, and five items with SIBTEST method were determined as having DIF. In the France - United Kingdom group, nine items with the MH method, thirteen items with the LR method, and eight items with the SIBTEST method were identified as having DIF. Finally, it was examined to what extent the results of the DIF determination methods used were in agreement with each other in terms of the items which having DIF. In general, although the number of items identified by the methods as having DIF is different, it has been found that the items identified as having DIF are mostly similar to each other. When the calculated Spearman-Brown rank difference correlation coefficient was examined, it was concluded that the highest agreement was seen between the MH - LR methods, while the lowest agreement was observed with SIBTEST.

Key Words : PISA, Mantel - Haenszel, Logistic Regression, SIBTEST.
Page Number : 114
Supervisor : Prof.Dr. Hakan Yavuz Atar

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU i
ETİK İLKELERE UYGUNLUK BEYANI ii
JÜRİ ONAY SAYFASI iii
TEŞEKKÜR v
ÖZ vi
ABSTRACT viii
İÇİNDEKİLER x
TABLOLAR LİSTESİ xiii
ŞEKİLLER LİSTESİ xv
BÖLÜM I 1
GİRİŞ 1
1.Problem Durumu 1
1.1.Yanlılık 7
1.2.Değişen Madde Fonksiyonu (DMF) 9
1.2.1.Tek Biçimli DMF (Uniform) 10
1.2.2.Tek Biçimli Olmayan DMF (Nonuniform) 11
1.3.DMF Belirleme Yöntemleri 12
1.3.1.Mantel – Haenszel (MH) 13
1.3.2.Lojistik Regresyon (LR) 15
1.3.3.SIBTEST (Simultaneous Item Bias Test) 17
1.4.PISA 19
2.Problem Cümlesi 24

2.1.Alt Problemler **25**

3.Araştırmanın Önemi **25**

4. Sayıltılar **28**

5. Sınırlılıklar **28**

BÖLÜM II **29**

İLGİLİ ARAŞTIRMALAR **29**

1.Yurt İçi Araştırmalar **29**

2.Yurt Dışı Araştırmalar **36**

3.DMF Belirleme Yöntemlerinin Karşılaştırılmasına Yönelik Araştırmalar **40**

BÖLÜM III **45**

YÖNTEM **45**

1.Araştırma Modeli **45**

2.Çalışma Grubu **45**

3.Verilerin Elde Edilmesi **47**

4.Verilerin Çözümlenmesi **47**

4.1 Çalışma Grubuna Yönelik Betimsel İstatistikler **51**

BÖLÜM IV **53**

BULGULAR VE YORUMLAR **53**

BÖLÜM V **68**

SONUÇ, TARTIŞMA VE ÖNERİLER **68**

1. Sonuç ve Tartışma **68**

1.1. Birinci Alt Probleme İlişkin Sonuç ve Tartışma **68**

1.2. İkinci Alt Probleme İlişkin Sonuç ve Tartışma **71**

2. Öneriler 73

KAYNAKLAR **75**

EKLER **88**

EK 1. Kayıp Veri Oranları **89**

EK 2. Fransa Örneklemi İçin Oluşturulan Uç Değer Analizi Kutu Grafiği **90**

EK 3. Türkiye Örneklemi İçin Oluşturulan Uç Değer Analizi Kutu Grafiği.....91
EK 4. Birleşik Krallık Örneklemi İçin Oluşturulan Uç Değer Analizi Kutu Grafiği 92
EK 5. Fransa ve Birleşik Krallık Örneklemi İçin DMF Belirleme Yöntemlerinin Spearman Brown Sıra Farkları Korelasyon Katsayısı 93
EK 6. Türkiye ve Fransa Örneklemi İçin DMF Belirleme Yöntemlerinin Spearman Brown Sıra Farkları Korelasyon Katsayısı 94
EK 7. Türkiye ve Birleşik Krallık Örneklemi İçin DMF Belirleme Yöntemlerinin Spearman Brown Sıra Farkları Korelasyon Katsayısı 95

TABLOLAR LİSTESİ

Tablo 1. *Herhangi Bir Madde İçin İ Yetenek Düzeyindeki Bireylere Ait Olasılık Tablosu*.14
Tablo 2. ΔMH *Değerleri İçin Yorumlama Ölçütleri* 15
Tablo 3. ΔR^2 *Değerleri İçin Yorumlama Ölçütleri* 17
Tablo 4. β_{UNI} *Değerlerini Yorumlama Ölçütleri*.....19
Tablo 5. *Çalışma Grubuna Ait Frekans Ve Yüzde Değerleri* 46
Tablo 6. *Çalışma Grubuna Yönelik Betimsel İstatistikler* 52
Tablo 7. *PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Birleşik Krallık Gruplarında MH Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler* 54
Tablo 8. *Tablo 7'de Yapılan Değerlendirme Özeti*.....55
Tablo 9. *PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Birleşik Krallık Gruplarında LR Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler*.....55
Tablo 10. *Tablo 9'da Yapılan Değerlendirme Özeti* 56
Tablo 11. *PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Birleşik Krallık Gruplarında SIBTEST Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler* 57
Tablo 12. *Tablo 11'de Yapılan Değerlendirme Özeti*.....58
Tablo 13. *PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Fransa Gruplarında MH Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler* 58
Tablo 14. *Tablo 13'te Yapılan Değerlendirme Özeti* 59
Tablo 15. *PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Fransa Gruplarında LR Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler* 59
Tablo 16. *Tablo 15'te Yapılan Değerlendirme Özeti* 60
Tablo 17. *PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Fransa Gruplarında SIBTEST Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler*.....60
Tablo 18. *Tablo 17'de Yapılan Değerlendirme Özeti*.....61

Tablo 19. *PISA 2018 Okuma Becerileri Alt Testinin Fransa ve Birleşik Krallık Gruplarında MH Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler*.....62
Tablo 20. *Tablo 19'da Yapılan Değerlendirme Özeti* 63
Tablo 21. *PISA 2018 Okuma Becerileri Alt Testinin Fransa ve Birleşik Krallık Gruplarında LR Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler* 63
Tablo 22. *Tablo 21'de Yapılan Değerlendirme Özeti*.....64
Tablo 23. *PISA 2018 Okuma Becerileri Alt Testinin Fransa ve Birleşik Krallık Gruplarında SIBTEST Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler* 64
Tablo 24. *Tablo 23'te Yapılan Değerlendirme Özeti* 65
Tablo 25. *Farklı Yöntemlerin Belirledikleri DMF'li Madde Sayısı* 66
Tablo 26. *DMF Belirleme Yöntemlerinin Gruplara Göre Hesaplanan Spearman Brown Sıra Farkları Korelasyon Katsayıları*.....66

ŞEKİLLER LİSTESİ

Şekil 1. DMF'li olmayan madde grafiği (Zumbo,1999, s.19).....10
Şekil 2. Tek biçimli DMF gösteren madde grafiği (Zumbo,1999, s.20).....11
Şekil 3. Tek biçimli olmayan DMF gösteren madde grafiği (Zumbo, 1999, s.21).....12
Şekil 4. PISA döngülerinde temel ve ağırlıklı alanlar. MEB. (2019). *PISA 2018 Türkiye Ön Raporu*, (Rapor No:10), Aralık 2019.....20
Şekil 5. Türkiye örnekleminin İBBS düzey 1 bölgelerine dağılımı. MEB. (2019). *PISA 2018 Türkiye Ön Raporu*, (Rapor No:10), Aralık 2019.....21
Şekil 6. PISA 2018 Türkiye örnekleminin okul türlerine göre dağılımı. MEB. (2019). *PISA 2018 Türkiye Ön Raporu*, (Rapor No:10), Aralık 2019.....22
Şekil 7. Temel, 1.aşama ve 2.aşama bölümlerini birbirine bağlayan yönlendirme yolları. OECD (2019). *PISA 2018 Technical Report Chapter* , PISA:OECD 23

SİMGELER VE KISALTMALAR LİSTESİ

PISA	Programme for International Student Assessment (Uluslararası Öğrenci Değerlendirme Programı)
OECD	Organisation for Economic Co-operation and Development (Ekonomik Kalkınma ve İşbirliği Örgütü)
TIMSS	Trends in International Mathematics and Science Study (Uluslararası Matematik ve Fen Eğilimleri Araştırması)
PIRLS	Progress in International Reading Literacy Study (Uluslararası Okuma Becerilerinde Gelişim Projesi)
MEB	Millî Eğitim Bakanlığı
DMF	Değişen Madde Fonksiyonu
MH	Mantel Haenszel
LR	Lojistik Regresyon
SIBTEST	Simultaneous Item Bias Test

BÖLÜM I

GİRİŞ

1.Problem Durumu

Yaşadığımız yüzyılda bilgi ve teknolojinin hızla değişim göstermesi nedeniyle yaşanan gelişmelere ayak uydurmak bütün ülkeler için ortak amaç haline gelmiştir. Bu amacı gerçekleştirmek ise ancak bilim ve teknolojiyi önemseyerek hayatın her alanında bu yenilikleri kullanabilecek bireyler yetiştirmekle mümkün olacaktır (Milli Eğitim Bakanlığı [MEB], 2019). Günümüzde küreselleşmenin hızlanması ve teknolojinin hızla gelişmesiyle, birbirinden uzak olan ülkeler ve toplumlar iç içe geçmeye başlamış, sınırlar ortadan kalkmıştır (Baumann, 2020). Bu durum, toplumların dünyada yaşanan gelişmelere ve yeniliklere karşı kendi içinde kapalı kalamamasına neden olmuş, toplumlar arası etkileşim kaçınılmaz bir durum haline gelmiştir. Bu nedenle toplumlar, yoğun bir etkileşime ve rekabet ortamının içine girmişlerdir. Toplumların bu rekabet alanında öne çıkabilmeleri, refah seviyelerini yükseklere taşıyabilmeleri ve kültürel kimliklerini sürdürebilmeleri, çağın gerektirdiği niteliklere sahip bireyleri yetiştirebilen bir eğitim sistemiyle mümkün olacaktır (Milli Eğitim Bakanlığı [MEB], 2011).

Eğitimde politika geliştiriciler, başarıyı yakalama ve rekabet ortamında varlığını hissettirebilmek adına, uluslararası başarı karşılaştırma çalışmalarına katılarak ülkelerindeki eğitim sisteminin kalitesi hakkında bilgiler elde etmektedirler (Hambleton, 2002; Hambleton & Patsula, 1998, Hambleton & Patsula, 1999; Sireci & Berberoğlu, 2000; Yayan & Berberoğlu, 2004; Zumbo, 2003). Ülkemizde, eğitim alanında belirli aralıklarla öğrencilerin seviyelerini belirlemek, belirlenen seviyeye dayalı karar almak ve mevcut olan eğitim sisteminin durumunu değerlendirmek amacıyla ulusal ve uluslararası boyutta sınavlar yapılmaktadır. Sınavların özellikle uluslararası boyutta gerçekleştirilen kısmı, eğitim sistemine bir ayna tutması ve bahsi geçen rekabet sürecindeki yerimizi görebilmemiz adına oldukça önemli olarak değerlendirilmektedir. Ulusal düzeyde yapılan

sınavlar, ülke sınırları içindeki durumu ortaya koyarken, uluslararası düzeyde yapılan sınavlar, en gelişmiş olarak nitelendirilen eğitim sistemlerine göre konumumuzu nesnel bir şekilde görebilmeyi sağlamaktadır. Eğitim sistemine yönelik tamamlanması gereken eksikliklerin ve alınması gereken önlemlerin neler olduğunu belirleyebilmek bakımından, uluslararası ölçekte uygulanan sınavların giderek artan bir değer kazandığını söylemek yanlış olmayacaktır (MEB, 2019).

Birçok ülkenin katıldığı uluslararası düzeyde yapılan sınavların başında Uluslararası Matematik ve Fen Eğilimleri Araştırması (Trends in International Mathematics and Science Study: TIMSS), Uluslararası Öğrenci Değerlendirme Programı (Programme for International Student Assessment: PISA) ve Uluslararası Okuma Becerileri Projesi (The Project of International Reading Language Skills: PIRLS) gelmektedir. PISA, TIMSS ve PIRLS gibi uluslararası çalışmalar, ülkelerin eğitim sistemlerini karşılaştırmak, değerlendirmek ve iyileştirmek için geniş bir perspektif sunar. Bu nedenle, ülkeler bu sınavlara önem vererek mevcut durumlarını ortaya koymakta ve uluslararası rekabet ortamında konumlarını benzer özellikler gösteren diğer ülkelere göre kolayca kıyaslayabilmektedirler. Bunun yanı sıra, başarıya yönelik verilerin dışında, başarıyı farklılaştıran değişkenleri de belirleyerek eğitim sistemlerinin mevcut durumunun ortaya çıkarılmasına da katkıda bulunmaktadır (Gök, Atalay-Kabasakal & Kelecioğlu, 2014).

Uluslararası düzeyde yapılan uygulamalardan biri olarak değerlendirilen PISA, Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından hazırlanıp üç yıl ara ile yapılan, 15 yaş grubuna dâhil olan öğrencilerin kazandıkları bilgi ve becerileri değerlendiren uluslararası düzeyde gerçekleştirilen en büyük eğitim araştırmalarından biridir. Bu uluslararası araştırma, öğrencilerin okulda öğrendikleri bilgilerin ne kadarını hatırlayabildiklerinden ziyade bu bilgileri günlük yaşama aktarabilme ve karşılaşacakları yeni durumlara var olan bilgilerini uyarlayabilme becerilerini ölçmektedir. PISA uygulamasında, katılım sağlayan öğrencilerin Okuma Becerileri, Matematik ve Fen alanlarına yönelik bilgi ve becerilerinin ölçülmesi dışında öğrencilerin kendileri hakkındaki fikirleri, içsel güdülenmeleri, öğrenme stilleri, aileleri ve okul ortamları ile ilgili bilgileri de toplanmaktadır (MEB, 2019). 2000 yılında uygulanmaya başlanan PISA uygulamasına ülkemiz OECD üyesi olarak ilk kez 2003 yılında dâhil olmuştur. Öğrencilerin başarı düzeylerini ve eğitim sistemini diğer katılımcı ülkelerin verileriyle kıyaslayarak güçlü ve zayıf taraflarını saptamak amacıyla PISA uygulamasına katılım sağlanmaktadır (MEB, 2019).

PISA uygulamasında ölçme araçları İngilizce ve Fransızca olmak üzere orijinal olarak iki dilde hazırlanmakta ve daha sonra katılımcı ülkelere kendi ulusal dillerine çevirmeleri için gönderilmektedir (The Organisation For Economic Co-Operation And Development [OECD], 2007). Farklı diller ve kültürler arası karşılaştırma yapmaya olanak sağlayan çalışmaların, bazı yöntem problemlerinden dolayı iyi bir şekilde uygulanması oldukça zordur (Hambleton, 2002). Testler ya da anketler farklı dillere çevrildiğinde yeni formun eşdeğerliğinin sağlanması gerekmektedir. Aksi halde çevrilen yeni formun zorluk ya da aşinalık derecesi uygulamaya katılan bireyler için farklı olabilir ve farklı bir yapıyı ölçebilir (Hambleton, 1994; Sireci & Berberoğlu, 2000). Testlerin ya da anketlerin çevrilmesindeki amaç; ayrı ayrı tüm maddelerin, bir kültürdeki birey için ne anlam taşıyorsa diğer kültürlerdeki bireyler için de maddenin aynı anlamı taşımasıdır (Hui & Triandis, 1985; Hambleton & de Jong, 2003). Bir çalışma farklı dil veya kültürden gelen bireyleri kıyaslamak amacı taşıyorsa, bu kıyaslamanın anlamlı olabilmesi için testle ölçülen yapının eşdeğer olması gerekmektedir. Bu amacı taşıyan çalışmalarda gruplar arası karşılaştırmaların yapılabilmesi için ölçme eşdeğerliğinin sağlanması temel bir sayıltıdır (Gierl, 2000; Reise, Widaman & Pugh, 1993; Van de Vijver & Tanzer, 2004; Vandenberg, 2002; Vandenberg & Lance, 2000). Testlerde ölçülen yapının eşdeğerlik göstermesi için, gözlenen değişkenlerle örtük değişkenler arasındaki ilişkilerin farklı gruplarda aynı olması gerekmektedir (Drasgow & Kanfer, 1985). Bir başka deyişle, ölçülen özellik bakımından eşit yetenek düzeyinde olan fakat farklı alt gruplardan gelen bireylerin bir maddeyi farklı yanıtlaması o maddenin eşdeğer olmadığı farklı fonksiyon gösterdiği anlamına gelir (Chan, 2000).

Ulusal ve uluslararası düzeyde düzenlenen uygulamalarda, uygulamaya katılan bireyler çeşitli özellikler bakımından birbirlerinden ayrılmaktadır. Sınav dilinin farklı olması, farklı demografik, sosyo-ekonomik ve kültürel özelliklere sahip olma, cinsiyet gibi özellikler katılımcıların performanslarına etki etmektedir (Asil & Gelbal, 2012; Ercikan, 1998). Uygulama sonucundan elde edilecek puanlar yorumlanırken bireylerin performansını etkileyebilecek olan bu değişkenler ile alt grupların özellikleri göz önünde bulundurulmalıdır (Uzun & Gelbal, 2017). Özellikle sonuçlarıyla ülkelerin eğitim sistemlerine yön veren, birçok ülkenin katıldığı uluslararası çapta yapılan uygulamaların minimum miktarda hata içermesi beklenir. Burada sonuçların tümüyle hatadan arınmış olması olanaksızdır. Fiziksel özelliklerin aksine zekâ, yaratıcılık, problem çözme gibi psikolojik özellikler doğrudan gözlemlenemeyeceği için sonuçlara hata karışması

kaçınılmazdır. Aksi durumda, ölçme araçlarından elde edilen puanların mükemmel biçimde güvenilir ve geçerli olması beklenirdi. Ancak tüm bu hata risklerine rağmen hatayı en düşük düzeyde sınırlandırabilen yüksek geçerlikte ölçme araçları geliştirilebilir (Crocker & Algina, 1986).

Farklı kültür ve dil gruplarını katılımcı olarak barındıran geniş ölçekli bir uygulamanın sonuçlarına dayanılarak, ülkelerin eğitim sistemleri hakkında doğru yorumlamalar ve ülkeler arası karşılaştırmalar yapabilmek için ölçülen yapının gruptan bağımsız olması gerekir. Gruptan bağımsız olmak, aynı yetenek seviyesindeki katılımcıların farklı alt gruplardan gelseler dahi test puanlarının eşit olması anlamını ifade eder (Osterlind, 1983; Hambleton & Rogers, 1994; Zumbo, 1999). PISA uygulamasında ölçülecek yapının gruptan bağımsız olması koşulu sağlanamazsa yapılacak karşılaştırmalarda ortaya çıkan farklılık gerçek durumdan mı yoksa yapının gruptan gruba farklılık göstermesinden mi kaynaklandığı bilenemeyecektir (Somer, Korkmaz, Dural & Can, 2009). Bu nedenle farklı kültür ve dile uyarlanan PISA uygulamasında ölçme aracının geçerliği, sonuçlara karışan hata miktarı, maddelerin yanlılığı gibi konular önem kazanmaktadır (Gök, Kabasakal & Kelecioğlu, 2014).

Ölçme araçlarının uygulanmasının ardından, çıkan sonuçlardan yola çıkarak verilecek kararların yerinde olması, ölçme sonuçlarının geçerli ve güvenilir olması ile yakından ilgilidir (Clauser & Mazor, 1998). Ancak bazı durumlarda öğrencilerin ölçülmek istenen özelliklerine farklı değişkenler karışabilir. Bu değişkenler, testin geçerliğini olumsuz yönde etkileyip testin yanlı olmasına neden olabilir (Gök vd., 2014). Yanlılık, sistematik hata kapsamına girer ve geçerliği tehdit eder (Camilli & Shepard, 1994). Bir test yanlıysa, bu test sonucuna dayanarak sınava girenler hakkında isabetli kararlar vermek mümkün olmayacaktır (Zumbo,1999).

Yanlılık, aynı yetenek seviyesinde bulunan ancak farklı alt gruplarda yer alan bireyler için test maddelerinin daha zor ya da daha kolay yanıtlanabilir olması durumdur (Zumbo, 1999). Bir başka deyişle, yanlılık, aynı yetenek seviyesinde yer alan ancak farklı alt gruplarda bulunan bireylerin bir maddeyi doğru yanıtlama olasılıklarının farklı olması olarak ifade edilmektedir (Osterlind, 1983; Hambleton & Rogers,1996; Zumbo,1999; Raju & Ellis, 2002). Bir ölçme aracı geliştirildiğinde, bu araçtan elde edilen puanların farklı alt gruplara ait olma durumuna göre değişiklik göstermemesi gerekir (Meredith & Millsap, 1992; Millsap, 2005). Uygulanan ölçme aracında yer alan maddelerin, hiçbir alt grubun lehine ya da aleyhine işlememesi, bir başka ifadeyle, ölçme aracının yanlılık içermemesi

istenir. Gerçekleştirilen bir uygulamada hata kaynaklarının alt gruplardan birine haksız bir avantaj sağlaması yanlılık için önemli bir göstergedir (Crocker & Algina, 1986).

Teste katılım sağlayan bireylerin eşit fırsatlara sahip olma durumunu belirlemenin ilk adımı, değişen madde fonksiyonunu (DMF) ortaya çıkarmaktır (Camilli & Shepard, 1994). Farklı gruplara ait puanların karşılaştırılabilir olması çoğunlukla DMF analizleri ile değerlendirilir (Ercikan, Gierl, McCreith, Puhan & Koh, 2004). DMF çalışmaları, yanlılığın istatistiksel bir göstergesi olarak düşünülebilir (Angoff, 1993). DMF, yanlı maddeleri tespit etmeye yardımcı olan istatiksel bir süreçtir (Zumbo & Gelin, 2005). Yanlılık araştırmalarında yapılacak ilk iş, testte yer alan maddelere istatiksel analizler uygulayarak DMF içerip içermediklerinin tespit edilmesidir. Bir maddenin yanlı olduğunu ileri sürebilmek için öncelikle o maddenin DMF göstermesi beklenir. İstatiksel olarak DMF gösteren maddeler belirlendikten sonra bu maddelerin farklılık gösterme kaynakları tespit edilmelidir. DMF'nin varlığı alt gruplar arasındaki gerçek farklılıktan ya da madde yanlılığından meydana geliyor olabilir (Camilli & Shepard, 1994). DMF içerdiği belirlenen maddelerin neden DMF içerdiğine yönelik olası nedenler tespit edilir. Ardından, maddenin ölçülmek istenen yapıdan ilgisiz olarak bazı alt gruplar için haksız bir kolaylık sağlayıp sağlamadığının uzmanlarca tespit edilmesi gerekir (Camilli & Shepard, 1994; Zumbo, 1999; Zumbo, 2007).

Geliştirilen bir ölçme aracı farklı gruplarda kullanılacaksa madde seçimi sürecinde DMF analizlerinden faydalanılmalıdır (Crocker & Algina, 1986; Mellenberg, 1983; Hambleton & Patsula, 1999; Zumbo, 2003; Zumbo, 2005). Aynı yetenek düzeyinde yer alan bireylerin maddeleri grup farklılıklarına göre yanıtlamasını ortaya çıkaran DMF analizleri, geçerliğin değerlendirilmesinde önemli bir göstergedir. Bir testte DMF'li madde ya da maddelerin yer alması, testin geçerliğini düşüren bir durum olarak kabul edilmektedir (Finch & French, 2013; French & Finch, 2015; Li & Zumbo, 2009; Thissen, Steinberg & Wainer, 1993). Özellikle eğitim ve psikoloji alanlarında testlerin değerlendirme araçları olarak kullanılması ve test sonuçları doğrultusunda önemli kararların alınması göz önünde bulundurulduğunda testte bulunan maddelerin DMF içerip içermeme durumu endişe kaynağı olarak ortaya çıkmaktadır (Balluerka, Gorostiaga, Gómez-Benito & Hidalgo, 2010; Thissen vd., , 1993; Thissen & Steinberg, 1988).

Bir testte yer alan maddelerin DMF içerip içermediğini ortaya çıkarmayı sağlayan birçok yöntem bulunmaktadır. Mantel-Haenszel (Holland & Thayer, 1988), Lojistik Regresyon (Swaminathan & Rogers, 1990), SIBTEST (Shealy & Stout, 1993) geliştirilen

yöntemlerden bazılarıdır. DMF belirleme yöntemleri kullanılırken "odak" ve "referans" olarak isimlendirilen iki grubun maddelere verdikleri yanıtlar kıyaslanmaktadır. Gerçekleştirilen analizlerin ardından her bir madde için DMF bulunup bulunmadığı ve DMF varsa hangi düzeyde olduğuna yönelik sınıflandırma yapılmaktadır (Swanson, Clauser, Case, Nungester & Featherman, 2002). Ancak bu yöntemler, maddeleri DMF'li olarak belirlerken izledikleri farklı matematiksel yöntemler, kullandıkları farklı algoritmalar ve kesme noktaları nedeniyle tamamıyla bir uyum içinde olmaları söz konusu değildir (Doğan & Öğretmen, 2008; Gök, Kelecioğlu & Doğan, 2010; Alatlı & Şenel, 2020; Bakan Kalaycıoğlu & Berberoğlu, 2010). DMF'yi belirlemek için güçlü bir istatistiksel yöntem üzerinde ortak bir fikre varılamamıştır. Bu nedenle yanlılık incelemelerinde birden fazla DMF belirleme yönteminden faydalanılması ve elde edilen sonuçların birbiriyle kıyaslanarak karar verilmesi tavsiye edilmektedir (Hambleton, 2006).

Testte DMF gösteren maddeler ortaya çıkarıldıktan sonra, bu maddelerin yanlılık içerip içermediğini belirlemek için DMF'ye kaynaklık eden değişkenlerin belirlenmesi gerekmektedir (Clauser & Mazor, 1998). DMF analizlerinde kişilerin ait olduğu ülke, kültür, dil, sosyo-ekonomik düzey, cinsiyet gibi değişkenler grup değişkeni olarak kategorize edilmektedir (Camili, 2006). Farklı kültür ve dil gruplarını karşılaştırmayı amaçlayan uluslararası çalışmalarda ölçme aracının yetersiz çevirisi, kavramların farklı kültürlerde farklı anlamları ifade etmesi, bazı kavramlara farklı kültürlerden gelen öğrencilerin farklı düzeyde aşina olması, ülkelerdeki eğitim programları ve öğretim yöntemlerinin farklılığı DMF kaynakları olarak görülmektedir (Hambleton, Merenda & Spielberger, 2005; Ercikan, 1998; Grisay, De Jong, Gebhardt, Berezner & Halleux-Monseur, 2007; Yıldırım & Berberoğlu, 2009; balde, 2010).

DMF gösteren bir madde, karşılaştırmak istediğimiz gruplardan birinde, tüm yetenek seviyelerinde daha yüksek oranda doğru yanıtlama olasılığına sahipse bu durumda tek biçimli DMF gerçekleşmiş olur. Bu tür maddelerin iki grup için belirlenen madde karakteristik eğrileri kesişmez (Swaminathan & Rogers, 1990). Madde karakteristik eğrileri kesişen maddeler tek biçimli DMF'nin aksine, tüm yetenek düzeyleri boyunca aynı gruba avantaj sağlamazlar.Belirli bir yetenek düzeyinde bir gruba kolaylık sağlanırken, başka yetenek düzeylerinde ise başka bir gruba kolaylık sağlanıyorsa bu tek biçimli olmayan DMF olarak adlandırılır (Hambleton, Swamanithan & Rogers, 1991; Osterlind & Everson, 2009). Buna göre, tek biçimli DMF gösteren maddeler tek biçimli olmayan DMF gösteren maddelere göre bir gruba daha önemli ölçüde avantaj sağlamaktadır. Ayrıca

kesişen madde karakteristik eğrilerine sahip maddelerin DMF bulgularının yorumlanması oldukça güçtür (Smith & Reise, 1998). Bu nedenlerle, bu çalışmada yalnızca tek biçimli DMF analizlerine yer verilmiştir.

Bu çalışmada, uluslararası eğitimsel değerlendirme çalışmalarından biri olan PISA 2018 uygulamasında yer alan Okuma Becerileri alt testinin farklı kültür ve dile göre değişen madde fonksiyonu gösterip göstermediğinin üç farklı yöntemle (MH, LR ve SIBTEST) incelenmesi amaçlanmıştır. PISA uygulamasında ölçme araçları İngilizce ve Fransızca olmak üzere orijinal olarak iki dilde hazırlanmakta ve daha sonra katılımcı ülkelere kendi ulusal dillerine çevirmeleri için gönderilmektedir (OECD, 2007). Bu nedenle bu çalışmada, kıyaslama yapabilmek adına PISA uygulamasını kendi ulusal dilinde alan Birleşik Krallık ve Fransa örneklemleri ile testi kendi ulusal dilinde bir çeviri aracılığıyla uygulayan Türkiye örneklemi tercih edilmiştir. MH, LR ve SIBTEST yöntemleri kullanılarak yürütülen analizlerde, DMF'ye yönelik ortaya çıkan sonuçların birbiriyle uyumu incelenmiştir.

1.1.Yanlılık

Yanlılık kavramı ilk olarak 1900'lü yılların başında Alfred Binet'in çocuklar üzerinde uyguladığı zekâ testi çalışmalarında ortaya çıkmıştır. Test maddelerine verilen cevapların, zekâ kapasitesinden çok kültürel farklılıklardan etkilenebileceği sonucuna ulaşılmıştır (Camilli & Shepard, 1994). 1916 yılında düzenlenen Stanford-Binet zekâ testinde ise madde yanıtlarının yaşanılan bölge ile cinsiyete göre değiştiği gözlenmiştir. Bu bağlamda, zekâ üzerinde dil, ırk, cinsiyet, yaş, ekonomik düzey, yaşanılan bölge gibi farklı değişkenlik kaynaklarının olabileceği çıkarımı yapılmıştır. Zekâ testlerinin bu değişkenlik kaynaklarına göre farklılık gösterdiğinin anlaşılması üzerine 1950 yıllarında yanlılık kavramına verilen önem artmıştır (Angoff, 1993). Eells, Allison, Havighurst, Herrick & Tyler (1951) değişkenlik kavramlarının yanında, zekâ testinde bulunan bazı maddelerin yapısından dolayı öğrencilerin yeteneklerini doğru bir biçimde yansıtmadığını söyleyerek yanlılıktan bahsetmiştir.

Camilli ve Shepard (1994) yanlılığı, bir grubun üyeleri için, test ölçümleri sonucunun gerçek yetenek düzeyini yansıtmayarak geçersiz ya da sistematik hatalı olması olarak ifade etmiştir. Clauser ve Mazor (1998) yanlılığı, ölçme aracından kaynaklanan ve geçerliği olumsuz yönde etkileyen bir tehdit olarak açıklamıştır. Zumbo (1999) ise yanlılığı, testte yer alan bazı maddelerin sahip olduğu özelliklerden ya da testin uygulanış amacına uygun

olmayan bazı test koşullarından dolayı, aynı yetenek düzeyindeki farklı gruplarda maddenin doğru cevaplanma olasılığının farklılaşması durumu olarak tanımlamıştır.

Camilli ve Shepard (1994) test maddelerini hazırlayan uzmanların bazen hazırladıkları maddeleri yetenek düzeyi yeterli olduğu halde sahip olduğu demografik özellikten dolayı bir grubun doğru yanıtlayamadığını gözlemlemiştir. Bu durum, testin ölçmek istediği yapıyı doğru bir şekilde ölçemediğini göstermekte ve yanlılığı ortaya çıkarmaktadır (Camilli & Shepard, 1994).

Yanlılık, belirli bir demografik özelliğe sahip gruba avantaj ya da dezavantaj sağlanmasına neden olan sistematik hata olarak tanımlanabilir (Osterlind & Everson, 2009). Çünkü test sonuçlarındaki hatanın yönü ve miktarı sadece belirli bir grubun üyelerine yönelik bozucu bir etki yaratmaktadır (Camilli & Shepard, 1994). Test yanlılığı, bir testin belirli bir gruba avantaj ya da dezavantaj sağlayarak sistematik hata göstermesi olarak tanımlanır (Camilli & Shepard, 1994). Madde yanlılığı ise, yanlılığın madde düzeyine indirgenmesidir (Osterlind & Everson, 2009).

Test sonuçlarına bakıldığında gruplardan hiçbirinin etkilenmemesi ya da her iki grubun da eşit etkilenmesi durumu söz konusu ise burada yanlılıktan bahsedilemeyeceği unutulmaması gereken bir noktadır. Yani eğer söz konusu madde ya da maddeler herkes için tehdit oluşturuyorsa bunu yanlılık olarak ifade etmek yanlış olacaktır (Zumbo, 1999).

Aynı yetenek düzeyindeki bireylerin hangi alt grupta olduklarına bakılmaksızın maddeyi doğru cevaplama olasılığı eşit ise maddenin yanlılık göstermediği söylenebilir; ancak aynı yetenek düzeyinde bulunmalarına rağmen başka alt gruplarda bulunan katılımcıların maddeyi doğru cevaplama olasılığı değişkenlik gösteriyorsa maddenin kesinlikle yanlı olduğunu söylemek doğru olmayacaktır. Bu noktada test puanlarındaki farklılığın gerçekten yanlılığı yansıtıp yansıtmadığı ve hangi maddelerin buna neden olduğu sorusu ortaya çıkmaktadır (Zumbo, 2007).

Maddenin yanlılığı ile madde etkisi kavramları sık sık birbiriyle karıştırılmaktadır. Madde etkisi ve madde yanlılığını ayırt edebilmek için şu soru yöneltilmelidir; Gruplarda ortaya çıkan puan farklılıkları, başarı veya yetenekteki gerçek farklılıktan mı, yoksa farklı değişkenlik kaynaklarının sebep olduğu farklılıktan mı ileri gelmektedir? (Sırgancı, 2012).

Madde etkisi, grupların farklı yetenek düzeylerinde yer aldıkları durumlarda görülür. Farklı gruplarda bulunan bireylerin maddeye doğru yanıt verme olasılıkları birbirinden farklı olacaktır. Çünkü gruplar arasındaki fark, madde ile ölçülmek istenen özelliğe yönelik yetenek düzeylerinin farklılaşmasından dolayı ortaya çıkar (Zumbo, 1999). Gruplar arası

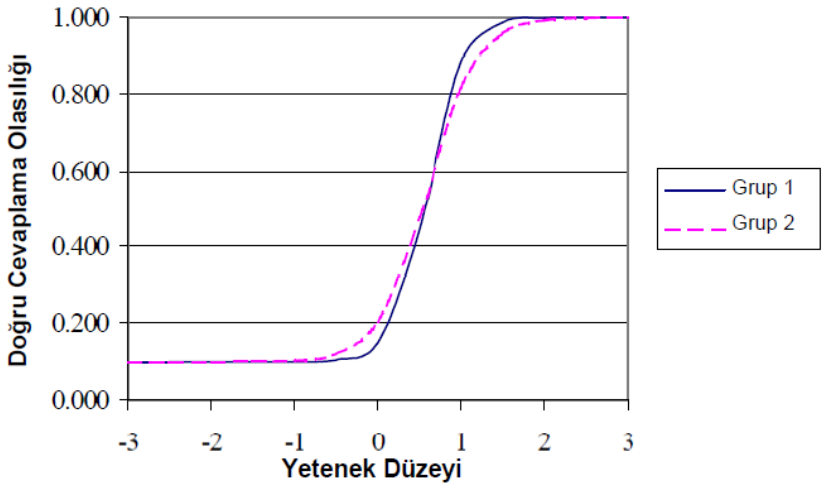
farklılıklar yetenek düzeylerinin farklı olmasından kaynaklanıyorsa yanlılık kanıtı olarak kabul edilemezler (Clauser & Mazor, 1998).

1.2.Değişen Madde Fonksiyonu (DMF)

Değişen madde fonksiyonu (DMF), testin ölçmeyi planladığı özellik bakımından aynı yetenek seviyesinde yer alan ancak ırk, cinsiyet, dil, sosyo-ekonomik düzey gibi demografik özellikler nedeniyle farklı alt gruplarda bulunan katılımcıların testte bulunan bir maddeyi doğru cevaplama olasılıklarının birbirinden farklı olması olarak ifade edilmektedir (Hambleton, Swaminathan & Rogers, 1991). Bir başka ifadeyle ise DMF, ölçülmek istenen yetenek düzeylerinde eşitlenmiş farklı alt gruplarda bulunan bireylerin maddeyi doğru cevaplama olasılıklarındaki farklılaşmadır (Clauser & Mazor, 1998).

Bir maddenin DMF göstermesinin iki sebebi bulunmaktadır. Bunlar; farklı alt gruplarda yer alan bireyler arasındaki gerçek yetenek farkı ya da madde yanlılığıdır (Camilli & Shepard, 1994). DMF, yanlılık için şarttır ancak yanlılığın kesin kanıtı değildir. Yani madde yanlı ise DMF gösterir; ancak DMF gösteren her madde yanlıdır demek yeterli değildir. Çünkü maddenin farklı fonksiyonlaşması gerçek yetenek farkından da kaynaklanabilir. Madde yanlılığı kavramsal ve yoruma açıktır, DMF ise istatiksel bir yöntemdir. Bu yüzden DMF içeren maddenin yanlı olup olmadığına karar vermek için uzman kanısına başvurulmalıdır (Camilli & Shepard, 1994).

Şekil 1'de DMF göstermeyen bir maddenin grafiği verilmiştir. Eğer her iki grubun madde karakteristik eğrisi özdeş ya da özdeşe yakınsa maddenin DMF göstermediği söylenebilir. Şekilde görüldüğü üzere madde karakteristik eğrileri arasında çok küçük bir alan kalmıştır ve eğrilere ilişkin parametreler neredeyse eşit bir durumdadır. Bu nedenlerle grafiğe bakarak maddenin DMF göstermediği söylenebilir (Zumbo, 1999).



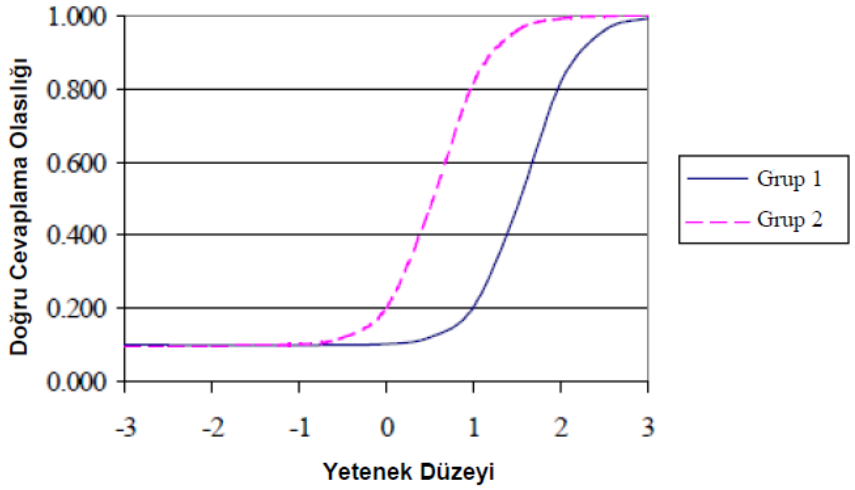
Şekil 1. DMF'li olmayan madde grafiği (Zumbo,1999, s.19).

Bir testte DMF'nin varlığını belirlemek için öncelikle karşılaştırılacak olan iki grup ölçülmek istenen yetenek düzeyinde eşitlenir. Gruplardan biri referans diğeri ise odak grup olarak belirlenir. Literatürde genellikle testin kolaylık sağladığı grup referans, dezavantaj sağladığı grup ise odak grup olarak belirlenmektedir. Daha sonra ise maddelere yönelik grup farklılıklarını belirlemek üzere istatiksel süreçler işe koşulur (Osterlind & Everson, 2009).

Bu kapsamda DMF, tek biçimli ve tek biçimli olmayan şeklinde iki grup olarak incelenmektedir.

1.2.1.Tek Biçimli DMF (Uniform)

Karşılaştırmak istediğimiz gruplardan biri, testte yer alan bir maddeye diğer gruba göre tüm yetenek seviyelerinde daha yüksek oranda doğru yanıtlama olasılığına sahipse tek biçimli DMF gerçekleşmiş olur (Swaminathan & Rogers, 1990). Yani bütün yetenek düzeylerinde tek bir grubun sabit biçimde hep daha avantajlı konumda yer almasını ifade eder.



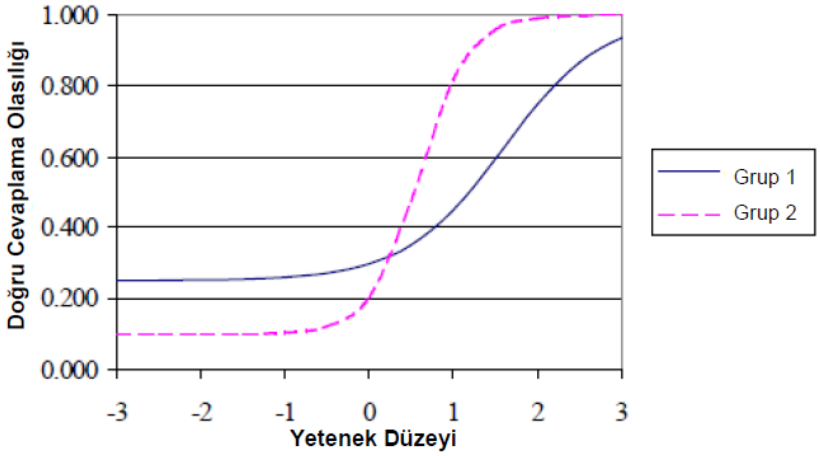
Şekil 2. Tek biçimli DMF gösteren madde grafiği (Zumbo,1999, s.20).

Şekil 2'de görüldüğü üzere maddeye doğru cevap verme olasılığı grafiğin her yerinde grup 2 için daha yüksek olmuştur. Yetenek düzeyleri farklılaşsa da maddeye yanıt verme olasılığı da aynı oranda bir grup lehine değişiklik göstermektedir. Ayrıca madde karakteristik eğrileri sınırları içerisinde yer alan DMF'nin derecesini gösterirken, madde karakteristik eğrilerinin birbirlerini kesmiyor oluşu ise tek biçimli DMF'yi işaret etmektedir (Zumbo, 1999).

Tek biçimli DMF gösteren maddelerin madde karakteristik eğrileri birbirleriyle kesişmezler. Kesişme durumunda ise bu; maddelerin kesişme noktasına kadar bir gruba, kesişmeden sonra ise diğer gruba avantaj sağladığı anlamına gelir ve buna tek biçimli olmayan DMF denir (Hambleton vd, 1991).

1.2.2.Tek Biçimli Olmayan DMF (Nonuniform)

Tek biçimli olmayan DMF, tek biçimli DMF'nin aksine, tüm yetenek düzeyleri boyunca aynı gruba kolaylık sağlamaz (Osterlind & Everson, 2009). Madde, belirli bir yetenek düzeyinde bir gruba kolaylık sağlarken, başka yetenek düzeylerinde ise başka bir gruba kolaylık sağlıyorsa bu tek biçimli olmayan DMF olarak adlandırılır (Swaminathan & Rogers, 1990). Kısacası bir grup, bazı yetenek düzeylerinde öteki gruba oranla daha yüksek doğru yanıtlama ihtimaline sahipken, bazılarında ise daha düşük doğru yanıtlama ihtimaline sahiptir.



*Şekil 3.*Tek biçimli olmayan DMF gösteren madde grafiği (Zumbo, 1999, s.21).

Şekil 3'te tek biçimli olmayan DMF gösteren maddenin, madde karakteristik eğrileri verilmiştir. Madde bir gruba tüm yetenek düzeylerinde tutarlı olarak aynı seviyede ve yönde avantaj sağlamamıştır (Osterlind & Everson, 2009). Madde karakteristik eğrileri sınırları içerisinde kalan alanın genişliği DMF'nin büyüklüğünü gösterirken, madde karakteristik eğrilerinin kesişmiş olmaları tek biçimli olmayan DMF'yi işaret etmektedir (Zumbo,1999).

1.3.DMF Belirleme Yöntemleri

Bir testte yer alan maddelerin DMF içerip içermediğini ortaya çıkarmayı sağlayan birçok yöntem bulunmaktadır. Mantel-Haenszel (Holland & Thayer, 1988), Lojistik Regresyon (Swaminathan & Rogers, 1990), SIBTEST (Shealy & Stout, 1993), Lord'un X^2 yöntemi (Lord, 1980), Raju'nun alan ölçümü (Raju, 1988; Raju, 1990) geliştirilen yöntemlerden bazılarıdır. DMF belirleme yöntemleri kullanılırken "odak" ve "referans" olarak isimlendirilen iki grubun maddelere verdikleri yanıtlar kıyaslanmaktadır. Gerçekleştirilen analizlerin ardından her bir madde için DMF bulunup bulunmadığı ve DMF varsa hangi düzeyde olduğuna yönelik sınıflandırma yapılmaktadır (Swanson vd., 2002). Ancak bu yöntemler, maddeleri DMF'li olarak belirlerken izledikleri farklı matematiksel yöntemler, kullandıkları farklı algoritmalar ve kesme noktaları nedeniyle tam bir uyum içinde değillerdir (Doğan & Öğretmen, 2008; Gök vd., 2010; Alatlı & Şenel, 2020; Bakan Kalaycıoğlu ve Berberoğlu, 2010). Bu nedenle, DMF'yi tespit etmek için güçlü bir istatistiksel yöntem üzerinde ortak bir fikre varılamamıştır. Bu nedenle yanlılık incelemelerinde birden fazla DMF belirleme yönteminden faydalanılması ve elde edilen

sonuçların birbiriyle kıyaslanarak karar verilmesi tavsiye edilmektedir (Hambleton, 2006). Bu çalışmada ise MH, LR ve SIBTEST olmak üzere üç farklı yöntem kullanılmıştır.

1.3.1.Mantel – Haenszel (MH)

Mantel – Haenszel yöntemi (MH), eşleştirilmiş gruplarla çalışmak üzere bir ki-kare testi olarak Nathan Mantel ve William Haenszel tarafından (1959) geliştirilmiş ve daha sonra Holand ve Thayer (1988) tarafından değişen madde fonksiyonu çalışmalarına uyarlanmıştır (Camilli & Shepard, 1994). Klasik test kuramı kapsamında, parametrik olmayan testlerde iki kategorili veriler için kullanılan MH istatistiği, yaygın olarak kullanılan DMF belirleme yöntemlerinden birisidir (Clauser & Mazor, 1998).

MH yöntemi, öğrencilerin maddelere verdikleri yanıt ile grup üyelikleri arasında, toplam test puanına bağlı olan bir ilişki bulunup bulunmadığını kontrol etmeyi hedefler (Magis, Beland, Teurlinckx & Boeck, 2010). MH yöntemi hesaplamaları karmaşık değildir ve çıkan sonuçları yorumlamak oldukça kolaydır. Bunun yanında, elde edilen DMF'nin büyüklüğü anlamlılık testi ile belirlenebilmektedir (Millsap, 2011). Ayrıca yöntemin büyük olmayan örneklemler (her grup için 200) üzerinde de kullanılabiliyor olması araştırmacılar tarafından sıkça tercih edilen bir yöntem olmasını sağlamıştır (Samuelsen, 2005; Clauser & Mazor, 1998). MH yönteminde, bireyler değişkenler temel alınarak referans ve odak grubu olmak üzere ikiye ayrılır ve böylece eşleştirilmiş gruplar oluşturulmuş olur. Referans ve odak grupları, bireylerin toplam test puanı üzerinden yetenek ya da yeterlik düzeylerine göre 4 ya da 5 yetenek seviyesine bölünürler (Osterlind & Everson, 2009). Aynı yetenek seviyesindeki bu iki grup ile madde arasında bir ilişki olmadığını belirten yokluk hipotezi test edilir (Gierl, Jodoin & Ackerman, 2000).

H_0: Aynı yetenek düzeyindeki bireyler ile test puanı arasında anlamlı düzeyde bir ilişki yoktur.

Grupların yetenek düzeylerine göre eşleştirilmesinin ardından her yetenek seviyesi ve testte yer alan her bir madde için bir olasılık tablosu oluşturulur. Tablo oluşturulurken bireylerin maddeye verdikleri doğru ve yanlış cevapların frekansları kullanılır (Gierl vd., 2000).

Tablo 1

Herhangi Bir Madde İçin İ Yetenek Düzeyindeki Bireylere Ait Olasılık Tablosu

Grup	Doğru	Yanlış	Toplam
R	A_i	B_i	N_{Ri}
O	C_i	D_i	N_{Oi}
Toplam	N_{Di}	N_{Yi}	$T_İ$

Tablo 1'de sırası ile A_i ve C_i, değerleri i yetenek düzeyindeki referans (R) ve odak (O) gruplarından maddeyi doğru cevaplayan kişi sayısını, B_i ve D_i değerleri maddeyi yanlış cevaplayan kişi sayısını göstermektedir. N_{Di} ve N_{Yi} değerleri sırasıyla referans ve odak gruplarında maddeyi doğru ve yanlış cevaplayan kişilerin toplamını, N_{Ri} ve N_{Oi} değerleri ise maddeye cevap veren toplam kişi sayısını göstermektedir (Holand & Thayer, 1988).

Olasılık tablosunun oluşturulmasının ardından, aynı yetenek seviyesinde olup farklı gruplardan gelen kişilerin maddeye doğru cevap verme olasılıklarını karşılaştırabilmek için olasılık oranı hesaplanır. α kestirim katsayısı adı verilen bu oran, referans gruptaki kişilerin maddeye doğru cevap verme olasılıklarının, odak gruptaki kişilerin maddeye doğru cevap verme olasılığına oranı şeklinde hesaplanır. α kestirim katsayısının hesaplanması aşağıdaki gibidir (Holand & Thayer, 1988; Clauser & Mazor, 1998).

$$\alpha_{MH} = \frac{\sum A_i D_i / T_i}{\sum B_i C_i / T_i}$$

Denklemden elde edilen α_{MH} değeri DMF etki büyüklüğünün bir kestiricisidir. Tek başına yorumlanması zor olan bu değer logaritmik dönüşümlerle daha kolay yorumlanabilecek farklı bir ölçeğe dönüştürülür. Doğal logaritması alınan α değeri -2,35 ile çarpılarak Δ_{MH} istatistiği hesaplanır (Clauser & Mazor, 1998).

$$\Delta_{MH} = -2{,}35 ln(\alpha MH)$$

Oluşturulan bu yeni katsayıda Δ_{MH} değerinin 0 olması DMF olmadığı anlamına gelmektedir. Eğer hesaplanan Δ_{MH} değeri pozitifse odak grup, negatifse referans grup lehine DMF bulunduğu tespit edilmiş olur (Holand & Thayer, 1988). Δ_{MH} değeri etki büyüklüğünü temsil ettiği için, aynı zamanda DMF içeren maddelerin düzeylerini yorumlamada da kullanılır. Zieky (1993) tarafından DMF miktarının büyüklüğünü yorumlamak için oluşturulan ve yaygın olarak kullanılan sınıflama ölçütü aşağıdaki gibidir.

Tablo 2

Δ_{MH} *Değerleri İçin Yorumlama Ölçütleri*

Değer Aralığı	DMF Düzeyi	Yorum
$\lvert \Delta_{MH} \rvert < 1$	A	DMF yoktur veya ihmal edilebilecek düzeydedir.
$1 < \lvert \Delta_{MH} \rvert < 1{,}5$	B	Orta düzeyde DMF vardır. Maddelerin gözden geçirilmesi gerekir.
$1{,}5 < \lvert \Delta_{MH} \rvert$	C	Yüksek düzeyde DMF vardır. Madde düzenlenmeli veya testten çıkarılmalıdır.

Tablo 2'de ifade edildiği üzere, Δ_{MH} değerinin mutlak değeri 1'den küçük olan maddeler A düzeyinde DMF içerir. Bu maddeler ihmal edilebilir düzeydedir ve testte yer almaya devam edebilir. Δ_{MH} değerinin mutlak değeri 1 ile 1,5 arasında yer alan maddeler B düzeyinde DMF içerir. Orta düzeyde DMF vardır ve maddeler üzerinde çalışılarak düzeltme yapılması gerekmektedir. Δ_{MH} değerinin mutlak değeri 1,5'den büyük olan maddeler C düzeyinde DMF içerir. Yüksek düzeyde DMF içeren bu maddelerin yerine başka madde eklenmeli ya da mutlaka testten çıkarılmalıdır (Wiberg, 2007).

MH yönteminin küçük örneklemler üzerinde kullanabilir olması, sonuçlarının kolaylıkla yorumlanabilmesi ve madde tepki kuramına dayalı yöntemlerle tutarlı sonuçlar sağlaması gibi avantajları bulunmaktadır. Bu nedenlerle sıkça kullanılan DMF belirleme yöntemlerinden biri olmuştur. Sadece tek biçimli DMF belirlemede kullanılabilmesi ise en büyük sınırlılığıdır (Bertrand & Boiteau, 2003).

1.3.2.Lojistik Regresyon (LR)

Lojistik regresyon (LR) analizinin DMF'li maddeleri belirlemek amacıyla uygulanması fikrini ilk olarak Swaminathan ve Rogers (1990) ileri sürmüştür (Camilli & Shepard, 1994). Bu öneri, bağımlı değişkenin kategorik olduğu, bağımsız değişkenin ise sürekli olduğu durumlarda kullanılmasına yönelik bir regresyon analiz türüdür (Osterlind & Everson, 2009). LR yöntemi kullanılarak hem tek biçimli hem de tek biçimli olmayan DMF tespit edilebilmektedir (Swaminathan ve Rogers, 1990). Tek biçimli DMF'yi belirlerken MH ve SIBTEST yöntemlerine, tek biçimli olmayan DMF'yi belirlerkenSIBTEST yöntemine kıyasla daha güçlü sonuçlar ortaya çıkarmaktadır (Li &

Stout, 1996; Rogers & Swaminathan, 1993; Swaminathan & Rogers, 1990). LR yöntemi kullanılarak yapılan DMF analizlerinde toplam puan, grup üyeliği ve toplam puan ile grup üyeliği etkileşimi kullanılan değişkenlerdendir. Maddeye verilen cevap (doğru=1, yanlış=0) iki kategorili bağımlı değişken, testten alınan toplam puan (X), grup değişkeni (G; referans grup=1, odak grup=0) ve grup değişkeni ile toplam puanın ortak etkileşimi bağımsız değişkenler olacak şekilde bir eşitlik kurulur (Zumbo, 1999). LR eşitliği;

$$Y = b_0 + b_1X + b_2G + b_3X*G$$

şeklinde ifade edilmektedir. Eşitlikte yer alan b_0 değeri kesim noktası, b_1 değeri toplam puanın madde üzerindeki etkisi, b_2 değeri ait olduğu grubun madde üzerindeki etkisi, b_3 değeri ise ait olduğu grubu ile toplam puanın madde üzerindeki etkisini ifade etmektedir. Analizi yapılan madde için b_2 ve b_3 değerlerin 0'a eşit olması maddenin DMF içermediği, sadece b_3 değerinin 0'a eşit olması maddenin tek biçimli DMF içerdiği, b değerlerinden hiçbirinin 0'a eşit olmaması ise maddenin tek biçimli olmayan DMF içerdiği anlamına gelmektedir. Ayrıca tek biçimli ve tek biçimli olmayan DMF belirlenirken serbestlik derecesi iki olan X^2 testi kullanılır (Penfield & Camilli, 2007; Swaminathan & Rogers, 1990).

Model karşılaştırmasına dayalı olan LR yönteminde üç aşamalı modelleme söz konusudur. Bu modeller bağımsız değişken olan toplam puan, grup ve toplam puan ile grup etkileşimi değişkenlerinin hiyerarşik olarak modele eklenmesiyle oluşturulur. Sonrasında modellerin uyumunu istatiksel olarak değerlendirmek için X^2 testi yapılır (Gierl vd., 2000; Zumbo,1999).

Model 1: $Y = b_0 + b_1X$

Model 2: $Y = b_0 + b_1X + b_2G$

Model 3: $Y = b_0 + b_1X + b_2G + b_3X*$G

Model 1'de toplam puan değişkeni bulunurken Model 2'de grup değişkeni eklenir. Bir sonraki aşama model 3'te ise daha önce eklenen değişkenlere ilaveten toplam puan ve grup etkileşimi değişkeni eklenir. Her modelde ki-kare değerinin hesaplanmasının yanı sıra standartlaştırılmış regresyon katsayısı olan R^2 değeri de hesaplanır. Bu değer etki büyüklüğü kestiricisi olarak kabul edilir ve DMF miktarını temsil eder. Etki büyüklüğü miktarı, iki model arasındaki R^2 farkı (ΔR^2) hesaplanarak bulunur. İkinci ve birinci modelden elde edilen ΔR^2 değerinin anlamlı olması tek biçimli DMF'nin, üçüncü ve ikinci modelden elde edilen ΔR^2 değerinin anlamlı olması ise tek biçimli olmayan DMF'nin varlığını işaret eder (Zumbo, 1999).

Zumbo ve Thomas (1996) ile Jodoin ve Gierl (2001), DMF düzeyinin belirlenmesi amacıyla etki büyüklüğü miktarının (ΔR^2) yorumlanması için Tablo 3'te verilen iki farklı sınıflamayı önermiştir.

Tablo 3

ΔR^2 Değerleri İçin Yorumlama Ölçütleri

DMF Düzeyi	Zumbo ve Thomas (1996) Değer Aralığı	Jodoin ve Gierl (2001) Değer Aralığı
A (Yok veya ihmal edilebilir düzeyde)	$\Delta R^2 < 0,13$	$\Delta R^2 < 0,035$
B (Orta düzeyde)	$0,13 \leq \Delta R^2 < 0,26$	$0,035 \leq \Delta R^2 < 0,070$
C (Yüksek Düzeyde)	$\Delta R^2 \geq 0,26$	$\Delta R^2 \geq 0,070$

LR yöntemi, hesaplama ve yorumlama kolaylığı bakımından MH yöntemiyle benzerlik göstermektedir. Ancak LR yönteminin hem tek biçimli hem de tek biçimli olmayan DMF'yi belirleyebilmesi bir avantaj sağlamaktadır (Swaminathan & Rogers, 1990). Tutarlı sonuçlar vermek adına MH yöntemine göre daha büyük örneklemler gerektiriyor olması ise en büyük sınırlılığıdır (Zumbo, 1999).

1.3.3.SIBTEST (Simultaneous Item Bias Test)

SIBTEST (Simultaneous Item Bias Test), iki kategorili verilerde DMF'yi ortaya çıkarmak için Shealy ve Stout (1993) tarafından geliştirilen bir modeldir (French & Finch, 2015; Osterlind & Everson, 2009; Potenza & Dorans, 1995). Öncesi Dorans ve Kulick (1986) tarafından geliştirilen Standartlaştırma yöntemine dayanmakla birlikte önemli bir farklılık içermektedir. Gözlenen toplam puanlara, regresyona dayalı bir düzeltme uygulanarak gerçek puanları kestirir ve odak ve referans grupları belirlerken bu puanlardan faydalanır. Böylece yöntem, 1. tip hatayı kontrol etmede kullanışlı hale gelir ve MH gibi benzer diğer DMF belirleme yöntemlerine göre odak ve referans grup eşleştirmesi bu yöntemde daha doğru bir şekilde yapılmış olur (Clauser & Mazor, 1998; Gierl, Gotzmann & Boughton, 2004; Abbott, 2007; Osterlind & Everson, 2009).

SIBTEST yöntemi, DMF belirlemede MH yöntemiyle benzer bir performans göstermesine rağmen ayırt ediciliği yüksek olan maddelerin analizinde 1. tip hatayı daha iyi kontrol eder. Bunun yanında SIBTEST, MH yöntemine kıyasla daha küçük örneklemlerde etkili olabilirken, odak ve referans gruplarının örneklem büyüklüğünün çok farklı olduğu

durumlarda MH yöntemi daha iyi performans sergilemektedir (Narayanan & Swaminathan, 1994; Roussos & Stout, 1996).

LR yönteminde olduğu gibi tek biçimli ve tek biçimli olmayan DMF'nin belirlenmesinde kullanılabilmektedir (French & Finch, 2015; Clauser & Mazor, 1998). DMF'yi hem madde hem de test düzeyinde belirleyebilen bir yöntemdir (Osterlind & Everson, 2009).

SIBTEST yönteminde öncelikle testte yer alan maddeler geçerli alt test ve tahmin edilen alt test olmak üzere iki alt gruba ayrılır. Geçerli alt test, varsayılan yetenek düzeyine uygun maddelerin ölçümünü içerirken tahmin edilen alt test, DMF'nin belirlenmesinde kullanılan maddeleri içerir. Geçerli alt testte, aynı yetenek grubundaki bireylerin puanları diğer grupla eşleştirmede kullanılır (Bolt, 2000).

Her bir madde için istatiksel hipotezler SIBTEST ile test edilir;

$H_0: B(T) = P_R(T) - P_F(T) = 0$ veya $H_0: \beta_{UNI} = 0$

$H_1: B(T) = P_R(T) - P_F(T) \neq 0$ veya $H_1: \beta_{UNI} \neq 0$

$B(T)$: Referans ve odak gruplarında gerçek puanları (T) temel alınarak eşleştirilmiş bireylerin maddeyi doğru cevaplama olasılığı arasındaki fark,

$P_R(T)$: Referans grubundaki bireylerin maddeyi doğru cevaplama olasılığı,

$P_F(T)$: Odak grubundaki bireylerin maddeyi doğru cevaplama olasılığı olarak tanımlanmaktadır (Gierl, Khaliq & Boughton, 1999).

H_0 hipotezi, maddeyi doğru cevaplama olasılığı açısından referans ve odak grupları arasında fark olmadığını, H_1 hipotezi ise fark olduğunu iddia etmektedir. β_{UNI} değeri ise etki büyüklüğünü belirtmektedir ve şu şekilde tanımlanır;

$$\beta_{UNI} = \int \beta(\theta) f_F(\theta) d(\theta)$$

$\beta(\theta)$: Referans ve odak gruplarındaki bireylerin maddeye doğru cevaplama olasılıkları

$f_F(\theta)$: Odak grubun maddeyi doğru cevaplama olasılığı için yoğunluk fonksiyonu,

$d(\theta)$: θ değerinin referans ve odak gruplarda farklılaşması olarak tanımlanmaktadır (Zhou, Gierl & Tan, 2006).

SIBTEST analizi sonucunda elde edilen β_{UNI} değeri her madde için tek tek hesaplanır. Elde edilen değer, DMF miktarı olarak yorumlanır. Bu değer eğer pozitifse maddenin referans gruba avantaj sağladığı, negatif ise odak gruba avantaj sağladığı söylenir. DMF'nin düzeyini belirlenmek için kullanılan ölçütler ise Roussos ve Stout (1996) tarafından önerilmiştir.

Tablo 4

β_{UNI} Değerlerini Yorumlama Ölçütleri

Değer Aralığı	DMF Düzeyi	Yorum
$\beta_{UNI} < 0{,}059$	A	DMF yoktur ya da ihmal edilebilir düzeydedir.
$0{,}059 \leq \beta_{UNI} < 0{,}088$	B	Orta düzeyde DMF vardır.
$0{,}088 \leq \beta_{UNI}$	C	Yüksek düzeyde DMF vardır.

Tablo 4'te ifade edildiği üzere $\boldsymbol{\beta_{UNI}}$ değeri 0,059'dan küçük olduğu durumda madde A düzeyinde DMF içerir. Bu madde ihmal edilebilir düzeydedir ve testte yer almaya devam edebilir. $\boldsymbol{\beta_{UNI}}$ değeri 0,059 ile 0,088 arasında yer aldığında madde B düzeyinde DMF içerir. Orta düzeyde DMF vardır ve madde üzerinde çalışılarak düzeltme yapılması gerekmektedir. $\boldsymbol{\beta_{UNI}}$ değeri 0,088'den büyük olduğu durumda ise madde C düzeyinde DMF içerir. Yüksek düzeyde DMF içeren bu maddenin yerine başka madde eklenmeli ya da mutlaka testten çıkarılmalıdır (Roussos & Stout, 1996).

1.4.PISA

Uluslararası Öğrenci Değerlendirme Programı (PISA), Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından geliştirilen ve uluslararası ölçekte üçer yıllık periyotlarla uygulanan bir izleme araştırmasıdır. 15 yaş grubunda örgün eğitime devam eden öğrencilerin, matematik okuryazarlığı, fen okuryazarlığı ve okuma becerilerini ölçmektedir. Gerçekleştirilen her PISA uygulamasında bu üç alandan birine ağırlık verilmektedir. Ağırlık verilen alana ilişkin yeni sorular hazırlanmakta ve derinlemesine analizler yapılmaktadır. PISA 2018 araştırmasında ise Okuma Becerileri alanına ağırlık verilmiştir (MEB, 2019).



Şekil 4. PISA döngülerinde temel ve ağırlıklı alanlar.[1] MEB. (2019). *PISA 2018 Türkiye Ön Raporu,* (Rapor No:10), Aralık 2019.

PISA 2018, diğer yıllarda uygulanan PISA araştırmalarına göre katılımın en fazla olduğu uygulama olmuştur. Uygulamaya, 79 ülkeden yarım milyonu geçkin öğrenci katılmıştır. Katılımcı ülke ve ekonomilerden 37'si OECD üyesidir. Bu ülkelerden elde edilen ortalama değerler, PISA sonuçlarını değerlendirirken karşılaştırma tablolarında referans noktası olarak kullanılmaktadır. Türkiye, PISA uygulamalarına 2003 yılından itibaren katılmaktadır. Katılımcı ülke ve ekonomilerden 70'i PISA 2018 uygulamasını bilgisayar tabanlı, 9'u ise kâğıt - kalem testi olarak uygulamıştır. Türkiye'de ise 2015 yılından bu yana PISA uygulaması bilgisayar tabanlı olarak gerçekleştirilmektedir (MEB, 2019).

PISA'dan elde edilen sonuçlara göre, ülkelerin ve ekonomilerin performansları değerlendirilmektedir. Her ülke ve ekonominin, ölçülen üç alanda diğer ülkeler ve OECD ortalamasına göre performans düzeyleri, önceki PISA uygulamalarına göre performanslarındaki değişimler, cinsiyete, okul türüne, bölgeye ve sosyo-ekonomik düzeye göre performansta görülen farklılıklar incelenmektedir (MEB, 2019).

PISA'nın temel amacı, öğrencilerin okulda öğrendikleri bilgi ve becerilerini günlük yaşama ilişkin problem durumlarında kullanma kabiliyetlerini ölçmektedir. Matematik okuryazarlığı, fen okuryazarlığı ve okuma becerileri alanlarının dışında, öğrencilerin motivasyonları, öğrenme alışkanlıkları, kendileri hakkında görüşleri, okul ortamları ve aileleriyle ilgili demografik, eğitimsel ve kültürel bilgilerin toplandığı geniş çaplı bir anket de uygulanmaktadır (Milli Eğitim Bakanlığı [MEB], 2017).

Orijinal olarak İngilizce ve Fransızca dillerinde hazırlanan ölçme araçları çeviri işlemlerinin yapılması için katılımcı ülkelere gönderilmektedir. Ülkeler bu iki versiyonu kullanarak kendi ulusal dillerinde iki çeviri yapmaktadır. Daha sonra ise bunları tek bir çeviriye dönüştürmektedirler. Çeviri işlemleri yapılırken ülkelerden, PISA çeviri ve uyarlama yönergelerine (OECD, 2007) uymaları istenmektedir. Çeviri işleminin ardından

[1] Ağırlıklı alanlar italik yazılmıştır.

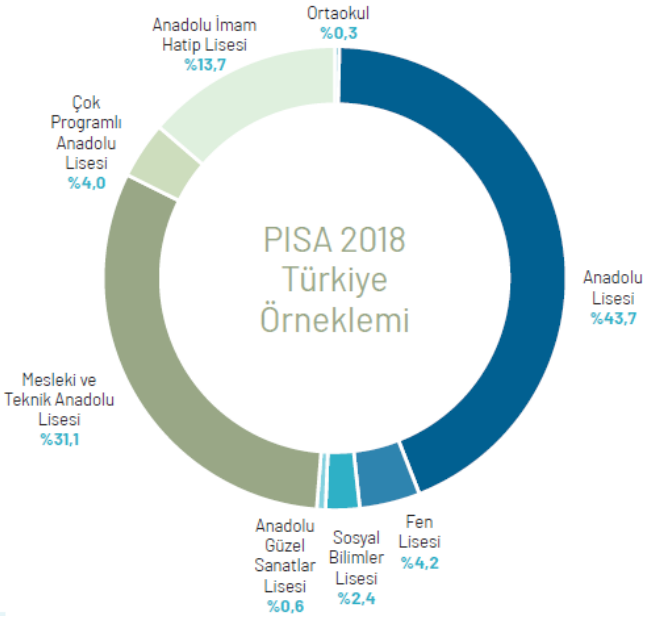
ölçme aracına, profesyonel çevirmen olan uzmanlarca son hali verilmektedir (OECD, 2007).

Türkiye, PISA 2018'e 186 okul ve 6890 öğrenci ile katılmıştır. 15 yaş grubundaki öğrencileri temsil edecek bu örneklem, seçkisiz örnekleme yöntemiyle seçilmektedir. İstatistiki Bölge Birimleri Sınıflaması (İBBS) Düzey 1'e göre 12 bölgeden okul belirlenmiştir. Seçilen okullarda uygulamaya katılacak öğrenciler ise seçkisiz olarak belirlenmektedir. PISA 2018 Türkiye örnekleminde yer alan öğrencilerin bölgelere göre katılım oranları şekilde verilmiştir. En fazla öğrencinin katıldığı bölge %20,2 ile İstanbul bölgesi olurken en az öğrencinin katıldığı bölge ise %1,6 ile Batı Anadolu Bölgesi olmuştur (MEB, 2019).



Şekil 5. Türkiye örnekleminin İBBS düzey 1 bölgelerine dağılımı. MEB. (2019). *PISA 2018 Türkiye Ön Raporu*, (Rapor No:10), Aralık 2019.

PISA 2018 Türkiye örneklemine göre okul türleri dağılımı aşağıdaki grafikte verilmiştir. Bu grafiğe göre temsil edilen 15 yaş grubundaki öğrencilerin %43,7'si Anadolu Lisesi, %31,1'i Mesleki ve Teknik Anadolu Lisesi, %13,7'si Anadolu İmam Hatip Lisesi, %4,2'si Fen Lisesi, %4'ü Çok Programlı Anadolu Lisesi, %2,4'ü Sosyal Bilimler Lisesi, %0,6'sı Anadolu Güzel Sanatlar Lisesinde eğitim görürken, %0,3'ü ise eğitimine ortaokulda devam etmektedir (MEB, 2019).



Şekil 6. PISA 2018 Türkiye örnekleminin okul türlerine göre dağılımı. MEB. (2019). *PISA 2018 Türkiye Ön Raporu,* (Rapor No:10), Aralık 2019.

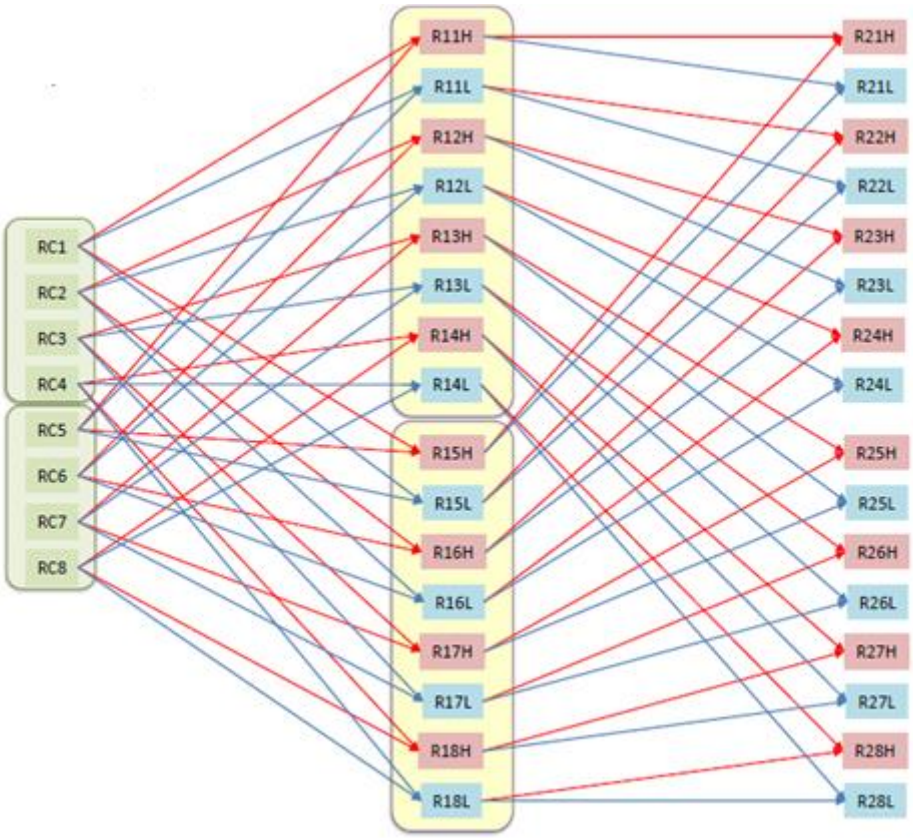
PISA 2018 Türkiye örnekleminde uygulamaya katılan öğrencilerin %49,6'sını kız öğrenciler %50,4'ünü ise erkek öğrenciler oluşturmaktadır. Sınıf düzeyinde bakıldığında ise uygulamaya katılanların %78,8'ini 10.sınıf, %17,7'sini 9.sınıf, %2,9'kunun 11.sınıf öğrencisi olduğu görülmüştür (MEB, 2019).

PISA 2018 uygulamasında diğer yıllarda yapılan uygulamalardan farklı olarak bireyselleştirilmiş test deseni (adaptive testing) kullanılmıştır. PISA 2015 ve daha önceki uygulamalarda maddeler sabit bir konuma sahipken PISA 2018'de ise bireyselleştirilmiş test deseni kullanılması nedeniyle maddeler dinamik bir yapıya sahiptir (OECD, 2019).

Bireyselleştirilmiş test deseninde, öğrencilerin önceki maddelere vermiş olduğu yanıtlara göre bir sonraki adımda farklı maddelerle karşılaşmaları sağlanmıştır. Bu doğrultuda maddeler üç bölümden oluşmaktadır: Temel Bölüm (Core), 1. Aşama (Stage 1), 2. Aşama (Stage 2). Temel bölümde yer alan madde paketleri benzer zorluk düzeyinde olacak şekilde hazırlanmıştır. 1. Aşama ve 2. Aşamada hazırlanan madde paketleri ise kısmen kolay ve kısmen zor olacak şekilde iki farklı düzeyde hazırlanmıştır.

PISA 2018 uygulamasına katılan öğrenciler öncelikle Temel bölümde yer alan maddelere yanıt vermişlerdir. Bu bölümde verdikleri doğru yanıt sayısına göre öğrenci başarıları düşük, orta ve yüksek olarak gruplandırılmıştır. Öğrencinin 1. Aşamada karşılaşacağı madde paketlerinin zorluk düzeyi, Temel bölümde verdiği doğru yanıt sayısına göre değişkenlik göstermektedir. Temel bölümde orta düzeyde başarı gösteren bir öğrencinin, 1. Aşamada kolay ya da zor düzeydeki madde paketleriyle karşılaşma ihtimali eşittir. Temel

bölümde yüksek başarı gösteren öğrencinin 1. Aşamada zor düzeydeki madde paketleriyle karşılaşma ihtimali %90 iken kolay düzeydeki madde paketleriyle karşılaşma ihtimali %10'dur. Temel bölümde düşük başarı gösteren bir öğrencinin 1. Aşamada kolay sorularla karşılaşma ihtimali %90 iken, zor düzeydeki sorularla karşılaşma ihtimali ise %10'dur. Öğrencilerin 2. Aşamada karşılaşacağı madde paketleri ise hem Temel bölüm hem de 1. Aşamada verdiği toplam doğru yanıt sayısına göre değişmektedir. Temel, 1. Aşama ve 2. Aşama bölümlerini birbirine bağlayan yönlendirme yolları Şekil 8'de yer almaktadır (OECD, 2019).



Şekil 7. Temel, 1.aşama ve 2.aşama bölümlerini birbirine bağlayan yönlendirme yolları. OECD (2019). *PISA 2018 Technical Report Chapter* , PISA:OECD

Şekil 8'de ilk sütun Temel bölümdekimadde paketlerini, ikinci sütun 1. Aşamadaki madde paketlerini, üçüncü sütun ise 2. Aşamadaki madde paketlerini göstermektedir. İkinci ve üçüncü sütunda görülen "H" harfi madde paketinin kısmen zor düzeyde (High), "L" harfi ise madde paketinin kısmen kolay düzeyde (Low) bir madde paketi olduğunu ifade etmektedir. Sütunlar arasındaki oklar öğrencilerin doğru yanıt sayılarına göre hangi düzeydeki madde paketleriyle karşılaşabilecekerini işaret etmektedir. Örneğin; Temel bölümde RC1 madde paketini yanıtlayan bir öğrenci, doğru yanıt sayısına göre 1. Aşamada kısmen zor düzeyde karşılaşabileceği madde paketleri R11H ve R15H iken, Kolay düzeyde

karşılaşabileceği madde paketleri ise R11L ve R15L'dir. Öğrencinin 2. Aşamada karşılaşabileceği madde paketleri de bu şekilde oklarla gösterilmiştir.

2.Problem Cümlesi

PISA, TIMSS, PIRLS gibi büyük ölçekte uygulanan uluslararası çalışmalar, ülkelerin eğitim sistemlerini karşılaştırmak, değerlendirmek ve geliştirmek için geniş bir bakış açısı sağlamaktadır (Gök vd., 2014). Özellikle sonuçlarıyla ülkelerin eğitim sistemlerine yön veren, birçok ülkenin katıldığı uluslararası çapta yapılan sınavların sonuçlarından doğru yorumlar ve ülkeler arası kıyaslamalar yapabilmek için ölçülen yapının gruptan bağımsız olması gerekir. (Crocker & Algina, 1986; Frey & Bernhardt, 2012). Ölçülecek yapının gruptan bağımsız olması koşulu sağlanmazsa yapılacak karşılaştırmalarda ortaya çıkan farklılık, gerçek durumdan mı yoksa yapının gruptan gruba farklılık göstermesinden mi kaynaklandığı bilenemeyecektir (Somer vd., 2009).

Bir testin belirli özelliklere sahip bir gruba sistematik olarak avantaj ya da dezavantaj sağlaması durumu yanlılık olarak ifade edilir (Camilli & Shepard, 1994; Clauser & Mazor, 1998; Zumbo, 1999; Osterlind & Everson, 2009). Bir maddenin yanlı olduğunu ileri sürebilmek için öncelikle o maddenin DMF içermesi beklenir. Yanlılık araştırmalarında yapılacak ilk iş, testte yer alan maddelere istatiksel analizler uygulayarak DMF içerip içermediklerinin tespit edilmesidir (Camilli ve Shepard, 1994; Zumbo, 1999; Zumbo, 2007).

Ülkemizin de katıldığı geniş ölçekli uygulamalardan biri olan PISA, birçok farklı kültür ve dil gruplarını katılımcı olarak barındırmaktadır (MEB, 2019). Bu denli farklı kültür ve dil gruplarını katılımcı olarak barındıran bir uygulamadan doğru çıkarımlar yapabilmek için uygulanan testin yanlılık içermediğinden emin olunmalıdır (Finch & French, 2013, French & Finch, 2015; Li & Zumbo, 2009). Bu kapsamda bu çalışmada aşağıdaki araştırma sorularına yanıt aranmıştır;

PISA 2018 uygulamasında yer alan okuma becerileri alt testindeki maddeler farklı kültür ve dil değişkenine göre değişen madde fonksiyonu göstermekte midir?

2.1.Alt Problemler

1. PISA 2018 okuma becerileri alt testinde yer alan maddeler, Türkiye, Birleşik Krallık ve Fransa örneği için MH, LR ve SIBTEST yöntemleriyle analiz edildiğinde DMF göstermekte midir?

   1.a. PISA 2018 okuma becerileri alt testinde yer alan maddeler, Türkiye ve Birleşik Krallık örneklemi için MH, LR ve SIBTEST yöntemleriyle analiz edildiğinde DMF göstermekte midir?

   1.b. PISA 2018 okuma becerileri alt testinde yer alan maddeler, Türkiye ve Fransa örneklemi için MH, LR ve SIBTEST yöntemleriyle analiz edildiğinde DMF göstermekte midir?

   1.c. PISA 2018 okuma becerileri alt testinde yer alan maddeler, Birleşik Krallık ve Fransa örneklemi için MH, LR ve SIBTEST yöntemleriyle analiz edildiğinde DMF göstermekte midir?
2. MH, LR ve SIBTEST yöntemleriyle yapılan analizlerin sonuçları birbiriyle uyum göstermekte midir?

3.Araştırmanın Önemi

Nitelikli işgücü ve kalitenin iyileştirilmesine yönelik artan talepler, insan özelliklerinin çeşitli yönlerinin kültürler arası değerlendirilmesi, geçtiğimiz on yıl içinde tüm dünyada daha önemli hale gelmiştir (Wu & Ercikan, 2006). İnsan özelliklerinin, farklı dil grupları ve kültürler arasında karşılaştırılmasına yönelik artan bu ilgi, uluslararası eğitimsel değerlendirme uygulamalarını da daha önemli hale getirmiştir (Yıldırım & Berberoğlu, 2009).

Uluslararası karşılaştırmalı eğitim çalışmalarında, aynı testin farklı dil versiyonları kullanılmaktadır. Bu versiyonlar arasında test uyarlama ve karşılaştırma sorunlarının yaşanması, ülkeler arasında öğrenci performansının ne kadar adil karşılaştırıldığı konusuna dikkatleri çekmiştir (Hambleton, 1994; Sireci & Berberoğlu, 2000; Yıldırım & Berberoğlu,2009).

Uluslararası eğitimsel değerlendirme çalışmalarının en önemli özelliği, çok dilli ve kültürler arası olmasıdır (OECD, 2007). Testler ya da anketler farklı dillere çevrildiğinde yeni versiyonun eşdeğerliğinin sağlanması gerekmektedir. Aksi halde çevrilen yeni formun zorluk ya da aşinalık derecesi uygulamaya katılan bireyler için farklı olabilir, farklı bir

yapıyı ölçebilir (Hambleton, 1994; Sireci & Berberoğlu, 2000). Bu değerlendirme çalışmalarının farklı dil versiyonları arasındaki ölçüm farklılığı, genellikle bu tür kültürler arası karşılaştırmaların geçerliliğine yönelik tehditler oluşturmaktadır. Bu nedenle, farklı dillerde değerlendirmeye giren öğrenciler arasında geçerli karşılaştırmalar yapabilmek için bu değerlendirmelerin ölçme eşdeğerliğini sağlamak önemli bir konu haline gelmiştir (Gierl, 2000; Vandenberg, 2002; Van de Vijver & Tanzer, 2004; Huang, Wilson & Wang, 2016). Bu amaçla, araştırmacılar son zamanlarda kültürler arası karşılaştırmalara, özellikle de farklı dilleri kullanan ülkeler arasındaki başarı testlerinin sonuçlarının karşılaştırmasını geçersiz kılacak kaynakları belirlemeye yoğunlaşmışlardır. Bu çalışmalar, uyarlanan testlerin test puanlarının karşılaştırılmasını geçersiz kılacak olası etkilerin anlaşılmasına büyük katkı sağlamıştır (Wu & Ercikan, 2006).

Daha önceki araştırmalar, test puanı karşılaştırmasını geçersiz kılan birçok olası kaynağı belirlemiş olsa da, bu kaynakların çoğu test çevirisinin veya müfredat farklılıklarının etkilerine dayandırılmıştır (Berberoglu & Sireci, 1996; Ellis, 1989; Ercikan, 1998; Ercikan & McCreith, 2002; Gierl, Rogers & Klinger, 1999; Sireci & Berberoglu, 2000; Sireci, Fitzgerald & Xing, 1998; van de Vijver & Tanzer, 1998). Ancak, öğrencilerin kültürlerinin ya da geçmiş yaşantılarının etkisini araştıran araştırmalar sınırlı kalmıştır (Wu & Ercikan, 2006). Uluslararası karşılaştırmalarda, iki ülke arasındaki test puanları farkı, testin ölçmek istediği özellikten ilgisiz olarak farklı bazı kültürel özellikleri yansıtıyorsa, burada yapılacak karşılaştırmada bazı eşitsizlikler söz konusudur. İstenmeyen bu kültürel etkiler, kültürler arası karşılaştırmalarda önyargı ve yanlış yorumlama gibi etkiler meydana getirebilir. Örneğin, bir ülkenin devlet okullarının diğerlerinden daha iyi performans gösterdiği sonucuna varmadan önce, gerçek dışı ve haksız yorumlara yol açabilecek bu tür etkileri hesaba katmak gereklidir. Ne yazık ki, araştırmacılar bu kültürel önyargı kaynaklarını, araştırma zorluğu nedeniyle sıklıkla ihmal etmektedir (Wu & Ercikan, 2006).

Hem çeviri hem de madde içeriğinde kullanılan kültüre özgü materyallerde eşdeğerlik gerekliliği, süreci adil karşılaştırmalar için daha da karmaşık hale getirmiştir (Hui & Triandis, 1985; Hambleton & de Jong, 2003; Yıldırım & Berberoğlu, 2009). Bir testte DMF'li maddeye rastlanması durumsal bir süreçtir ve o testin içsel bir özelliği değildir, ancak herhangi bir kültürler arası karşılaştırmaya özgü olabilir (Yıldırım & Berberoğlu, 2009). Başka bir deyişle DMF, herhangi bir test maddesinin sabit bir özelliği olmadığı gibi, kültürel bir faktör de DMF'nin doğal bir kaynağı değildir (van de Vijver & Tanzer, 1998;

Yıldırım & Berberoğlu, 2009). Bir DMF kaynağının varlığı, kültürler arası karşılaştırma için belirlenen değişkenlere ve amaca dayanır (Wu & Ercikan, 2006).

Bu çalışmada, uluslararası eğitimsel değerlendirme çalışmalarından biri olan PISA 2018 uygulamasında yer alan Okuma Becerileri alt testinin farklı kültür ve dile göre değişen madde fonksiyonu gösterip göstermediğinin üç farklı yöntemle incelenmesi amaçlanmıştır.

PISA, sonuçlarıyla ülkelerin eğitim politikalarına yön veren uluslararası bir izleme araştırmasıdır. Uygulamaya katılan ülkelerin, kendi eğitim sistemlerini değerlendirmelerine, diğer ülkelerin eğitim sistemleri ile karşılaştırma yapmalarına imkân tanımaktadır. Ülkelerin eğitim konusunda ne düzeyde bulunduklarının, tamamlanması gereken eksikliklerin ve alınması gereken önlemlerin neler olduğunun belirlenmesine yardımcı olmaktadır (MEB, 2019). Sonuçlarına dayanarak önemli kararların alınmasına katkı sağlayan bu denli büyük ölçekli uluslararası bir araştırmada öğrencilerin gerçek düzeylerinin belirlenebilmesi için uygulanan ölçme araçlarının geçerli ve güvenilir olması gerekmektedir (Gierl, 2000; Van de Vijver & Tanzer, 2004; Vandenberg, 2002). Testlerde yanlı maddelerin bulunması, gruplar arasında sistematik hataya neden olarak ölçülmek istenen yapıya ilişkisiz değişkenler karıştırabilir. Yanlılığın bir göstergesi olan DMF'nin tespit edilmesi, yansız ve geçerli bir test ortaya çıkarmada önemli bir göreve sahiptir (Clauser & Mazor, 1998; Camilli & Shepard, 1994; Ercikan vd., 2004).

PISA uygulamasında ölçme araçları İngilizce ve Fransızca olarak hazırlanmakta daha sonrakatılımcı ülkelere çeviri işlemlerinin yapılması için gönderilmektedir (OECD, 2007). Bu nedenle bu çalışmada, kıyaslama yapabilmek adına PISA uygulamasını kendi ulusal dilinde alan Birleşik Krallık ve Fransa örneklemleri ile kendi ulusal dilinde bir çeviriye dönüştürerek uygulayan Türkiye örneklemi tercih edilmiştir.

Bu araştırmanın önemi, PISA 2018 uygulamasındaki okuma becerileri alt testinde bulunan maddelerin farklı kültür ve dil değişkenine bağımlı olup olmadıklarını saptayarak ilerideki PISA uygulamalarını geliştirici yönde katkılar sağlamaktır. Daha sonraki yıllarda yapılacak olan PISA uygulamalarında yer alacak maddeleri seçerken, kültürlerarası DMF'yi göz önüne alarak test sonuçlarının güvenirliğini ve geçerliğini artırıcı yönde bir etki yapacağı düşünülmektedir.

İlgili alan yazın incelendiğinde, DMF'li maddeleri saptamak için kullanılan yöntemlerin birbirine kıyasla avantaj ve dezavantajlarının bulunduğu görülmüştür. Her yöntemin temel aldığı istatiksel modeller veya kabul ettikleri varsayımlar bakımından diğer yöntemlere kıyasla üstünlüğü ya da sınırlılığı bulunabilmektedir. Bu nedenle sadece bir yöntem

kullanılarak yapılan araştırmalar yerine birden çok yöntemin kullanıldığı araştırmaların gerçekleştirilmesi tavsiye edilmektedir (Holland & Wainer, 1993; Osterlind & Everson, 2009). Bundan dolayı bu araştırmada DMF'li maddeleri belirlemede MH, LR ve SIBTEST olmak üzere üç farklı DMF belirleme yöntemi kullanılmıştır. Ayrıca yapılan analizlerin sonuçları karşılaştırılarak DMF'li maddeler bakımından yöntemler arasındaki benzerlik ya da farklılıkların ortaya konması amaçlanmıştır. Bu bakımdan DMF belirleme yöntemlerinin karşılaştırılmasına dayalı yapılan çalışmalara katkı sağlaması açısından önemli olduğu düşünülmektedir. Son olarak, 2006, 2009, 2012 ve 2015 yıllarında uygulanan PISA'ya ilişkin birçok DMF çalışması bulunurken 2018 yılında yapılan uygulamaya ilişkin herhangi bir DMF çalışmasına rastlanılmamıştır.

4. Sayıltılar

1. PISA 2018 uygulamasına katılan öğrencilerin okuma becerileri testindeki maddelere verdikleri cevapların gerçek durumlarını yansıttığı varsayılmıştır.
2. Okuma becerileri testinin tüm öğrencilere aynı koşullar altında uygulandığı varsayılmıştır.
3. Öğrencilerin, tüm maddeleri bilgi düzeylerini ortaya çıkaracak ciddiyette ve içtenlikte kopya çekmeden cevapladıkları varsayılmıştır.

5. Sınırlılıklar

1. Bu araştırmada yapılan DMF analizleri Mantel Haenszel, Lojistik Regresyon ve SIBTEST yöntemleri ile sınırlı tutulmuştur.
2. Araştırmada, PISA uygulamasına katılan her öğrencinin aynı soruyu görmemesi durumundan kaynaklı, DMF analizine alınan maddeler bir kitapçıkla sınırlı tutulmuştur.
3. Bu araştırmadaki kıyaslamalar PISA 2018'e katılım sağlayan Türkiye, Fransa ve Birleşik Krallık ülkelerinden seçilen örneklemlerle sınırlı tutulmuştur.

BÖLÜM II

İLGİLİ ARAŞTIRMALAR

Bu bölümde, PISA uygulamasına yönelik farklı kültür ve dil değişkenine dayalı olarak gerçekleştirilen değişen madde fonksiyonu analizleri yürüten çalışmalar ile farklı değişen madde fonksiyonu belirleme yöntemlerini kıyaslamayı amaçlayan çalışmalara yer verilmiştir. Bu çalışmalar, yurt içi, yurt dışı ve DMF belirleme yöntemlerinin karşılaştırılmasına yönelik olan çalışmalar olmak üzere üç başlıkta toplanmış ve araştırma yılına göre sıralanmıştır.

1.Yurt İçi Araştırmalar

Çet (2006) çalışmasında, PISA 2003 uygulaması Matematik alt testinde yer alan maddelerin Türkçe ve İngilizce formları arasındaki madde yanlılığını araştırmıştır. Yayınlanmış madde sayısının çoğunlukta olduğu iki kitapçık seçilmiş ve bu kitapçıklarda yer alan matematik alt testlerinin faktör yapıları faktör çözümlemesi yöntemleriyle belirlenmiştir. Faktör yapılarının belirlenmesinin ardından seçilen maddeler analiz edilmiştir. Tek boyutlu DMF analizleri ile çok boyutlu eşleştirme analizlerinden elde edilen sonuçlar Mantel Haenszel ve Lojistik Regresyon yöntemleri kullanılarak karşılaştırılmıştır. Bunun sonucunda çok boyutlu eşleştirme analizinden elde edilen sonuçlarda her iki kitapçıkta da yanlılık gösteren maddeler tespit edilmiştir. Yanlılık gösterdiği belirlenen maddelerde ölçtükleri bilişsel yeterlilikleri ve matematiksel beceriler, madde kökü, madde türü, görsel ve uzamsal unsurlar göz önüne alınarak madde yanlılığının nedenleri belirlenmeye çalışılmıştır. Yapılan niteliksel analizler sonucunda yanlılığın kaynakları müfredat farklılıkları, kültürel farklılıklar ve çeviriden kaynaklanan farklılıklar olarak üç başlık altında toplanmıştır. Türkiye ve Amerika alt gruplarında yer alan öğrencilerin eşit yeteneklerde olmalarına rağmen bazı maddelere doğru cevap verme olasılıklarının farklı olması ülkelerdeki matematik programlarının farklılığından ya da

İngilizce'den Türkçe'ye çeviri yapılırken maddelerindeki bazı nicelik bildiren kelimelerin anlamlarının değişime uğramasından kaynaklandığı görülmüştür.

Ulutaş (2012) çalışmasında, PISA 2006 uygulaması Fen Okuryazarlığı alt testinde bulunan maddelerin farklı kültür ve cinsiyet değişkenlerine dayalı olarak değişen madde fonksiyonu içerip içermediğini incelemiş ve maddelerin farklı fonksiyonlaşma nedenlerini ortaya koymayı amaçlamıştır. PISA uygulamasında kullanılan 13 adet kitapçıktan maddelerinin kamuoyuna açıklandığı 1 ve 5 numaralı kitapçıklar incelenmiş, Türkiye'den katılan 757öğrenci ile ABD'den katılan 856 öğrencinin bu kitapçıklara verdikleri cevaplar kullanılmıştır. Maddeler analiz edilirken Mantel-Haenszel (MH), Simultaneous Item Bias Test (SIBTEST) ve Madde Tepki Kuramı Olabilirlik Oran Analizi (MTK-OOA) tekniklerini kullanarak analizlerini yürütmüştür. Analiz sonuçlarına göre üç yöntemde de ortak olarak DMF gösteren maddeler olası yanlı maddeler olarak kabul edilmiştir. Kültürler arası değişkenine göre, 1 numaralı kitapçıkta 16, 5 numaralı kitapçıkta 24 adet DMF gösteren madde bulunmuştur. Maddelerden 15'i Türk öğrencilere avantaj sağlarken, 25 madde ise ABD'li öğrencilere avantaj sağlamıştır. Cinsiyet değişkenine göre ise sadece Mantel-Haenszel ve SIBTEST yöntemleri kullanılarak yapılan analizlerin sonucunda cinsiyete göre DMF gösteren madde bulunmuş ve bunların 4'ünün kızlar, 3'ünün ise erkekler lehine çalıştığı tespit edilmiştir. Olası yanlılık nedenlerini ortaya çıkarmak için uzman kanısına başvurulmuş ve bunun sonucunda ise Türkiye ve ABD'den katılan gruplara bazı maddelerin içerik ve beceriler bakımından avantaj veya dezavantaj sağlamış olabileceği görüşü uzman kişilerce ileri sürülmüştür.

Sırgancı (2012) çalışmasında, PISA 2006 uygulaması kapsamında öğrencilere uygulanan anketin kültür ve dil değişkenine göre değişen madde fonksiyonu gösterip göstermediğini incelemiştir. Araştırmanın çalışma grubunu benzer ve farklı kültürel - dilsel özelliklere sahip olan Avustralya, Yeni Zelanda, ABD ve Türkiye örneklemleri oluşturmaktadır. Çalışmaya Avustralya'dan 14.170, Yeni Zelanda'dan 482, ABD'den 5.605 ve Türkiye'den 4.942 öğrenci katılmıştır. DMF analizleri yürütülürken Sıralı Lojistik Regresyon ve Poly-SIBTEST yöntemleri kullanılmıştır. Analizler yapılırken Avustralya referans ülke alınmış ve diğer ülkeler ile aynı dil- benzer kültür (Avustralya- Yeni Zelanda), aynı dil-farklı kültür (Avustralya- ABD) ve farklı dil-farklı kültür (Avustralya- Türkiye) olarak ikişerli olarak gruplandırılmıştır. Yapılan analizler sonucunda, aynı dil- benzer kültür (Avustralya-Yeni Zelanda) örneklemlerine uygulanan SLR analizinde bir maddenin B düzeyinde, dört maddenin A düzeyinde, Poly-SIBTEST analizinde bir maddenin B düzeyinde, üç

maddenin A düzeyinde DMF içerdiği görülmektedir. Aynı dil-farklı kültür (Avustralya-ABD) örneklemlerine uygulanan SLR analizinde bir maddenin C düzeyinde, Poly-SIBTEST analizinde yedi maddenin A düzeyinde; bir maddenin C düzeyinde, iki maddenin B düzeyinde, dört tanesi ise A düzeyinde DMF göstermiştir. Farklı dil-farklı kültür (Avustralya- Türkiye) örnekleminde tüm maddeler SLR ve Poly-SIBTEST yöntemlerinde C düzeyinde DMF göstermiştir.

Atalay Kabasakal ve Kelecioğlu (2012) çalışmasında, PISA 2006 uygulaması öğrenci anketinde yer alan fen bilimlerine karşı tutum maddelerinin cinsiyet ve kültür değişkenine göre değişen madde fonksiyonu içerip içermediğini incelemiştir. Araştırmanın çalışma grubunu Türkiye'den katılan 4.942 öğrenci ile ABD'den katılan 5.611 öğrenci oluşturmaktadır. Değişen madde fonksiyonu analizleri yürütülürken Ordinal Lojistik Regresyon ve Poly – SIBTEST yöntemleri kullanılmıştır. Yapılan analizler sonucunda cinsiyet değişkenine göre Türkiye örnekleminde OLR yöntemi ile 12 maddede, Poly-SIBTEST yöntemi ile 13 maddede DMF gözlenmiştir. Kullanılan yöntemlerin uyumuna bakıldığında ise yöntemlerin DMF belirlemede bir uyum gösterdiği; ancak düzeylerinde farklılıklar olduğu belirlenmiştir. ABD örnekleminde cinsiyet değişkenine göre her iki yöntemde de 15 maddede DMF gözlenmiştir.. Kullanılan yöntemlerin uyumuna bakıldığında ise yöntemlerin DMF'li madde sayısı bakımından uyum gösterdiği; ancak DMF'li maddelerin düzeylerinde farklılıklar olduğu belirlenmiştir. Türkiye ve ABD örneklemleri arasında kültüre göre OLR ve Poly-SIBTEST yöntemleri kullanılarak yapılan analizlerde her iki yöntemde de 30 maddede DMF görülmüştür. Kullanılan yöntemlerin uyumuna bakıldığında ise yöntemlerin DMF'li madde sayısı bakımından uyum gösterdiği; ancak DMF'li maddelerin düzeylerinde farklılıklar olduğu belirlenmiştir.

Özmen (2013) çalışmasında, PISA 2009 uygulaması Okuma Becerileri alt testinde yer alan maddelerin Türkiye-ABD, Türkiye-Birleşik Krallık grupları arasında DMF içerip içermediğini inceleyip olası yanlılık nedenlerini ortaya koymayı amaçlamıştır. Maddeler analiz edilirken MH, SIBTEST, MTK temelli MTK-OO ve b parametrelerinin birbiriyle kıyaslanması teknikleri işe koşulmuştur. Okuma becerileri alt testine ait dört ve altı numaralı kitapçıklar incelenmiş ve bunları cevaplayan Türk grubunda 767, ABD grubunda 796 ve Birleşik Krallık grubunda 765 öğrencinin cevapları kullanılmıştır. Analiz sonucunda, Türk ve ABD'li, Türk ve Birleşik Krallık'taki öğrenci grupları arasında en az DMF'li madde altı numaralı kitapçıkta tespit edilirken, en fazla DMF'li madde ise dört numaralı kitapçıkta ortaya çıkmıştır. DMF içeren maddelerin, kelimelerin zor anlaşılması,

madde formatının bilindik olması, uyarlama hataları ve kültürel farklılıklardan kaynaklandığı kanısı uzmanlar tarafından ileri sürülmüştür. Kullanılan DMF belirleme yöntemleri karşılaştırıldığında, SIBTEST ve MTK temelli analizlerin birbirleriyle uyumlu sonuçlar ortaya çıkardığı, MH yönteminin ise diğerlerine göre daha az sayıda DMF'li madde belirlediği ortaya konmuştur.

Demir ve Köse (2014) araştırmasında, PISA 2009 uygulaması Matematik okuryazarlığı alt testinde bulunan maddelerin cinsiyet ve kültür değişkenlerine göre DMF içerip içermediğini belirlemeye yönelik bir çalışma ortaya koymuştur. Analizleri yaparken, Mantel Haenszel, Lojistik Regresyon ve SIBTEST yöntemlerini kullanmıştır. Araştırmanın çalışma grubu, cinsiyet ve kültür değişkeni için ayrı ayrı oluşturulmuştur. Cinsiyet değişkeni için çalışma grubuna PISA 2009 uygulamasında Türkiye'de 5. kitapçığın uygulandığı 332 öğrencinin matematik okuryazarlığı alt testindeki maddelere verdikleri yanıtlar kullanılırken, kültür değişkeni için ise Türkiye, Almanya, Finlandiya ve ABD'den katılan öğrencilerin 5. kitaptaki maddelere verdikleri cevaplar kullanılmıştır. Cinsiyete göre yapılan DMF incelemelerine bakıldığında MH tekniği kullanılarak yapılan analizde iki maddede DMF bulunurken, LR tekniğinde üç, SIBTEST tekniğinde ise dört maddede DMF bulunmuştur. Toplamda beş DMF'li maddenin dördüü erkeklere, biri ise kızlara avantaj sağlamıştır. Kültür değişkenine göre bakıldığında, MH tekniği kullanılarak yapılan analizde 10 madde DMF'li olarak belirlenmiş ve bu maddelerin beşi Almanya'dan katılan öğrencilere, diğer beşi maddenin ise Türkiye'den katılan öğrencilere avantaj sağladığı görülmüştür. LR tekniği kullanılarak yapılan analizde 10 maddede DMF tespit edilmiş, bu maddelerin yedi tanesi Almanya'dan katılan öğrencilerin lehine olurken üç tanesi ise Türkiye'den katılan öğrencilerin lehine olduğu görülmüştür. SIBTEST tekniği kullanılarak yapılan analizde toplamda 12 DMF'li madde bulunmuş ve bu maddelerin yedi tanesi Almanya'dan katılan öğrencilerin lehine olurken beş tanesi ise Türkiye'den katılan öğrencilerin lehine olduğu görülmüştür. Türkiye ile Finlandiya'dan katılan öğrencilerin yanıtları incelendiğinde MH tekniği kullanılarak yapılan analizde 12 maddede DMF belirlenmiş, bu maddelerden altısı Finlandiya'dan katılan öğrencilerin, altısı ise Türkiye'den katılan öğrencilerin lehine olmuştur. LR tekniği kullanılarak yapılan analizde toplamda 10 adet maddede DMF olduğu belirlenmiş, bu maddelerin hepsinin Finlandiya'dan katılan öğrencilere avantaj sağladığı görülmüştür. SIBTEST tekniği kullanılarak yapılan analizde toplamda 13 adet maddenin DMF'li olduğu belirlenmiş ve bu maddelerden yedisinin Finlandiya'dan katılan öğrencilerin lehine, altısının ise Türkiye'den

katılan öğrencilerin lehine çalıştığı görülmüştür. Türkiye ile ABD'den katılan öğrencilerin yanıtları incelendiğinde MH tekniği kullanılarak yapılan analizde toplamda 11 maddede DMF belirlenmiş, bu maddelerin beş tanesi ABD'den katılan öğrencilerin lehine, altı tanesi ise Türkiye'den katılan öğrencilerin lehine çalıştığı görülmüştür. LR tekniği kullanılarak yapılan analizde toplamda dokuz maddede DMF bulunmuş, bu maddelerin dördü ABD'den katılan öğrencilere avantaj sağlarken, beşi ise Türkiye'den katılan öğrencilere avantaj sağlamıştır. SIBTEST tekniği kullanılarak yapılan analizde toplamda 13 maddede DMF bulunmuş ve bu maddelerin sekizi ABD'den katılan öğrencilerin lehine çalışırken beşi ise Türkiye'den katılan öğrencilerin lehine çalıştığı tespit edilmiştir.

Gök, Kabasakal ve Kelecioğlu (2014) çalışmasında, PISA 2009 kapsamında öğrencilere uygulanan 11 maddelik okumaya yönelik tutum anketinin kültür ve dil değişkenine göre değişen madde fonksiyonu içerip içermediğini incelemiştir. DMF analizleri yürütülürken poly-SIBTES, Ordinal Lojistik Regresyon ve madde tepki kuramı olabilirlik oranı yöntemleri kullanılmıştır. Araştırmanın çalışma grubunu Türkiye, Finlandiya, Rusya, Avustralya, Yeni Zelanda ve Singapur örneklemlerinden katılan öğrencilerin yanıtları ele alınmıştır. DMF analizleri yürütülürken çalışma grubunda yer alan ülkeler ikişerli olarak gruplandırılmıştır. PISA 2009 uygulamasında 39. sırada yer alan Türkiye ile 3. Sırada yer alan Finlandiya farklı başarı düzeylerine ve farklı kültür ve dile sahip olmaları nedeniyle karşılaştırılmıştır. Bunun yanında Türkiye başarı düzeyi benzer ancak farklı kültür ve dil özelliklerine sahip olan 41. Sırada yer alan Rusya ile karşılaştırılmıştır. Bunlara ek olarak başarı düzeyleri yüksek ve benzer kültür ve dil özelliklerine sahip olan 7.sıradaki Avustralya ile 9.sıradaki Yeni Zelanda ve başarı düzeyleri yüksek ve farklı kültür ve aynı dil özelliklerine sahip olan 5.sıradaki Singapur ile 7.sıradaki Yeni Zelanda ikişerli gruplar halinde karşılaştırılmıştır. DMF analizlerine başlamadan önce okumaya yönelik tutum anketinin faktör yapısı açımlayıcı faktör analizi ile incelenmiş ve ölçeğin tek boyutta toplandığı ortaya çıkmıştır. Yapılan DMF analizleri sonucunda Singapur – Yeni Zelanda örnekleminde dokuz maddenin, Türkiye - Finlandiya örnekleminde 10, Türkiye - Rusya örnekleminde 10, Yeni Zelenda - Avustralya örnekleminde dört maddede değişen madde fonksiyonu olduğu tespit edilmiştir. Buna göre kültürel ve dilsel farklılıklar arttıkça değişen madde fonksiyonu gösteren maddelerin sayısının da arttığı görülmektedir. Farklı dil ve farklı kültüre sahip ülkelerde maddelerin %91'i DMF gösterirken, bu oran farklı kültür ve aynı dilde %82'ye, aynı kültür ve aynı dilde %37'ye düşmüştür.

Arslan (2014) çalışmasında, MIMIC (Multiple Indicators Multiple Causes) (Çoklu Göstergeler Çoklu Nedenler, ÇGÇN) ve Lojistik Regresyon (LR) tekniklerinin DMF belirleme başarısını 1. tip hata ve güç oranlarını temel alarak değerlendirmiştir. Çalışmada, grupların yetenek dağılımları (N (0,1) ve N(-0.5,1)), örneklem büyüklüğü (2000 ve 4000 bireylik) ve DMF gösteren madde yüzdesi (%10 ve %20) değişken olarak ele alınırken, testin uzunluğu (30 madde), referans gruptaki bireylerin yetenek parametreleri (N(0,1)), odak ve referans grup oranı (1:1) ve DMF gösterdiği saptanan maddeler için gruplar arasındaki madde güçlük parametrelerindeki değişim (0,6) sabit tutulan koşullar olarak belirlenmiştir. Yapılan incelemeler sonucunda bazı durumlarda MIMIC tekniğinin, bazı durumlarda ise LR tekniğinin 1. tip hata oranları daha düşük çıkmıştır. Bazı durumlarda ise tekniklerin 1. tip hata oranları birbirine çok benzer değerler almıştır. Güç oranlarındaki farklılıklara bakıldığında ise; iki modelde de tüm koşullar için güç oranları kabul edilebilir seviyenin (0,70) üzerinde yer aldığı anlaşılmıştır. Fakat MIMIC tekniğinin güç oranının daha yüksek olduğu koşul sayısı, LR tekniğinin güç oranının daha yüksek olduğu koşul sayısından fazladır. Örneklem büyüklüğünün artması MIMIC tekniğinde 1. tip hata oranını önemli miktarda azaltırken, LR tekniğinde çok fazla etki yaratmamıştır. DMF'li madde yüzdelerinin artması MIMIC tekniğinin 1. tip hata oranına çok fazla etki yaratmazken, LR tekniğinde 1.tip hata oranını önemli ölçüde artırmıştır. Odak grupta bulunan kişilerin yetenek dağılımlarının farklı olması ise her iki teknik içinde 1. tip hata oranı miktarına çok fazla etki etmemiştir. Köse (2015) çalışmasında, PISA 2009'da uygulanan öğrenci anketinde yer alan okul öğrenmelerine ilişkin algı alt ölçeği (q32) ve öğretmenlere yönelik algı alt ölçeğinde (q33) bulunan maddeleri cinsiyet ile farklı kültür ve dil değişkenleri açısından incelemiştir. Araştırmanın çalışma grubuna İngiltere, İrlanda, ABD ve Türkiye alt gruplarından katılan öğrenciler dâhil edilmiştir. Cinsiyet değişkeni açısından maddeleri incelemek için tüm ülkelerden elde edilen bir veri grubu ile cinsiyete göre maddelerin DMF gösterip göstermediği incelenmiştir. Farklı kültür ve dil değişkeni için ise İrlanda ve İngiltere aynı dil – benzer kültür, ABD ve İngiltere aynı dil - farklı kültür, ABD ve Türkiye farklı dil – farklı kültür özelliklerini karşılaştırmak için ikişerli olarak gruplandırılmıştır. DMF analizlerine geçmeden önce MTK'nın tek boyutluluk varsayımı test edilmiştir. Veri gruplarına ayrı ayrı doğrulayıcı faktör analizi uygulanmış ve tek boyutluluk varsayımı doğrulanmıştır. DMF analizleri yürütülürken aşamalı tepki modeli altında kullanılan olabilirlik oranı testi (likelihood ratio test) tercih edilmiştir. Yapılan analizler sonucunda okul öğrenmelerine ilişkin algı alt ölçeğinde cinsiyete göre iki, İrlanda-İngiltere

örneklemlerinde bir, ABD-İngiltere'de iki ve ABD-Türkiye'de dört maddede, öğretmenlere yönelik algı alt ölçeğinde ise cinsiyete göre bir, İrlanda-İngiltere örneklemlerinde bir, ABD-İngiltere'de iki ve ABD-Türkiye'de dört maddede DMF bulunduğu tespit edilmiştir. DMF bulunan maddeler incelendiğinde kültürel ve dilsel benzerlik artıkça DMF oranın azaldığı belirlenmiştir. Bu durum maddelerdeki ifadelerin kültürel ve dilsel benzerliği olan ülkelerde aynı anlamı ifade etmesi ile açıklanabilir.

Başusta ve Gelbal (2015) çalışmasında, PISA 2006 uygulaması Fen Okuryazarlığı alt testinde yer alan maddelerin farklı kültür ve dil değişkenine göre yanlılık gösterip göstermediğini incelemiştir. DMF analizlerini gerçekleştirirken Mantel-Haenszel (MH), Lojistik Regresyon (LR) ve Alan İndeksleri yöntemlerini kullanmıştır. Araştırmanın çalışma grubunu Kanada, Avustralya, İngiltere ve Türkiye'den beşinci kitapçığı alan öğrencilerin yanıtları ile İngiltere ve Türkiye'den birinci kitapçığı yanıtlayan öğrencilerin yanıtları kullanılmıştır. DMF analizleri yürütülürken ülkeler aynı kültür - farklı dil (İngilizce ve Fransızca form kullanan Kanada örneği), farklı kültür - aynı dil (Avustralya-İngiltere örneği) ve farklı kültür - farklı dil (İngiltere – Türkiye örneği) olacak şekilde ikişerli olarak gruplandırılmıştır. Yapılan analizler sonucunda Kanada örnekleminde Fransızca formu ve İngilizce formu alan gruplar arasında; MH yöntemine göre üç maddede B düzeyinde, üc maddede A düzeyinde; LR yöntemine göre üç maddede A düzeyinde; alan indeksleri yöntemine göre beş maddede DMF bulunmuştur. Avustralya–İngiltere örnekleminde MH yöntemine göre; dört maddede A düzeyinde, bir maddede B düzeyinde ve bir maddede C düzeyinde; LR yöntemine göre dört maddede A düzeyinde; alan indeksleri yöntemine göre iki maddede DMF bulunmuştur. Birinci kitapçık için yapılan analizlerde ise Türkiye - İngiltere örnekleminde MH yöntemine göre beş maddede A düzeyinde, iki maddede B düzeyinde ve üç maddede C düzeyinde; LR yöntemine göre; beş maddede A düzeyinde ve bir maddede B düzeyinde; alan indeksleri yöntemine göre ise altı maddede DMF bulunmuştur. Kullanılan DMF belirleme yöntemlerinden elde edilen sonuçların birbiri ile uyumu incelenmiştir. Bunun için LR yöntemi için etki büyüklüğü değerleri; MH için hesaplanan ki-kare değerleri, işaretli alan indeksleri ve işaretsiz alan indeksleri ile elde edilen değerler arasındaki sıra farkları korelasyonu incelenmiştir. LR ve MH arasındaki korelasyon ve alan indeksleri yöntemleri arasındaki korelasyon aynı kültür - farklı dil gruplaması için 0,01 düzeyinde anlamlıdır. Farklı kültür - aynı dil için 0,05 düzeyinde ve farklı kültür - farklı dil için MH ve LR arasındaki korelasyon 0,01

düzeyinde; alan indeksleri yöntemleriyle elde edilen korelasyon ise 0,05 düzeyinde anlamlı bulunmuştur.

Gür (2019) çalışmasında, PISA 2015 öğrenci anketini ele almıştır. Ankette bulunan fen bilimlerine karşı tutum maddelerinin, farklı kültür ve dil değişkenlerine göre DMF içerip içermediğini incelemiştir. DMF analizlerini yürütürken genelleştirilmiş Mantel Haenszel (GMH), Ordinal Lojistik Regresyon (OLR) ve Poly-SIBTEST yöntemlerinden faydalanmıştır. Araştırmanın veri setini PISA 2015 uygulamasına Türkiye, İrlanda, ABD ve İngiltere'den katılan 19.663 öğrenci oluşturmaktadır. DMF analizleri gerçekleştirilirken ülkeler aynı dil - benzer kültür, aynı dil - farklı kültür ve farklı dil - farklı kültür şeklinde ikişerli olarak gruplandırılmıştır. Elde edilen bulgular doğrultusunda, dokuz tutum maddesinden, İngiltere-İrlanda karşılaştırmasında GMH ve OLR yöntemleriyle yapılan analizde altı maddede, poly-SIBTEST yöntemiyle yapılan analizde yedi maddede DMF belirlenmiştir. İngiltere-ABD karşılaştırmasında GMH ve OLR yöntemleriyle yapılan analizde yedi maddede, poly-SIBTEST yöntemiyle yapılan analizde altı maddede DMF belirlenmiştir. İngiltere-Türkiye karşılaştırmasında ise GMH yöntemiyle sekiz maddede, OLR yöntemiyle dokuz maddede, poly-SIBTEST yöntemiyle yedi maddede DMF tespit edilmiştir. Kullanılan yöntemlerin uyumuna bakıldığında ise OLR, GMH ve poly-SIBTEST yöntemlerinin DMF olarak belirledikleri madde sayısında uyum görülürken, yöntemlerin DMF'li maddelerin düzeylerini belirlemede farklılıklar gösterdiği belirlenmiştir.

2.Yurt Dışı Araştırmalar

Schulz (2003) tarafından gerçekleştirilen araştırmada PISA 2000 uygulaması temel alınarak ölçülmesi amaçlanan bazı yapıların geçerliği ve eşdeğerliği üzerine bir inceleme gerçekleştirilmiştir. İnceleme sürecinde ÇG-DFA ve Madde Tepki Kuramı yöntemleri işe koşulmuştur. Araştırmada zayıf ölçme değişmezliği sınanmış ve sınama sonucunda bazı ülkeler için çeviriden kaynaklı sorunlar olduğu ortaya çıkmıştır. Bunun haricinde, bazı maddelerin ölçülmesi amaçlanan yapıya hizmet etmediğine ulaşılmıştır. Bu durumun ülke kıyaslamalarında önemli bir soruna neden olduğu vurgulanmıştır. Son olarak test yapılarını doğrulamada yapısal eşitlik modelinden de yararlanılabileceği aktarılmıştır.

Schulz (2005) tarafından gerçekleştirilen diğer bir araştırmada PISA 2006 ön verisi kullanılmış ve ölçme değişmezliğinin sınanması amaçlanmıştır. Bu süreçte öğrenci anketinde bulunan maddelerin ölçme değişmezliği grupların her birinde iki benzer dil ve

kültür olacak biçimde sekiz farklı grup kapsamında sınanmıştır. Analiz sürecinde Doğrulayıcı Faktör Analizi (DFA) ve Madde Tepki Kuramı (MTK) kullanılmıştır. Araştırma sonucunda, benzer dil ve kültür grubunda yer alan ülkeler arasında zayıf değişmezlik bulunmuştur. Fakat faktör yükleri tüm ülkeler arasında eşit biçimde tutulduğunda istenilen uyuma ulaşılamamıştır. Son olarak DFA ve MTK yöntemleriyle yapılan sınamalarda aynı sonuçlara ulaşılmasının beklenmemesi gerekliliği belirtilmiştir.

Diğer bir araştırma Yıldırım ve Berberoğlu (2009) tarafından gerçekleştirilmiştir. Araştırmacılar PISA 2003 uygulaması matematik okuryazarlığı testi maddelerini değişen madde fonksiyonu açısından incelemişlerdir. Araştırmada yöntem olarak, Sınırlandırılmış Faktör Analizi, Mantel-Haenszel ve Madde Tepki Kuramı olabilirlik oranı kullanılmıştır. Araştırmanın örneklemi Türk ve Amerikan öğrencilerden oluşmuştur. Sınırlandırılmış Faktör Analizi, Mantel-Haenszel ve Madde Tepki Kuramı olabilirlik oranı yöntemleriyle gerçekleştirilen analiz sonucunda üç yöntem tarafından da 152 maddenin 10'unun değişen madde fonksiyonu içerdiği tespit edilmiştir. Araştırma sonucunda değişen madde fonksiyonunun nedenleri uzmanlar tarafından çeviriden kaynaklı, nicelik bildirilen ve birebir karşılığı olmayan sözcüklerin tercih edilmesi ve testte yer alan maddelerin bilişsel özellikleri olarak ifade edilmiştir.

Le (2006) tarafından gerçekleştirilen diğer bir araştırma PISA 2006 ön uygulaması değişen madde fonksiyonu açısından incelenmiştir. Bu kapsamda, PISA 2006 fen okuryazarlığı testi maddelerinin farklı ülke, dil ve cinsiyet değişkenleri açısından değişen madde fonksiyonu içerip içermediğinin tespit edilmesi amaçlanmıştır. Araştırmanın örneklemini 50 ülkeden 840.000 öğrenci oluşturmuştur. Değişen madde fonksiyonunu belirlemek için Madde Tepki Kuramı işe koşulmuştur. Analiz sonucunda cinsiyete bağlı %10, ülke farklılığına bağlı %25 ve farklı dile bağlı olarak %39-59 aralığında değişen madde fonksiyonu tespit edilmiştir. Bu durumun nedenlerinin testin çevirisinden kaynaklı olarak kültürel alan ve eğitim programlarındaki farklılıkların olabileceği ifade edilmiştir.

Diğer bir araştırma Grisay ve Monseur (2007) tarafından gerçekleştirilmiştir. Araştırmacılar PISA 2000 ve 2001'de yer alan okuma becerileri parçalarının teste katılan ülkelerin farklı dil gruplarındaki eşdeğerliklerini incelemişlerdir. Araştırma sonucunda testteki madde formatının ve aynı okuma parçasında bulunan maddelerin yerel bağımlılığının değişen madde fonksiyonuna neden olabileceğine ulaşılmıştır. Bunun haricinde, kültürel ve bölgesel farklılıklar azaltılsa bile (Örneğin, Kanada ülkesinde İngilizce ve Fransızca konuşulan alanlar) testlerin çeviriden kaynaklı değişen madde

fonksiyonunun güçlü bir biçimde ortaya çıkabileceği ortaya konulmuştur. Araştırmacılar ayrıca, bir testin her ne kadar nitelikli bir çeviriye sahip olsa da eşdeğer olmama problemiyle karşılaşabileceğini belirtmişlerdir. Araştırma sonucunda aynı dili konuşulan ülkelere yönelik ortak bir formatın geliştirilmesi gerekliliği ve madde yapısıyla kültürel ve dilsel özellikle arasındaki ilişkinin derinlemesine çalışması araştırmacılara öneri olarak sunulmuştur.

PISA testiyle ilgili diğer bir eşdeğerlik araştırması Grisay vd. (2007) tarafından gerçekleştirilmiştir. Araştırmacılar, PISA 2006 ön deneme veri seti fen bilimleri sorularının ölçek temelinde eşdeğerliğini incelemişlerdir. Analiz sürecinde kümeleme analizi işe koşulmuştur. Analiz sonucunda, coğrafi açıdan yakın ülkeler, aynı dili konuşan ülkeler, kültürel açıdan benzer ülkeler ve sosyo-ekonomik düzey açısından birbirine yakın ülkeler aynı küme içerisinde yer almıştır. Bu sonucun ardından kümeleme analizi PISA 2006 gerçek testine de uygulanmıştır. Analiz sonucunda PISA 2006 ön deneme veri setinden elde edilen sonuçlara benzer sonuçlara ulaşılmıştır. Ayrıca testteki soruların tüm katılımcı ülkeler için benzer biçimde değişen madde fonksiyonu göstermediği ortaya çıkarılmıştır. Son olarak, dilsel farklılık nedeniyle ortaya çıkan yanlılığın büyüklüğü incelenmiştir. Bu incelemede faktör analizinden yararlanılarak madde zorluklarında açıklanan ve açıklanamayan varyans oranları karşılaştırılmıştır. Bunun sonucunda, Orta Doğu ve Asya'da bulunan ülkelerin açıklanamayan varyans oranlarının diğer bölgelerdeki ülkelere kıyasla daha yüksek olduğuna ulaşılmıştır.PISA uygulaması değişen madde fonksiyonu açısından inceleyen diğer bir araştırma Le (2009) tarafından yapılmıştır. Araştırmacı, PISA 2006 ön uygulamasındaki fen okuryazarlığı testinde yer alan maddelerin cinsiyet değişkeni açısından değişen madde fonksiyonu içerip içermediğini sınamıştır. Bu sınama 60 farklı dil ve 50 ülke verisi kapsamında yapılmıştır. Analiz sürecinde Madde Tepki Kuramı kullanılmıştır. Analiz sonucunda, farklı dil grupları kapsamında fen okuryazarlığı testinde yer alan maddelerin %24'ünde cinsiyetlerden birine avantaj tanıyan biçimde değişen madde fonksiyonuna ulaşılmıştır. Bu sonucun nedenleri, maddelerin formatı ve içerik olarak belirtilmiştir. Ayrıca araştırmada ülkenin ve testin dilinin cinsiyet temelli değişen madde fonksiyonu üzerinde etkili olduğu belirtilmiştir.

Finch ve French (2011) yaptıkları simülasyon çalışmasında gizil değişken modellerin performansını değerlendirmeyi amaçlamıştır. Analizler, gizil ve gözlemlenen değişkenleri ilişkilendirmek için esnek ve etkili bir araç olan çoklu göstergeler çoklu nedenler (MIMIC) modeli kullanılarak yürütülmüştür. Veriler, çok düzeyli bir modelleme yaklaşımının uygun

olması için hiyerarşik bir çerçevede simüle edilmiştir. Bu tür çok düzeyli veri yapılarını yapısal eşitleme modellerinde göz ardı etmenin etkisini belirlemek için iç içe verileri hem hesaba katarak hem de katmayarak analizler yapılmıştır. Sonuç olarak standart yöntemlerle gizil değişken kullanan yöntemler arasında 1.tip hata ve güç oranları bakımından benzerlik görülmüştür.

Asil ve Brown (2016) çalışmasında PISA uygulamasında yer alan okuma becerileri alt testinin farklı uluslar, kültürler ve diller arasında kullanılmasını eleştirmiştir. Okuma becerileri alt testinin tasarımının potansiyel olarak altında yatan dilbilimsel ve kültürel önyargılara işaret ederek, ekonomiler arasında yapılacak karşılaştırmaların ne denli geçerli olduğuna dikkat çekmiştir. Çalışmanın örneklemini Avustralya İngilizcesi konuşan grup referans grup olmak üzere 55 ülkeden 32.000 kişi oluşturmuştur. Çalışmada Avustralya İngilizcesi konuşan grubun performansı baz alınarak dil, kültür ve ekonomik gelişmeye göre PISA okuma becerileri alt testindeki değişmezliğe, değişmezliğin türüne ve büyüklüğüne odaklanılmıştır. Ortalamalara ve kovaryans yapısına (MACS) dayalı Çoklu Grup Doğrulayıcı Faktör Analizi, değişmezlik derecesinde her ekonomi için bir dMACS etki büyüklüğü indeksi oluşturmak için kullanılmıştır. Yalnızca üç ülke, Avustralya ile skaler değişmezliğe sahip çıkmıştır. Karşılaştırmaların sadece %31'inde orta veya büyük etkiler gözlenmiştir. PISA ekonomik, sosyal ve kültürel durum indeksi, dMACS ile orta derecede ters bir korelasyona sahip çıkmıştır. Bu durum eğitimin sosyoekonomik kaynağının ölçüm değişmezliğinde önemli bir rol oynadığını gösterirken, eğitim uygulaması ve dil faktörlerinin değişmezlikte daha küçük bir rol oynadığı görülmüştür.

Huang, Wilson ve Wang (2016) çalışmaların PISA 2006 fen okuryazarlığı testindeki maddeleri dil ve kültür değişkenleri açısından incelemişlerdir. Araştırmaya Amerika Birleşik Devletleri, Kanada, Hong Kong ve Çin ülkeleri dâhil edilmiştir. Katılımcı grubu ise Amerika Birleşik Devletleri'nden 5.611, Kanada'dan (İngilizce konuşan) 17.555, Çin'den 4892 ve Hong Kong'dan 4.645 öğrenci oluşturmuştur. Karşılaştırmalar, Amerika Birleşik Devletleri ve Kanada arasında, Çin ve Hong Kong arasında ve Amerika Birleşik Devletleri ve Çin arasında gerçekleştirilmiştir. Araştırmada incelenen veri seti PISA 2006 fen okuryazarlığı alt testinde yer alan 192 maddedir. Maddelerin 103'ü bilişsel süreçleri ölçerken, 89 madde ise tutumları ölçmektedir. Verilerin analizinde çok boyutlu rastgele katsayı multinominal logit prosedürü işe koşulmuştur. Bunun haricinde değişen madde fonksiyonu nedenlerini belirlemek için içerik analizi gerçekleştirilmiştir. Araştırma sonucunda Amerika Birleşik Devletleri ve Kanada arasında düşük düzeyde değişen madde

fonksiyonu tespit edilmiştir. İçerik analizi sonucunda ise değişen madde fonksiyonu ortaya çıkaran en önemli nedenin öğretim programı kapsamı olduğuna ulaşılmıştır.

Dzul-Garcia ve Atar (2020) araştırmalarında PISA 2015 fen okuryazarlığı alt testinde yer alan maddelerin kültür değişkenine göre değişen madde fonksiyonu içerip içermediğini tespit etmeyi amaçlamışlardır. Araştırmacılar Şili, Kosta Rika ve Meksika ülkeleri arasında bir karşılaştırma yapmışlardır. Araştırmanın çalışma grubuna Şili'den 954, Kosta Rika'dan 827 ve Meksika'dan 1047 öğrenci dâhil edilmiştir. Veri seti seçiminde ise öncelikle PISA 2015 teknik raporu incelenmiş, ardından fen okuryazarlığı alt testinde en fazla iki değerli ölçümlenen maddeye sahip testler araştırılmış ve 16 madde içeren 12. kitapçık araştırmada kullanılmak üzere seçilmiştir. Araştırmaya dâhil edilen maddeleri değişen madde fonksiyonu açısından incelemek üzere Mantel-Haenszel ve Lojistik Regresyon analizleri işe koşulmuştur. Analiz sürecinde SPSS 23.0 programından yararlanılmıştır. Araştırma sonucunda tüm ülkeler için maddelerin %25 oranında yüksek etki büyüklüğüne sahip değişen madde fonksiyonu içerdiği tespit edilmiştir. Bu sonuç doğrultusunda maddelerin değişen madde fonksiyonu gösterme nedenlerini tespit etmek amacıyla uzman görüşüne başvurulabileceği önerilmiştir.

3.DMF Belirleme Yöntemlerinin Karşılaştırılmasına Yönelik Araştırmalar

Narayanan ve Swaminathan (1996) çalışmasında, 1.tip hata oranı ve tek biçimli olmayan DMF'li maddeleri belirleme gücü bakımından MH, LR ve SIBTEST yöntemlerini karşılaştırmıştır. Araştırmada kullanılan veri seti örneklem büyüklüğü, odak ve referans grupları arasındaki yetenek dağılımı farklılığı, DMF'li maddelerin oranı, DMF etki büyüklüğü ve madde tipi özellikleri manipüle edilerek 384 koşul incelenmiştir. Araştırma sonucunda birçok koşulda LR ve SIBTEST tek biçimli olmayan DMF'nin belirlenmesinde eşit derecede güçlü çıkarken, MH yöntemi çok etkili bulunmamıştır. 1.tip hata oranı bakımından ise MH beklenen sınırlar dâhilinde, LR ve SIBTEST ise beklenenin üzerinde sonuçlar vermiştir. Genel itibariyle SIBTEST sonuçları LR sonuçlarına göre %1'lik bir artış göstermiştir.

Gierl, Khaliq ve Boughton (1999) çalışmasında, Kanada'daki Alberta Eğitim Öğrenci Değerlendirme Merkezinde 1994, 1995, 1996, 1997 ve 1998 yıllarında uygulanan matematik ve fen bilgisi başarı testlerini cinsiyet değişkeni açısından farklı DMF belirleme yöntemleriyle incelemeyi amaçlamıştır. Çalışmada 6000 kız 6000 erkek olmak üzere rast gele seçilen 12000 öğrencinin verileri kullanılmıştır. Analizler yapılırken MH, SIBTEST

ve LR yöntemleri kullanılmıştır. İncelenen 469 matematik maddesinden MH 34 (%7,3), SIBTEST 40 (%8,5), LR 49 (%10,5) adet tek biçimli DMF ve 1 (%0,2) tane tek biçimli olmayan DMF tespit etmiştir. Analiz edilen 428 fen bilgisi maddesinden MH 36 (%8,4), SIBTEST 63 (%14,7), LR 74 (%17,3) tane tek biçimli DMF ve 2 (%0,5) adet tek biçimli olmayan DMF belirlemiştir. Her iki alanda yapılan analiz sonuçları dikkate alındığında maddelerin büyük çoğunluğunun B ve C düzeyinde DMF göstermediği ve fen bilgisinin matematiğe kıyasla daha fazla sayıda DMF'li madde içerdiği görülmüştür.

Gierl, Jodoin ve Ackermann (2000) çalışmasında, testlerin uyarlanması ya da başka dile çevrilmesi sonucunda artan DMF'li madde miktarının, üç popüler DMF belirleme yönteminin - MH, SIBTEST ve LR – performansını nasıl etkilendiğini incelemiştir. Maddelerin 1.tip hata oranlarındaki değişim ve yöntemlerin güç performansı ele alınmıştır. Yapılan simülasyon çalışmasında 40 maddelik testte yer alan DMF'li madde sayısı (%20, %40 ve %60 oranlarında), örneklem büyüklüğü (tüm gruplar için 250, 500 ve 1.000 katılımcı) ve grupların yetenek dağılımı farklılıkları (eşit ve eşit olmayan) manipüle edilmiştir. Yapılan analizlerde iki büyük sonuca ulaşılmıştır. İlk olarak DMF'li madde miktarındaki artış (maddelerin %60'ına kadar) üç yöntemi de 1.tip hata oranı bakımından olumsuz etkilememiş hatta küçük örneklemlerde bile hata oranı nominal alfa seviyesinin altında kalmıştır. Diğer bir sonuç ise yöntemlerin güç performansının bu koşullar altında farklılık göstermesidir. Örneklem sayısı 250'ye ulaştığında, üç yöntemin de DMF içeren maddeleri belirleyebilme güçleri orta seviyede bulunmuştur. Örneklem sayısı 500'e çıkarıldığında, SIBTEST'in gücü eşit ve eşit olmayan yetenek dağılımlarında önemli ölçüde artmıştır. Örneklem sayının 1.000'e ulaştığı durumda ise, SIBTEST yönteminin gücü daha fazla artış göstermiş ve bu üç yöntem arasında en güçlü DMF belirleme yöntemi olduğu belirtilmiştir.Farklı yöntemleri kullanarak değişen madde fonksiyonunu belirlemeye yönelik diğer bir araştırma Hidalgo ve Lopez-Pina (2004) tarafından gerçekleştirilmiştir. Araştırmacılar, değişen madde fonksiyonunu belirlemede lojistik regresyon, Mantel-Haenszel ve yenilenmiş Mantel-Haenszel yöntemlerini kullanmışlardır. Ayrıca araştırma kapsamında etki büyüklüğü ölçümlerinin de karşılaştırılması gerçekleştirilmiştir. Araştırmanın veri seti simülatif olarak oluşturulmuştur. Bu kapsamda, 1.000 kişiden oluşan referans grup ve yine 1000 kişiden oluşan odak grup araştırmaya dâhil edilmiştir. Araştırma sonucunda lojistik regresyon yönteminin standart Mantel-Haenszel ve yenilenmiş Mantel-Haenszel yöntemlerine göre daha fazla değişen madde fonksiyonlu madde tespit ettiği sonucuna ulaşılmıştır. Tüm bunlara rağmen, lojistik

regresyonun etki büyüklüğü noktasında değişen madde fonksiyonunu belirlemedeki kadar duyarlı olmadığı görülmüştür. Bu sonuçlar, değişen madde fonksiyonunu ve etki büyüklüğünü belirlemede farklı yöntemlerin farklı sonuçlar ortaya çıkarabileceği biçiminde yorumlanmıştır.

Finch (2005) çalışmasında, DMF belirleme yöntemlerinden ÇGÇN doğrulayıcı faktör analizi modeli ile Mantel-Haenszel, SIBTEST ve MTK olabilirlik oranı yöntemlerini 1.tip hatalarını ve güçlerini incelemiştir. İki kez gerçekleştirilen simülasyon çalışmalarıyla yöntemlerin DMF belirleme performansları test edilmiştir. Birinci simülasyon çalışmasında gizil özellik her iki grup için de 0 ortalamaya sahipken, ikincide referans grubunun ortalaması 1 ve odak grubunun ortalaması 0'dır. Referans grubunun her koşulda 500 katılımcısı varken, odak grubu 100 ve 500 katılımcıdan oluşan iki grubu kullanmıştır. Araştırma, 20 ve 50 maddelik iki test kullanılarak gerçekleştirilmiştir. Yapılan analizler sonucunda, ÇGÇN modelinin bazı durumlarda kullanılabilir bazı durumlarda ise kullanılamaz olduğu ortaya çıkmıştır. Özelikle 3-PL modele sahip kısa testler (20 madde) için ÇGÇN modelinin kullanılması durumunda 1.tip hata oranı oldukça yüksek çıkmıştır. Aynı koşullarda diğer yöntemler ise ÇGÇN modelinden daha az hata oranına sahiptir. Test 50 madde içerdiğinde ve/veya 2PL model sahipken, ÇGÇN modelinin odak grubu boyutuna bakılmaksızın 1.tip hata oranı ve güç açısından diğer DMF belirleme modellerine kıyasla daha iyi performans gösterdiği sonucuna varılmıştır.

Zheng, Gierl, Cui (2007) tarafından gerçekleştirilen araştırmada gerçek veri üzerine odaklanılmıştır. Bu araştırmada üç farklı DMF belirleme yönteminin (LR, MH ve SIBTEST) cinsiyet ve dil değişkeni açısından DMF'yi belirlemede karşılaştırılması yapılmıştır. İlgili yöntemleri karşılaştırmada DMF yöntemlerinin etki büyüklüğü korelasyonları, eşleştirme yüzdeleri ve göreceli eşleştirme yüzdeleri kullanılmıştır. Araştırmanın veri seti geniş ölçekli bir değerlendirmeden oluşmaktadır. Veri setinin Çince ve İngilizce olmak üzere iki dil versiyonu bulunmaktadır. Araştırmaya rastgele örnekleme ile 2.000 katılımcı dâhil edilmiştir. Araştırma sonucunda, DMF belirleme yöntemleri arasında etki büyüklüğü ölçümlerinde yüksek korelasyona ulaşılmış, eşleştirme yüzdelerinde orta ve yüksek düzeyde uyum görülmüş ve en yüksek eşleştirme yüzdesi uyumunun LR ve SIBTEST yöntemleri arasında olduğuna ulaşılmıştır.

Doğan ve Öğretmen (2008) tarafından gerçekleştirilen araştırmada DMF belirleme yöntemlerinden Ki-Kare, MH ve LR teknikleri karşılaştırılarak benzerlik ve farklılıkların ortaya çıkarılması amaçlanmıştır. Araştırmada 2003 Ortaöğretim Kurumları Seçme ve

Yerleştirme Sınavı fen bilgisi alt testi verileri temel alınmıştır. Araştırmanın örneklemini Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı'na giren adaylardan yansız bir biçimde seçilen 3.345 öğrenci oluşturmuştur. DMF analizi cinsiyet değişkeni üzerine gerçekleştirilmiştir. Araştırma sonucunda değişen madde fonksiyonunu belirlemede yöntemler arasında manidar farklılıklar olduğuna ulaşılmıştır. Gerçekleştirilen analizlere göre en fazla DMF, MH yöntemiyle tespit edilmiştir. Ardından işaretli ki-kare yöntemiyle dokuz maddenin DMF içerdiği belirlenmiştir. MH ve işaretli ki-kare yöntemlerinin aksine işaretsiz ki-kare ve LR yöntemleriyle gerçekleştirilen analizlerde değişen madde fonksiyonuna ulaşılmamıştır.

Gök, Kelecioğlu ve Doğan (2010) çalışmasında, DMF belirleme yöntemlerinden MH ve LR yöntemlerini karşılaştırmayı amaçlamışlardır. Çalışmada 2005 yılında yapılan Ortaöğretim Kurumları Sınavı'nın (OKS) matematik ve fen bilgisi testleri ele alınmıştır. Testteki maddelerin cinsiyet ve okul türü değişkenine göre DMF gösterip göstermediği araştırılmıştır. Analiz sonuçları doğrultusunda teknikler arasında çoğunlukla düşük seviyede bir uyum tespit edilmiştir.

Atalay, Gök, Kelecioğlu ve Arsan (2012) çalışmasında, farklı koşullar altında DMF belirleme yöntemlerinin karşılaştırılmasını amaçlayan bir simülasyon çalışması gerçekleştirmiştir. Gözlenen puana dayalı yöntemlerden LR ve MH ile örtük puana dayalı yöntemlerden SIBTEST ve MTK-OO yöntemlerini karşılaştırmıştır. Örneklem büyüklüğü, odak ve referans gruplarının yetenek düzeyi ve testteki DMF'li madde oranı değişkenlerinin DMF belirlemede yöntemlere olan etkileri incelenmiştir. Yapılan analizler sonucunda DMF'li maddeleri belirlemede SIBTEST ve MTK-OO yöntemlerinin LR ve MH yöntemlerine karşı daha duyarlı ve etkili olduğu belirlenmiştir. Araştırmada kullanılan dört yöntem içinde DMF'li maddeleri belirlemede LR en düşük orana sahipken, MTK-OO yöntemi ise en yüksek orana sahip olmuştur. Tek biçimli DMF'yi belirlemede yöntemlerin uyumuna bakıldığında MH, SIBTEST ve MTK-OO uyumlu bulunurken tek biçimli olmayan DMF'yi belirlemede ise LR, SIBTEST ve MTK-OO yöntemleri birbiriyle uyumlu sonuçlar vermiştir.

Akın Arıkan, Uğurlu ve Atar (2014) çalışmasında, DMF belirlemede kullanılan MIMIC, LR, MH ve SIBTEST yöntemlerinin benzerlik ve farklılıklarını incelemiştir. DMF analizleri yapıldıktan sonra yöntemlerin tutarlılığını saptamak için elde edilen sonuçlar uzman görüşüne sunulmuştur. Analizler 2009 yılında SBS'na katılan 340.000 kişi arasından oluşturulan 300, 600, 1.200 ve 2.000 kişilik farklı alt örneklemler üzerinden

yürütülmüştür. Elde edilen sonuçlar doğrultusunda büyük örneklemlerde yapılan analizlerde yöntemlerin daha uyumlu sonuçlar verdiği, küçük örneklemlerde ise daha düşük benzerlikte sonuçlara ulaşıldığı görülmüştür. Ayrıca uzman görüşlerine bakıldığında yöntemlerden elde edilen sonuçlarla tutarlı olduğu tespit edilmiştir.

Erdem (2015) çalışmasında, TEOG Ortak Sınav'da bulunan maddelerin kitapçık türü değişkenini ele alarak DMF analizlerini yürütmüştür. DMF analizlerini yaparken Mantel-Haenszel (MH), Lojistik Regresyon (LR) ve SIBTEST yöntemlerini kullanmış ve üç yöntem kullanılarak elde edilen sonuçları inceleyerek kıyaslamasını yapmıştır. 2014-2015 eğitim öğretim yılının sonbahar döneminde yapılan TEOG Ortak Sınav'a katılan öğrencilerden seçkisiz yöntemle belirlenen 12.000 öğrenci araştırmanın örneklemini meydana getirmiştir. Uygulanan analizler sonucunda, Matematik alt testinde kitapçık türüne göre hiçbir maddenin DMF göstermediği belirlenmiştir. Türkçe ve Fen ve Teknoloji testlerinde az sayıda DMF'li madde tespit edilirken, T.C. İnkılap Tarihi ve Atatürkçülük, İngilizce ve Din Kültürü ve Ahlak Bilgisi alt testlerinde ise daha fazla DMF'li madde olduğu görülmüştür. Yöntemlerden çıkarılan sonuçlar karşılaştırıldığında ise SIBTEST yöntemi ile 120 madde içinden 28 tanesinde DMF bulunurken, MH yönteminde 17, LR yönteminde 8 maddede DMF bulunmuştur. DMF'li olduğu belirlenen madde miktarına bakıldığında yöntemlerin benzer sonuçlar çıkarmadığı görülmüştür. Yöntemlerin aynı maddeyi DMF'li olarak tespit edebilmesi açısından uyumları incelendiğinde, en çok uyum MH ve SIBTEST yöntemleri arasında görülürken, en az uyum MH-LR yöntemleri arasında görülmüştür.

Arslan (2020) çalışmasında, 2016 yılında uygulanan TEOG sınavı İngilizce alt testinin öğrenim görülen kurum (devlet okulu-özel okul) ve cinsiyet değişkenine göre DMF gösterip göstermediğini ve kullandığı DMF belirleme yöntemlerinin birbiriyle ne derece uyum gösterdiğini incelemiştir. Çalışmada DMF belirleme yöntemlerinden MH, LR ve SIBTEST kullanılmıştır. Sınava katılan 1.133.857 öğrencinden rastgele olarak seçilen 9.108 öğrencinin verdiği yanıtlar kullanılmıştır. Yapılan analizler sonucunda cinsiyet değişkenine göre dört, okul türü değişkenine göre bir maddede DMF tespit edilmiştir. Yöntemlerin uyumunu incelemek için Spearman rho katsayısından faydalanılmıştır. Buna göre en yüksek uyum MH-SIBTEST yöntemleri arasında görülürken LR yönteminin diğer yöntemlerle uyumunun düşük olduğu görülmüştür.

BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın modeli, çalışma grubu, verilerin elde edilmesi, verilerin analizi ve verilerin analizi sırasında kullanılan bilgisayar programlarına ilişkin bilgilere yer verilmiştir.

1.Araştırma Modeli

Bu araştırmada, PISA 2018 okuma becerileri testinde bulunan maddelerin farklı kültür ve dil değişkenine göre DMF gösterip göstermediği incelenmektedir. Bu amaç doğrultusunda betimsel yöntemden yararlanılmıştır. Araştırmada Uluslararası Öğrenci Değerlendirme Programı (PISA) 2018 okuma becerileri testi verileri üç farklı yöntemle (SIBTEST, Lojistik Regresyon, Mantel-Haenszel) DMF analizine tabi tutulmuş ve sonuçlar betimlenmiştir.

2.Çalışma Grubu

PISA, Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından geliştirilen ve uluslararası ölçekte üçer yıllık periyotlarla uygulanan bir izleme araştırmasıdır. PISA uygulamasına, 15 yaş grubunda olup örgün eğitime devam eden öğrenciler katılmaktadır. PISA uygulaması kapsamında öğrencilerin matematik okuryazarlığı, fen okuryazarlığı ve okuma becerileri ölçülmektedir (OECD, 2019). PISA, ülkemizde nisan ayı içerisinde uygulanmaktadır. Son olarak 2018 yılında gerçekleştirilmiş ve 12 istatistiki bölgeden PISA merkezi tarafından seçkisiz olarak belirlenen 186 okuldan 6.890 öğrenci katılmıştır (Milli Eğitim Bakanlığı, 2019). PISA'ya dâhil edilecek öğrencilerin seçiminde örneklem belirleme süreci çeşitli aşamalardan oluşmaktadır. Bu kapsamda, ilk olarak ulusal merkezler tarafından tabakalı örnekleme yöntemi kullanılmıştır. Bu süreçte, PISA

uygulamasına dâhil edilecek öğrencilerin örnekleme dâhil edilme kriterleri (bölge, okul türü vb.) saptanmıştır. Bu sürecin tamamlanmasının ardından tabakalı örnekleme yöntemiyle belirlenen örneklem ulusal merkez tarafından PISA'nın uluslararası merkezine iletilmiştir. Belirlenen okullardan 15 yaş grubu içerisinde yer alan öğrencilerden seçkisiz örnekleme yoluyla sınavda yer alacak öğrencileri belirlemiştir. Okulların ve öğrencilerin kesinleşmesinin ardından süreç tamamlanmıştır (Demir, 2013).

Uygulamaya 79 ülkeden 600.000 üzerinde öğrenci katılmıştır (OECD, 2019). Bu çalışmada yapılan analizlerde, bireyselleştirilmiş test deseni göz önünde bulundurularak PISA 2018 uygulaması okuma becerileri alt testinde temel bölüm, 1.aşama ve 2.aşama bölümlerinde aynı demette yer alan madde paketleri ele alınmıştır. Temel bölüm için 1.paket (Core RC1), 1.Aşama için 1.paket (Stage 1 - R11H), 2.Aşama için 1.paket (Stage 2 - R21H) seçilmiştir. Çalışma grubu olarak ise Türkiye, Birleşik Krallık ve Fransa örneklemlerinden bu paketleri yanıtlayan öğrenciler analize dahil edilmiştir. Çalışmaya dahil olan kişiler aynı madde paketlerini yanıtlamış ve maddeleri aynı sıra ile görmüşlerdir. Çalışma grubunda yer alan ülkeler; Türkiye – Birleşik Krallık, Türkiye – Fransa ve Fransa – Birleşik Krallık şeklinde ikişerli gruplarda analiz edilmiştir. Çalışma grubuna yönelik frekans ve yüzde değerleri Tablo 5'te gösterilmiştir.

Tablo 5

Çalışma Grubuna Ait Frekans Ve Yüzde Değerleri

Gruplar	F	%
Türkiye	125	21,04
Birleşik Krallık	326	54,88
Fransa	143	24,07

PISA uygulamasına katılan ülkeler arasında maddelerin DMF göstermesinin kültür ve dile yönelik farklılıklardan kaynaklanıp kaynaklanmadığını incelemek için Türkiye, Birleşik Krallık ve Fransa örneklemleri amaçlı olarak incelenmiştir. İngilizce ve Fransızca olarak hazırlanan ölçme araçları daha sonra çeviri işlemlerinin yapılması için katılımcı ülkelere gönderilmektedir. Ülkeler bu iki versiyonu kullanarak kendi ulusal dillerinde bir çeviriye dönüştürmektedirler (OECD, 2007). Bu nedenle kıyaslama yapabilmek adına PISA uygulamasını kendi ulusal dilinde alan Birleşik Krallık ve Fransa örneklemleri ile kendi ulusal dilinde bir çeviriye dönüştürerek uygulayan Türkiye örneklemi tercih edilmiştir.

3.Verilerin Elde Edilmesi

Araştırmada kullanılan veriler, Ekonomik İşbirliği ve Kalkınma Örgütü tarafından 2018 yılında gerçekleştirilen PISA 2018 Okuma Becerileri alt testi öğrenci cevapları kullanılarak elde edilmiştir. Verilere, OECD PISA internet sitesi üzerinden erişim sağlanmıştır (OECD, 2018).

4.Verilerin Çözümlenmesi

Verilerin çözümlenmesine yönelik aşağıdaki adımlar izlenmiştir;

OECD PISA internet sitesi üzerinden erişilen PISA 2018 uygulamasına ait veri setinden Türkiye, Birleşik Krallık ve Fransa veri seti seçilmiştir. PISA 2018 uygulamasında, diğer yıllarda yapılan uygulamalardan farklı olarak bireyselleştrilmiş test deseni kullanılmıştır. PISA 2018 okuma becerileri alt testinde öğrencinin önceki maddelere vermiş olduğu yanıtların doğruluğuna göre belirlenen dinamik bir yapı söz konusudur (OECD, 2019). Bu nedenle PISA 2018 okuma becerileri alt testinde yer alan Temel, 1.Aşama ve 2.Aşama bölümlerinde aynı demette yer alan madde paketleri ele alınmıştır. Temel bölüm için 1.paket (Core RC1), 1.Aşama için 1.paket (Stage 1 - R11H), 2.Aşama için 1.paket (Stage 2 - R21H) seçilmiştir. Seçilen madde paketlerinde toplamda 39 madde yer almaktadır. Ancak madde paketlerinde 1-0 şeklinde iki kategorili olarak puanlanmayan 4 madde analize dahil edilmeyecekleri için veri setinden çıkarılmış ve toplam madde sayısı 35 olarak belirlenmiştir.

Verilerin analizine başlamadan önce veri seti düzenlenmiştir. Bu süreçte veri seti kayıp veri (missing value) ve uç değerler (outliers) açısından incelenmiştir. SPSS paket programı kullanılarak gerçekleştirilen kayıp veri analizi sonucunda her bir ülke verisine ait kayıp değerlere ilişkin oranlar Ek 1'de listelenmiştir. Yapılan kayıp veri analizi sonucunda, Little ve Rubin (2002) tarafından yapılan sınıflandırmaya göre veri seti içerisinde yer alan kayıp verilerin rastlantısal olarak dağılmadığı görülmüştür. Bu noktada, kayıp verileri gidermek amacıyla, kayıp verilerle baş etme yöntemlerinden çoklu atama (multiple imputation) yöntemi kullanılarak kayıp veriler giderilmiş ve her bir örneklem için tam bir veri seti elde edilmiştir.

Çalışmada kullanılan veri setlerine yönelik uç değerlerin bulunup bulunmadığını analiz etmek için SPSS paket programından faydalanılmıştır. Tek değişkenli uç değerlerin belirlenmesi için toplam puanlar, z standart puanına dönüştürülerek analiz edilmiştir. Z

puanı, gözlenen değişkenin ortalama değere kaç standart sapma kadar uzakta olduğunu gösterir. Z puanı değerinin ±3 aralığını aşması durumunda veri uç değer olarak değerlendirilir (Tabachnick & Fidell, 2015). Bu çalışmada yapılan tek değişkenli uç değer analizleri sonucunda, ±3 aralığını -3,02 z puanı ile aşan tek bir değere rastlanılmıştır. Birleşik Krallık örnekleminde yer alan bu veri, veri setinden çıkarılmayarak yapılacak olan analizlere dahil edilmiştir. Her bir veri setine yönelik oluşturulan uç değer analizi kutu grafiği Ek 2, Ek 3, Ek 4'te yer almaktadır.

Alt problemlere ilişkin istatistiksel analizlere başlamadan önce grup dağılımının genel özelliklerini belirlemek için betimsel istatistik kullanılmıştır. Ayrı ayrı tüm alt gruplara yönelik madde sayısı, öğrenci sayısı, ortalama, standart sapma, çarpıklık, basıklık, en yüksek puan, en düşük puan ve Cronbach Alpha değerlerine ilişkin bilgiler verilmiştir. Betimsel istatistiğe yönelik işlemler yapılırken SPSS paket programından faydalanılmıştır.

1.alt problemlere ilişkin bulguların elde edilmesi aşamasında, maddelerin farklı kültür ve dil değişkenine göre DMF içerip içermediği MH, LR ve SIBTEST yöntemleri ile belirlenmeye çalışılmıştır.

MH yönteminde, bireyler değişkenler temel alınarak referans ve odak grubu olmak üzere iki gruba ayrılmış ve eşleştirilmiş gruplar oluşturulmuştur. Oluşturulan gruplar toplam test puanı üzerinden yetenek düzeylerine göre 5 yetenek seviyesine bölünmüştür (Osterlind & Everson, 2009). Aynı yetenek seviyesindeki bu iki grup ile madde arasında bir ilişki olmadığını belirten yokluk hipotezi test edilmiştir (Gierl vd., 2000).

H_0: Aynı yetenek düzeyindeki bireyler ile test puanı arasında anlamlı düzeyde bir ilişki yoktur.

Grupların yetenek düzeylerine göre eşleştirilmesinin ardından her yetenek seviyesi ve testte yer alan her bir madde için bir olasılık tablosu oluşturulmuştur. Tablo oluşturulurken bireylerin maddeye verdikleri doğru ve yanlış cevapların frekansları kullanılmıştır (Gierl vd., 2000). Aynı yetenek seviyesinde olup farklı alt gruplardan gelen kişilerin maddeye doğru cevap verme olasılıklarını karşılaştırabilmek için olasılık oranı hesaplanmıştır. α kestirim katsayısı adı verilen bu oran, referans gruptaki kişilerin maddeye doğru cevap verme olasılıklarının, odak gruptaki kişilerin maddeye doğru cevap verme olasılığına oranı şeklinde hesaplanmıştır. α_{MH} değeri DMF etki büyüklüğünün bir kestiricisidir. Tek başına yorumlanması zor olan bu değer logaritmik dönüşümlerle daha kolay yorumlanabilecek farklı bir ölçeğe dönüştürülmüştür. Doğal logaritması alınan α değeri -2,35 ile çarpılarak Δ_{MH} istatistiği hesaplanmıştır (Clauser & Mazor, 1998).

$$\Delta_{MH} = -2{,}35ln(\alpha MH)$$

Oluşturulan bu yeni katsayıda Δ_{MH} değerinin 0 olması DMF olmadığı anlamına gelmektedir. Eğer hesaplanan Δ_{MH} değeri pozitifse odak grup, negatifse referans grup lehine DMF bulunduğu tespit edilmiş olur (Holand & Thayer, 1988). Δ_{MH} değeri etki büyüklüğünü temsil ettiği için, aynı zamanda DMF içeren maddelerin düzeylerini yorumlamada da kullanılmıştır. Zieky (1993) tarafından DMF miktarının büyüklüğünü yorumlamak için oluşturulan ve yaygın olarak kullanılan sınıflama ölçütüne göre her bir madde için hesaplanan Δ_{MH} değerinin mutlak değeri 1'den küçükse DMF yok veya ihmal edilebilecek düzeyde, 1 ile 1,5 değerleri arasındaysa orta düzey DMF, 1,5'tan büyük ise yüksek düzey DMF vardır şeklinde yorumlanmıştır.

LR yönteminde model karşılaştırmasına dayalı olarak üç aşamalı modelleme kullanılmıştır. Bu modeller toplam puan, grup ve toplam puan ile grup etkileşimi değişkenlerinin hiyerarşik olarak modele eklenmesiyle oluşturulmuştur (Gierl vd., 2000).

Model 1: $Y = b_0 + b_1X$

Model 2: $Y = b_0 + b_1X + b_2G$

Model 3: $Y = b_0 + b_1X + b_2G + b_3X$*G

Model 1'de toplam puan değişkeni yer alırken Model 2'de grup değişkeni eklenmiştir. Bir sonraki aşama model 3'te ise daha önce eklenen değişkenlere ilaveten toplam puan ve grup etkileşimi değişkeni eklenmiştir. Her modelde ki-kare değerinin hesaplanmasının yanı sıra standartlaştırılmış regresyon katsayısı olan R^2 değeri R programı kullanılarak hesaplanmıştır. Bu değer etki büyüklüğü kestiricisi olarak kabul edilmiştir ve DMF miktarını temsil etmektedir. Etki büyüklüğü miktarı iki model arasındaki R^2 farkı (ΔR^2) hesaplanarak bulunmuştur. İkinci ve birinci modelden elde edilen ΔR^2 değerinin anlamlı olması tek biçimli DMF'nin, üçüncü ve ikinci modelden elde edilen ΔR^2 değerinin anlamlı olması ise tek biçimli olmayan DMF'nin varlığını işaret etmektedir. DMF düzeyinin belirlenmesi amacıyla etki büyüklüğü miktarının (ΔR^2) yorumlanması için ise Jodoin ve Gierl (2001) değer aralığı kullanılmıştır. Buna göre ΔR^2 değeri 0,035'den küçük ise A düzeyinde (yok veya ihmal edilebilir düzey) DMF, 0,035 ile 0,070 arasında ise B düzeyinde (orta düzey) DMF, 0,070'ten büyük ise C düzeyinde (yüksek düzey) DMF vardır şeklinde yorumlanmıştır.

SIBTEST yönteminde, testte yer alan maddeler geçerli alt test ve tahmin edilen alt test olmak üzere iki alt gruba ayrılmıştır. Geçerli alt testte, aynı yetenek grubundaki bireylerin puanları diğer grupla eşleştirmede kullanılmıştır. Tahmin edilen alt test ise DMF'nin

belirlenmesinde kullanılan maddeleri içerir (Bolt, 2000). Daha sonra her bir madde için istatiksel hipotezler test edilmiştir. H_0 hipotezi, maddeyi doğru cevaplama olasılığı açısından referans ve odak grupları arasında fark olmadığını, H_1 hipotezi ise fark olduğunu iddia etmektedir (Zhou, Gierl & Tan, 2006). R programı kullanılarak hesaplanan β_{UNI} değeri ise etki büyüklüğünü belirtmektedir ve şu şekilde tanımlanır;

$$\beta_{UNI} = \int \beta(\theta) f_F(\theta) d(\theta)$$

Elde edilen β_{UNI} değeri her madde için tek tek hesaplanır ve DMF miktarı olarak yorumlanır. Bu değer eğer pozitifse maddenin referans grubu lehine, negatif ise odak grubu lehine çalıştığı söylenir. DMF'nin düzeyini belirlenmek için ise Roussos ve Stout (1996) tarafından önerilen ölçütler kullanılmıştır. Buna göre elde edilen β_{UNI} değeri 0,059'dan küçükse DMF düzeyi A (DMF yoktur ya da ihmal edilebilir düzey); 0,059 ile 0,088 arasında ise DMF düzeyi B (orta düzey); 0,088'den büyükse DMF düzeyi C (yüksek düzey) olarak belirlenmiştir.

MH, LR ve SIBTEST yöntemlerine yönelik analizler R programı kullanılarak gerçekleştirilmiştir. Açık kaynak kodlu ve ücretsiz bir yazılım olması nedeniyle tercih edilen R, resmi web sitesinden (https://cran.r-project.org/) indirilmiştir. R'ın kullanıcı dostu bir arayüze sahip olmaması nedeniyle R Studio entegre programı da ilgili web sitesinden (RStudio, 2020) indirilerek kurulmuştur. R Studio, R programının konsol, kaynak düzenleme, grafik gibi bileşenlerini daha kullanışlı bir ortamda birleştirmeyi amaçlayan açık kaynak kodlu bir yazılımdır. R ve R Studio kurulumundan sonra bu programların Windows işletim sisteminde kullanıma hazır olması için R araçları (Rtools), ilgili siteden inidirilerek kurulumu gerçekleşrilmiştir (Using Rtools4.0 on Windows, 2020). DMF analizlerinin gerçekleştirilmesi adımında ise MH ve LR yöntemleri için difR paketi, SIBTEST yöntemi için ise mirt paketi kullanılmıştır (Alatlı & Şenel, 2020).

Yapılan analizler sonucunda DMF gösteren maddeler ile bu maddelerin DMF düzeyleri belirlenmiş ve sonuçlar tablolar kullanılarak gösterilmiştir. Son olarak kullanılan DMF belirleme yöntemlerinin DMF gösterdikleri maddeler bakımından sonuçlarının birbiriyle ne derece uyum içinde olduğu incelenmiştir. Yöntemlerden elde edilen sonuçların birbiriyle ne derece uyum içinde olduğunu belirlemek için Spearman Brown sıra farkları korelasyonundan faydalanılmıştır. Bu korelasyon, sıralı puanlar kullanılarak ölçülen iki değişken arasındaki doğrusal ilişkiyi açıklamaktadır. Eşit aralık ya da eşit oran düzeyindeki ölçümler analiz edilmesi gerektiğinde sıralı ölçeğe dönüştürülerek uygulanabilir. Spearman Brown sıra farkları korelasyonu şu durumlarda kullanılmaktadır; nesne ya da bireye ilişkin

yargılar sıra değerleri ile gösterildiğinde, değişkenlerden biri sıralı diğeri eşit aralıklı ya da oranlı ölçek düzeyinde olduğunda, en az eşit aralık düzeyinde olan ancak normallik varsayımının karşılanmadığı değişkenler var ise kullanılır (Tan, 2016). Bu çalışmada ise ülkeler, Türkiye-Fransa, Türkiye-Birleşik-Krallık, Fransa-Birleşik Krallık olmak üzere ikişerli olarak gruplandırılmış ve her bir grup için MH, LR ve SIBTEST yöntemlerine göre DMF analizleri yapılmıştır. Yapılan bu analizlerin sonuçlarına dayanarak yöntemlerin DMF'li olarak tespit ettiği maddeler 1, diğerleri 0 olacak şekilde kodlanmıştır. Her bir grup için yöntemlerin sonuçları MH-LR, MH-SIBTEST, LR-SIBTEST olmak üzere ikişerli gruplandırılarak DMF gösterdikleri maddeler bakımından sıralamaları alınıp Spearman Brown sıra farkları korelasyon katsayısı hesaplanmıştır. Örneğin; Türkiye – Fransa örneklemi için yapılan DMF analizleri sonuçlarının birbiriyle uyumuna bakarken MH-LR, MH-SIBTEST ve LR-SIBTEST olmak üzere üç kere Spearman Brown sıra farkları korelasyon katsayısı hesaplanmıştır. Spearman-Brown sıra farkları korelasyon katsayıları hesaplanırken SPSS paket programından faydalanılmıştır.

4.1 Çalışma Grubuna Yönelik Betimsel İstatistikler

Bu bölümde, çalışma grubuna yönelik betimsel istatistiklere yer verilmiştir. Bu çalışmada yapılan analizlerde, bireyselleştirilmiş test deseni göz önünde bulundurularak PISA 2018 uygulaması okuma becerileri alt testinde temel, 1.aşama ve 2.aşama bölümlerinde aynı demette yer alan madde paketleri ele alınmıştır. Temel bölüm için 1.paket (Core RC1), 1.Aşama için 1.paket (Stage 1 - R11H), 2.Aşama için 1.paket (Stage 2 - R21H) seçilmiştir. Çalışma grubu olarak ise Türkiye, Birleşik Krallık ve Fransa'dan bu paketleri yanıtlayan öğrenciler seçilmiştir. PISA 2018 uygulamasında Türkiye, Birleşik Krallık ve Fransa ülkelerinden yukarıda belirtilen madde paketlerini yanıtlayan 594 kişi çalışma grubuna dâhil edilmiştir. Katılımcıların verdikleri yanıtlar göz önünde bulundurularak SPSS paket programı aracılığı ile grubun tamamının ve alt grupların genel özellikleri belirlenmiştir. Hesaplanan betimsel istatistikler Tablo 6'da sunulmuştur.

Tablo 6

Çalışma Grubuna Yönelik Betimsel İstatistikler

	Türkiye	Birleşik Krallık	Fransa
Madde Sayısı	35	35	35
Öğrenci Sayısı	125	326	143
Ortalama	20,71	23,37	23,88
Standart Sapma	5,696	5,421	4,971
Çarpıklık	-0,349	-0,345	-0,198
Basıklık	-0,275	-0,488	-0,388
En Yüksek Puan	33	34	35
En Düşük Puan	6	7	11
Cronbach Alpha	0,807	0,800	0,762

Tablo 6 incelendiğinde en düşük test ortalaması 20,71 ile Türkiye'ye aittir. Fransa 23,88, Birleşik Krallık ise 23,37 test ortalamasına sahiptir. Standart sapma değerleri göz önüne alındığında, Türkiye 5,696 ile ilk sırada, Birleşik Krallık 5,421 ile ikinci sırada, Fransa ise 4,971 ile en düşük sapma değerine sahip ülke konumundadır. Çarpıklık ve basıklık değerlerini incelemek veri setinin normal dağılım gösterip göstermediğini belirlemede etkili yollardan biridir. Alan yazında çarpıklık ve basıklık değerlerinin +1,5 ile -1,5 arasında dağılım göstermesi normal dağılım olarak değerlendirilmektedir (Tabachnick & Fidell, 2015). Bu açıdan bakıldığında, Türkiye, Birleşik Krallık ve Fransa'nın veri setindeki basıklık ve çarpıklık değerlerinin normal dağılım şartını karşıladığına ulaşılmıştır. Ülkelerin testten aldıkları en yüksek puanlar incelendiğinde, Türkiye'nin 33, Birleşik Krallık'ın 34 ve Fransa'nın 35 puan aldığı görülmektedir. En düşük puanların ise Türkiye için 6, Birleşik Krallık için 7, Fransa için 11 puan olduğu görülmüştür. Veri seti, Cronbachs' Alpha iç tutarlılık katsayısı açısından incelendiğinde tüm ülkelerin 0,70 üzeri bir iç tutarlılık değerine sahip olduğu görülmüştür. Bu değer, araştırma kapsamında incelenen ülke veri setlerinin yeterli bir iç güvenirliğe sahip olduğunun göstergesidir.

BÖLÜM IV

BULGULAR VE YORUMLAR

Bu bölümde, araştırmanın alt problemlerine ilişkin bulgulara ve yorumlara yer verilmiştir. Yapılan analizler sonucunda her bir alt probleme yönelik elde edilen bulgular ayrı ayrı tablolaştırılarak yorumlanmıştır.

1. Birinci Alt Probleme İlişkin Bulgular Ve Yorumlar

Birinci Alt Problem: PISA 2018 okuma becerileri alt testinde yer alan maddeler, Türkiye, Birleşik Krallık ve Fransa örneği için MH, LR ve SIBTEST yöntemleriyle analiz edildiğinde DMF göstermekte midir?

1a. PISA 2018 okuma becerileri alt testinde yer alan maddeler, Türkiye ve Birleşik Krallık örneklemi için MH, LR ve SIBTEST yöntemleriyle analiz edildiğinde DMF göstermekte midir?

PISA 2018 okuma becerileri alt testinin farklı kültür ve dil değişkenine göre Türkiye ve Birleşik Krallık gruplarında MH analizi sonucunda DMF gösteren maddeler Tablo 7'de listelenmiştir.

Tablo 7

PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Birleşik Krallık Gruplarında MH Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler

Madde No	Karşılaştırılan Ülkeler	*A*	$\chi 2$	*p*	Δ_{MH}	DMF Düzeyi	Avantajlı Grup
2	TR - UK	10,555	72,713	0,000	-5,538	C	UK
3	TR - UK	3,166	13,345	0,000	-2,709	C	UK
5	TR - UK	2,441	8,519	0,004	-2,097	C	UK
7	TR - UK	0,179	25,301	0,000	4,043	C	TR
9	TR - UK	0,405	8,163	0,004	2,125	C	TR
13	TR - UK	2,533	11,489	0,001	-2,184	C	UK
16	TR - UK	2,214	8,635	0,003	-1,867	C	UK
20	TR - UK	0,435	8,105	0,004	1,956	C	TR
21	TR - UK	0,309	21,437	0,000	2,761	C	TR
25	TR - UK	0,445	9,820	0,002	1,905	C	TR
26	TR - UK	0,355	12,711	0,000	2,431	C	TR
29	TR - UK	0,394	8,281	0,004	2,191	C	TR
31	TR - UK	0,538	7,032	0,008	1,459	B	TR
35	TR - UK	1,685	4,640	0,031	-1,226	B	UK

Türkiye ve Birleşik Krallık örneklemlerinin karşılaştırılmasına yönelik yapılan MH analizinde, PISA 2018 uygulamasını kendi ulusal dilinde bir çeviriye dönüştürerek uygulayan taraf olması nedeniyle Türkiye dezavantajlı grup olarak kabul edilmiştir. Bu nedenle analizlerde Türkiye odak grup, Birleşik Krallık ise referans grup olarak belirlenmiştir. Analiz sonucunda, Tablo 7'de okuma becerileri alt testinde yer alan maddelerden *p* değeri 0,05 düzeyinde manidar çıkanlara yer verilmiştir. Buna göre Tablo 7'de toplamda 14 madde DMF göstermektedir.

Zieky (1993) tarafından DMF miktarının büyüklüğünü yorumlamak için oluşturulan sınıflama ölçütüne göre maddelerin Δ_{MH} değerleri ele alınarak DMF düzeyleri belirlenmiştir. Hesaplanan Δ_{MH} değeri pozitif ise odak grup, negatif ise referans grup lehine DMF bulunduğu tespit edilmektedir (Holand & Thayer, 1988). Buna göre Türkiye örnekleminin lehine olduğu görülen sekiz maddeden biri B düzeyinde (Orta Düzey), yedisi C düzeyinde (Yüksek Düzey) DMF göstermektedir. Birleşik Krallık örnekleminin lehine olan altı maddeden birinin B düzeyinde (Orta Düzey), beşinin C düzeyinde (Yüksek Düzey) DMF gösterdiği belirlenmiştir. Oluşturulan tabloda, DMF'li olduğu belirlenen

maddelerin hangi gruba avantaj sağladığına yönelik yapılan değerlendirme özeti ise Tablo 8'de sunulmuştur.

Tablo 8

Tablo 7'de Yapılan Değerlendirme Özeti

Avantajlı Grup	DMF Düzeyi			
	A	B	C	Toplam
Türkiye	-	1	7	8
Birleşik Krallık	-	1	5	6

PISA 2018 okuma becerileri alt testinin farklı kültür ve dil değişkenine göre Türkiye ve Birleşik Krallık gruplarında R programı kullanılarak yapılan LR analizi sonucunda DMF gösteren maddeler Tablo 9'da listelenmiştir.

Tablo 9

PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Birleşik Krallık Gruplarında LR Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler

Madde No	Karşılaştırılan Ülkeler	X^2	p	ΔR^2	DMF Düzeyi
2	TR - UK	84,042	0,000	0,201	C
3	TR - UK	17,785	0,000	0,058	B
5	TR - UK	8,153	0,017	0,022	A
7	TR - UK	33,493	0,000	0,104	C
9	TR - UK	11,424	0,003	0,036	B
13	TR - UK	13,297	0,001	0,030	A
16	TR - UK	10,777	0,005	0,024	A
20	TR - UK	9,023	0,011	0,021	A
21	TR - UK	26,301	0,000	0,064	B
23	TR - UK	8,924	0,012	0,024	A
24	TR - UK	7,251	0,027	0,026	A
25	TR - UK	15,598	0,000	0,042	B
26	TR - UK	17,685	0,000	0,035	B
28	TR - UK	9,026	0,011	0,028	A
29	TR - UK	10,238	0,006	0,037	B
31	TR - UK	11,005	0,004	0,028	A
33	TR - UK	7,383	0,025	0,020	A
35	TR - UK	6,888	0,032	0,019	A

Türkiye ve Birleşik Krallık örneklemlerinin karşılaştırılmasına yönelik yapılan LR analizinde Türkiye odak grup, Birleşik Krallık ise referans grup olarak belirlenmiştir.

Analiz sonucunda, Tablo 9'da okuma becerileri alt testinde yer alan maddelerden *p* değeri 0,05 düzeyinde manidar çıkanlara yer verilmiştir. Buna göre Tablo 9'da 18 maddede DMF olduğu görülmektedir.

Etki büyüklüğü miktarı iki model arasındaki R^2 farkı (ΔR^2) hesaplanarak bulunur (Zumbo, 1999). Buna göre Tablo 9'da yer alan maddelerin ΔR^2 değerleri ele alınarak DMF'li maddelerin etki büyüklüğü miktarı tespit edilmiştir. Etki büyüklüğünün yorumlanabilmesi için ise Jodoin ve Gierl (2001) tarafından önerilen ölçüt temel alınmıştır. Buna göre, DMF gösteren 18 maddeden ikisinin C düzeyinde (Yüksek Düzey), altısının B düzeyinde (Orta Düzey), 10'unun ise A düzeyinde (İhmal Edilebilir Düzey) olduğu belirlenmiştir. DMF'li maddelere yönelik değerlendirme özeti ise Tablo 10'da sunulmuştur.

Tablo 10

Tablo 9'da Yapılan Değerlendirme Özeti

	DMF Düzeyi			
	A	B	C	Toplam
DMF'li Madde Sayısı	10	6	2	18

PISA 2018 okuma becerileri alt testinin farklı kültür ve dil değişkenine göre Türkiye ve Birleşik Krallık gruplarında yapılan SIBTEST analizi sonucunda DMF gösteren maddeler Tablo 11'de listelenmiştir.

Tablo 11

PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Birleşik Krallık Gruplarında SIBTEST Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler

Madde No	Karşılaştırılan Ülkeler	β	SE	X^2	*P*	DMF Düzeyi	Avantajlı Grup
2	TR – UK	0,400	0,074	29,260	0,000	C	UK
3	TR – UK	0,317	0,069	21,039	0,000	C	UK
4	TR – UK	0,274	0,085	10,400	0,001	C	UK
9	TR – UK	-0,163	0,067	5,840	0,016	C	TR
13	TR – UK	0,168	0,065	6,641	0,010	C	UK
20	TR – UK	-0,185	0,067	7,564	0,006	C	TR
21	TR – UK	-0,221	0,079	7,792	0,005	C	TR
26	TR – UK	-0,232	0,079	8,568	0,003	C	TR
29	TR – UK	-0,154	0,071	4,641	0,031	C	TR
31	TR – UK	-0,160	0,061	6,760	0,009	C	TR
35	TR – UK	0,133	0,067	3,985	0,046	C	UK

Analizlerde Türkiye odak grup, Birleşik Krallık ise referans grup olarak belirlenmiştir. Tablo 11'de, SIBTEST analizi sonucunda *p* değeri 0,05 düzeyinde manidar çıkan maddelere yer verilmiştir. Buna göre Türkiye ve Birleşik Krallık örneklemlerinin karşılaştırılmasına yönelik yapılan SIBTEST analizi sonucunda 11 maddede DMF olduğu görülmektedir. Gerçekleştirilen analizler sonucunda ulaşılan β değeri etki büyüklüğünü ifade etmektedir. Elde edilen bu değerin pozitif bir değer alması maddenin referans gruba avantaj sağlandığı, negatif bir değer alması odak gruba avantaj sağlandığı yönünde DMF'nin gözlendiği şeklinde değerlendirilmektedir (Stout & Roussos, 1995). Buna göre altı madde Türkiye lehine, beş madde ise Birleşik Krallık lehine DMF göstermektedir. Bunun yanında β değerine göre DMF düzeyi de belirlenebilmektedir (Stout & Roussos, 1995). Buna göre Tablo 11'de yer alan 11 DMF'li maddenin tamamında yüksek düzeyde (C Düzeyi) DMF vardır. Tablo 11'de yapılan değerlendirme özetine Tablo 12'de yer verilmiştir.

Tablo 12

Tablo 11'de Yapılan Değerlendirme Özeti

Avantajlı Grup	DMF Düzeyi			
	A	B	C	Toplam
Türkiye	-	-	6	6
Birleşik Krallık	-	-	5	5

Türkiye ve Birleşik Krallık gruplarında yöntemlere göre DMF gösteren maddelerin yüzdelerine bakıldığında, 14 madde ile MH yöntemi maddelerin %40'ını, 18 madde ile LR yöntemi maddelerin %51'ini, 11 madde ile SIBTEST yöntemi maddelerin %31'ini DMF'li olarak tespit etmiştir. Buna göre en çok DMF'li madde tespit eden yöntem LR olurken en az DMF'li madde tespit eden yöntem ise SIBTEST olmuştur.

1b. PISA 2018 okuma becerileri alt testinde yer alan maddeler, Türkiye ve Fransa örneklemi için MH, LR ve SIBTEST yöntemleriyle analiz edildiğinde DMF göstermekte midir?

PISA 2018 okuma becerileri alt testinin farklı kültür ve dil değişkenine göre Türkiye ve Fransa gruplarında MH analizi sonucunda DMF gösteren maddeler Tablo 13'te listelenmiştir.

Tablo 13

PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Fransa Gruplarında MH Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler

Madde No	Karşılaştırılan Ülkeler	*A*	$\chi 2$	*p*	Δ_{MH}	DMF Düzeyi	Avantajlı Grup
2	TR – FR	6,780	31,453	0,000	-4,498	C	FR
5	TR – FR	2,428	7,124	0,008	-2,085	C	FR
13	TR – FR	2,214	5,767	0,016	-1,868	C	FR
14	TR – FR	2,264	6,593	0,010	-1,921	C	FR
16	TR – FR	2,276	6,111	0,013	-1,933	C	FR
21	TR – FR	0,312	12,790	0,000	2,735	C	TR

Türkiye ve Fransa örneklemlerinin karşılaştırılmasına yönelik yapılan MH analizinde, PISA 2018 uygulamasını kendi ulusal dilinde bir çeviriye dönüştürerek uygulayan taraf olması nedeniyle Türkiye dezavantajlı grup olarak kabul edilmiştir. Bu nedenle analizlerde Türkiye odak grup, Fransa ise referans grup olarak belirlenmiştir. Analiz sonucunda, Tablo

13'te okuma becerileri alt testinde yer alan maddelerden *p* değeri 0,05 düzeyinde manidar çıkanlara yer verilmiştir. Buna göre Tablo 13'te altı maddede DMF olduğu görülmektedir. Hesaplanan Δ_{MH} değeri pozitif ise odak grup, negatif ise referans grup lehine DMF bulunduğu tespit edilmektedir (Holand & Thayer, 1988). Buna göre Tablo 13'te yer alan maddelerden biri Türkiye, beşi ise Fransa örnekleminin lehine DMF göstermektedir.

DMF miktarının büyüklüğünü yorumlamak için maddelerin Δ_{MH} değerleri dikkate alınarak, Zieky (1993) tarafından oluşturulan sınıflama ölçütüne göre DMF düzeyleri belirlenmiştir (Holand & Thayer, 1988). Buna göre Tablo 13'te yer alan DMF'li maddelerin tamamının C düzeyinde (Yüksek Düzey) olduğu görülmektedir. DMF'li olduğu belirlenen maddelerin hangi gruba avantaj sağladığına yönelik yapılan değerlendirme sonucu ise Tablo 14'te özetlenmiştir.

Tablo 14

Tablo 13'te Yapılan Değerlendirme Özeti

Avantajlı Grup	DMF Düzeyi			Toplam
	A	B	C	
Türkiye	-	-	1	1
Fransa	-	-	5	5

PISA 2018 okuma becerileri alt testinin farklı kültür ve dil değişkenine göre Türkiye ve Fransa gruplarında LR analiz sonucunda DMF gösteren maddeler Tablo 15'te listelenmiştir.

Tablo 15

PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Fransa Gruplarında LR Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler

Madde No	Karşılaştırılan Ülkeler	X^2	*p*	ΔR^2	DMF Düzeyi
2	TR – FR	39,626	0,000	0,150	C
5	TR – FR	11,310	0,004	0,051	B
13	TR – FR	8,655	0,013	0,033	A
14	TR – FR	11,485	0,003	0,043	B
21	TR – FR	12,901	0,002	0,054	B
27	TR – FR	7,442	0,024	0,034	A
31	TR – FR	12,028	0,002	0,052	B
33	TR – FR	10,133	0,006	0,046	B

Türkiye ve Fransa örneklemlerinin karşılaştırılmasına yönelik yapılan LR analizinde Türkiye odak grup, Fransa ise referans grup olarak belirlenmiştir. Analiz sonucunda, Tablo 15'te okuma becerileri alt testinde yer alan maddelerden *p* değeri 0,05 düzeyinde manidar çıkanlara yer verilmiştir. Buna göre sekiz maddede DMF olduğu görülmektedir.

Etki büyüklüğü miktarını değerlendirmek maddelerin ΔR^2 değerleri ele alınmıştır. Etki büyüklüğünün yorumlanabilmesi için ise Jodoin ve Gierl (2001) tarafından önerilen ölçüt temel alınmıştır. Buna göre DMF gösteren sekiz maddeden birinin C düzeyinde (Yüksek Düzey), beşinin B düzeyinde (Orta Düzey), ikisinin ise A düzeyinde (İhmal Edilebilir Düzey) olduğu belirlenmiştir. Tablo 15'te yapılan değerlendirmenin öz eti Tablo 16'da gösterilmiştir.

Tablo 16

Tablo 15'te Yapılan Değerlendirme Özeti

	DMF Düzeyi			
	A	B	C	Toplam
DMF'li Madde Sayısı	2	5	1	8

PISA 2018 okuma becerileri alt testinin farklı kültür ve dil değişkenine göre Türkiye ve Fransa gruplarında yapılan SIBTEST analizi sonucunda DMF gösteren maddeler Tablo 17'de listelenmiştir

Tablo 17

PISA 2018 Okuma Becerileri Alt Testinin Türkiye ve Fransa Gruplarında SIBTEST Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler

Madde No	Karşılaştırılan Ülkeler	β	SE	X^2	P	DMF Düzeyi	Avantajlı Grup
2	TR - FR	0,306	0,086	12,745	0,000	C	FR
14	TR - FR	0,232	0,086	7,261	0,007	C	FR
20	TR - FR	0,175	0,078	5,037	0,025	C	FR
21	TR - FR	-0,222	0,091	5,988	0,014	C	TR
28	TR - FR	-0,169	0,085	3,979	0,046	C	TR

Analizlerde Türkiye odak grup, Fransa referans grup olarak belirlenmiştir. Tablo 17'de, SIBTEST analizi sonucunda *p* değeri 0,05 düzeyinde manidar çıkan maddelere yer verilmiştir. Buna göre Türkiye ve Birleşik Krallık örneklemlerinin karşılaştırılmasına

yönelik yapılan SIBTEST analizi sonucunda beş maddede DMF olduğu görülmektedir. Gerçekleştirilen analizler sonucunda ulaşılan β değeri etki büyüklüğünü ifade etmektedir. Elde edilen bu değerin pozitif bir değer alması maddenin referans gruba avantaj sağlandığı, negatif bir değer alması odak gruba avantaj sağlandığı yönünde DMF'nin gözlendiği şeklinde değerlendirilmektedir. Buna göre DMF'li olduğu beirlenen beş maddenin üçü Fransa, ikisi Türkiye örneklemi lehinedir. Bunun yanında β değerine göre DMF düzeyi de belirlenebilmektedir (Stout & Roussos, 1995). Buna göre DMF'li olduğu tespit edilen tüm maddelerde C düzeyinde (yüksek düzey) DMF görülmektedir. Tablo 17'de yapılan değerlendirme özeti tablo 18'de gösterilmiştir.

Tablo 18

Tablo 17'de Yapılan Değerlendirme Özeti

Avantajlı Grup	DMF Düzeyi			Toplam
	A	B	C	
Türkiye	-	-	2	2
Fransa	-	-	3	3

Türkiye ve Fransa gruplarında yöntemlere göre DMF gösteren maddelerin yüzdelerine bakıldığında, altı madde ile MH yöntemi maddelerin %18'ini, sekiz madde ile LR yöntemi maddelerin %8'ini, beş madde ile SIBTEST yöntemi maddelerin %14'ünü DMF'li olarak tespit etmiştir. Buna göre en çok DMF'li madde tespit eden yöntem LR olurken en az DMF'li madde tespit eden yöntem ise SIBTEST olmuştur.

1c. PISA 2018 okuma becerileri alt testinde yer alan maddeler, Birleşik Krallık ve Fransa örneklemi için MH, LR ve SIBTEST yöntemleriyle analiz edildiğinde DMF göstermekte midir?

PISA 2018 okuma becerileri alt testinin farklı kültür ve dil değişkenine göre Birleşik Krallık ve Fransa gruplarında MH analizi sonucunda DMF gösteren maddeler Tablo 19'de listelenmiştir.

Tablo 19

PISA 2018 Okuma Becerileri Alt Testinin Fransa ve Birleşik Krallık Gruplarında MH Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler

Madde No	Karşılaştırılan Ülkeler	α	$\chi 2$	P	Δ_{MH}	DMF Düzeyi	Avantajlı Grup
3	FR - UK	2,655	9,076	0,003	-2,294	C	FR
7	FR - UK	0,210	20,071	0,000	3,666	C	UK
9	FR - UK	0,428	7,222	0,007	1,992	C	UK
20	FR - UK	0,300	17,647	0,000	2,828	C	UK
22	FR - UK	1,908	8,255	0,004	-1,519	C	FR
26	FR - UK	0,578	3,907	0,048	1,289	B	UK
27	FR - UK	1,764	6,306	0,012	-1,334	B	FR
33	FR - UK	1,643	5,348	0,021	-1,167	B	FR
35	FR - UK	2,004	9,507	0,002	-1,634	C	FR

Birleşik Krallık ve Fransa örneklemlerinin karşılaştırılmasına yönelik yapılan MH analizinde Birleşik Krallık odak grup, Fransa ise referans grup olarak belirlenmiştir. Analiz sonucunda, Tablo 19'da okuma becerileri alt testinde yer alan maddelerden *p* değeri 0,05 düzeyinde manidar çıkanlara yer verilmiştir. Buna göre Tablo 19'da DMF'li olduğu tespit edilen dokuz madde yer almaktadır.

Hesaplanan Δ_{MH} değeri pozitif ise odak grup, negatif ise referans grup lehine DMF bulunduğu tespit edilmektedir (Holand & Thayer, 1988). Buna göre Tablo 19'da yer alan maddelerden beşinin Fransa, dördünün ise Birleşik Krallık örneklemi lehine olduğu görülmektedir. Zieky (1993) tarafından DMF miktarının büyüklüğünü yorumlamak için oluşturulan sınıflama ölçütüne göre Δ_{MH} değerine bakıldığında, Fransa örneklemi lehine olan beş maddeden ikisinin B düzeyinde (Orta Düzey), üçünün C düzeyinde (Yüksek Düzey) DMF gösterdiği görülmektedir. Birleşik Krallık örneklemi lehine olan dört maddeden ise biri B düzeyinde (Orta Düzey), üçü C düzeyinde (Yüksek Düzey) DMF göstermektedir. DMF'li olduğu belirlenen maddelere yönelik yapılan değerlendirme sonucu ise Tablo 20'de özetlenmiştir.

Tablo 20

Tablo 19'da Yapılan Değerlendirme Özeti

Avantajlı Grup	DMF Düzeyi			Toplam
	A	B	C	
Fransa	-	2	3	5
Birleşik Krallık	-	1	3	4

PISA 2018 okuma becerileri alt testinin farklı kültür ve dil değişkenine göre Birleşik Krallık ve Fransa gruplarında LR analiz sonucunda DMF gösteren maddeler Tablo 21'de listelenmiştir.

Tablo 21

PISA 2018 Okuma Becerileri Alt Testinin Fransa ve Birleşik Krallık Gruplarında LR Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler

Madde No	Karşılaştırılan Ülkeler	X^2	P	ΔR^2	DMF Düzeyi
3	FR - UK	11,011	0,004	0,040	B
7	FR - UK	27,995	0,000	0,082	C
9	FR - UK	8,231	0,016	0,026	A
14	FR - UK	6,559	0,038	0,015	A
19	FR - UK	6,372	0,041	0,031	A
20	FR - UK	22,222	0,000	0,049	B
22	FR - UK	8,406	0,015	0,022	A
26	FR - UK	7,515	0,023	0,015	A
27	FR - UK	9,307	0,010	0,026	A
28	FR - UK	7,075	0,029	0,022	A
29	FR - UK	7,906	0,019	0,026	A
33	FR - UK	6,296	0,043	0,018	A
35	FR - UK	19,773	0,000	0,053	B

Fransa ve Birleşik Krallık örneklemlerinin karşılaştırılmasına yönelik yapılan LR analizinde Fransa odak grup, Birleşik Krallık ise referans grup olarak belirlenmiştir. Analiz sonucunda, Tablo 21'de okuma becerileri alt testinde yer alan maddelerden *p* değeri 0,05 düzeyinde manidar çıkanlara yer verilmiştir. Buna göre Tablo 21'de 13 maddede DMF olduğu görülmektedir.

Etki büyüklüğü miktarı iki model arasındaki R^2 farkı (ΔR^2) hesaplanarak bulunur (Zumbo, 1999). Buna göre Tablo 21'de yer alan maddelerin ΔR^2 değerlerine bakılarak DMF'li maddelerin etki büyüklüğü miktarı tespit edilmiştir. Etki büyüklüğünün yorumlanabilmesi

için ise Jodoin ve Gierl (2001) tarafından önerilen ölçüt temel alınmıştır. Buna göre DMF gösteren 13 maddeden birinin C düzeyinde (Yüksek Düzey), üçünün B düzeyinde (Orta Düzey), dokuzunun ise A düzeyinde (İhmal Edilebilir Düzey) olduğu belirlenmiştir. Tablo 21'e yönelik yapılan değerlendirme özeti tablo 22'de gösterilmiştir.

Tablo 22

Tablo 21'de Yapılan Değerlendirme Özeti

	DMF Düzeyi			
	A	B	C	Toplam
DMF'li Madde Sayısı	9	3	1	13

PISA 2018 okuma becerileri alt testinin farklı kültür ve dil değişkenine göre Fransa ve Birleşik Krallık gruplarında yapılan SIBTEST analizi sonucunda DMF gösteren maddeler Tablo 23'te listelenmiştir

Tablo 23

PISA 2018 Okuma Becerileri Alt Testinin Fransa ve Birleşik Krallık Gruplarında SIBTEST Yöntemi İle Analiz Edilmesi Sonucunda DMF'li Bulunan Maddeler

Madde No	Karşılaştırılan Ülkeler	β	SE	X^2	P	DMF Düzeyi	Avantajlı Grup
3	FR - UK	0,160	0,067	5,731	0,017	C	FR
7	FR - UK	-0,145	0,061	5,566	0,018	C	UK
9	FR - UK	-0,163	0,056	8,450	0,004	C	UK
20	FR - UK	-0,222	0,060	13,787	0,000	C	UK
22	FR - UK	0,144	0,062	5,327	0,021	C	FR
28	FR - UK	0,137	0,063	4,744	0,029	C	FR
29	FR - UK	-0,135	0,049	7,624	0,006	C	UK
35	FR - UK	0,192	0,061	9,889	0,002	C	FR

Analizlerde Fransa odak grup, Birleşik Krallık referans grup olarak belirlenmiştir. Tablo 23'te SIBTEST analizi sonucunda *p* değeri 0,05 düzeyinde manidar çıkan maddelere yer verilmiştir. Buna göre Fransa ve Birleşik Krallık örneklemlerinin karşılaştırılmasına yönelik yapılan SIBTEST analizi sonucunda sekiz maddede DMF olduğu görülmektedir. Gerçekleştirilen analizler sonucunda ulaşılan β değeri etki büyüklüğünü ifade etmektedir. Elde edilen bu değerin pozitif bir değer alması maddenin referans gruba avantaj sağlandığı, negatif bir değer alması odak gruba avantaj sağlandığı yönünde DMF'nin gözlendiği şeklinde değerlendirilmektedir Bunun yanında β değerine göre DMF düzeyi de

belirlenebilmektedir (Stout & Roussos, 1995). Buna göre DMF'li olduğu tespit edilen sekiz maddede de yüksek düzeyde (C Düzeyi) DMF vardır. Bu maddelerden dördü odak grup olan Birleşik Krallık, dördü ise referans grup olan Fransa lehine çalışmaktadır. Tablo 23'te yapılan değerlendirmeye ilişkin özet tablo 24'te gösterilmiştir.

Tablo 24

Tablo 23'te Yapılan Değerlendirme Özeti

Avantajlı Grup	DMF Düzeyi			
	A	B	C	Toplam
Fransa	-	-	4	4
Birleşik Krallık	-	-	4	4

Fransa ve Birleşik Krallık gruplarında yöntemlere göre DMF gösteren maddelerin yüzdelerine bakıldığında, dokuz madde ile MH yöntemi maddelerin %26'sını, 13 madde ile LR yöntemi maddelerin %37'sini, sekiz madde ile SIBTEST yöntemi maddelerin %23'ünü DMF'li olarak tespit etmiştir. Buna göre en çok DMF'li madde tespit eden yöntem LR olurken en az DMF'li madde tespit eden yöntem ise SIBTEST olmuştur.

2.İkinci Alt Probleme İlişkin Bulgular Ve Yorumlar

MH, LR ve SIBTEST yöntemleriyle yapılan analizlerin sonuçları birbiriyle uyum göstermekte midir?

Çalışmanın birinci alt probleminde, PISA 2018 okuma becerileri alt testinde yer alan maddelerin, Türkiye - Birleşik Krallık, Türkiye – Fransa ve Fransa – Birleşik Krallık grupları arasında MH, LR ve SIBTEST yöntemleriyle analiz edildiğinde DMF gösterip göstermediği incelenmiştir. Buna göre yöntemlerin belirledikleri DMF'li madde sayısı Tablo 25'te gösterilmiştir.

Tablo 25

Farklı Yöntemlerin Belirledikleri DMF'li Madde Sayısı

	DMF'li Madde Sayısı		
	MH	LR	SIBTEST
Türkiye - Birleşik Krallık	14	18	11
Türkiye – Fransa	6	8	5
Fransa - Birleşik Krallık	9	13	8

Tablo 25'e göre tüm gruplarda en çok DMF'li madde tespit eden yöntem LR olurken en az DMF'li madde tespit eden yöntem ise SIBTEST olmuştur. Gruplar arasında ise en çok DMF'li madde Türkiye ve Birleşik Krallık grubunda tespit edilirken en az DMF'li madde Türkiye ve Fransa grubunda tespit edilmiştir.

DMF belirlemek için kullanılan analiz yöntemlerinin, DMF'li olarak belirledikleri maddeler bakımından sonuçlarının birbiriyle ne derece uyum içinde olduğu incelenmiştir. Yöntemler arasındaki uyumu belirlemek için Spearman Brown sıra farkları korelasyon katsayısından faydalanılmıştır. Ülkeler, Türkiye-Fransa, Türkiye-Birleşik-Krallık, Fransa-Birleşik Krallık olmak üzere ikişerli olarak gruplandırılmış ve her bir grup için MH, LR ve SIBTEST yöntemlerine göre DMF analizleri yapılmıştır. SPSS paket programı kullanılarak hesaplanan Spearman Brown sıra farkları korelasyon katsayısına ait çıktılar Ek 5'te sunulmuştur. DMF belirleme yöntemlerinin gruplara göre hesaplanan korelasyon katsayısı Tablo 26'da yer almaktadır.

Tablo 26

DMF Belirleme Yöntemlerinin Gruplara Göre Hesaplanan Spearman Brown Sıra Farkları Korelasyon Katsayıları

	MH-LR		MH-SIBTEST		LR-SIBTEST	
	r_s	p	r_s	p	r_s	P
TR-UK	0,793	0,000	0,704	0,000	0,535	0,001
TR-FR	0,655	0,000	0,464	0,005	0,361	0,033
FR-UK	0,765	0,000	0,614	0,000	0,708	0,000

Tablo 26'dan yola çıkarak Guilford (1956)'un belirttiği korelasyon yorumlama kriterlerine göre Türkiye ve Birleşik Krallık grubunda p=0,000 manidarlık düzeyinde 0,793 korelasyon katsayısı ile MH-LR ve 0,704 korelasyon katsayısı ile MH-SIBTEST yöntemleri arasında

yüksek düzeyde bir ilişki görülmektedir. LR-SIBTEST yöntemleri arasında ise p=0,001 manidarlık düzeyinde 0,535 korelasyon katsayısı ile orta düzeyde bir ilişki bulunmaktadır. Türkiye – Fransa grubunda ise p=0,000 manidarlık düzeyinde 0,655 korelasyon katsayısı ile MH-LR ve p=0,005 manidarlık düzeyinde 0,464 korelasyon katsayısı ile MH-SIBTEST yöntemleri arasında orta düzeyde bir ilişki görülmektedir. LR-SIBTEST yöntemleri arasında ise p=0,033 manidarlık düzeyinde 0,361 korelasyon katsayısı ile zayıf düzeyde bir ilişki bulunmaktadır. Fransa – Birleşik Krallık grubunda ise p=0,000 manidarlık düzeyinde 0,765 korelasyon katsayısı ile MH-LR ve 0,708 korelasyon katsayısı ile LR-SIBTEST yöntemleri arasında yüksek düzeyde bir ilişki görülmektedir. MH-SIBTEST yöntemleri arasında ise p=0,000 manidarlık düzeyinde 0,614 korelasyon katsayısı ile orta düzeyde bir ilişki bulunmaktadır.

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde, araştırma soruları kapsamında çalışmadan elde edilen bulgulara dayalı olarak elde edilen sonuçlara yer verilmiştir. Ardından tartışma ve çalışmanın sonuçları ışığında gelecekte yapılacak olan araştırmalara yönelik önerilere yer verilmiştir.

1. Sonuç ve Tartışma

Bu çalışmada PISA 2018 okuma becerileri alt testinde yer alan maddelerde farklı kültür ve dil değişkeni ele alınarak DMF araştırılmıştır. DMF belirleme yöntemlerinden MH, LR ve SIBTEST kullanılmıştır. Elde edilen bulgular ışığında alt problemlere yönelik sonuç ve tartışmalar aşağıda sıralanmıştır.

1.1. Birinci Alt Probleme İlişkin Sonuç ve Tartışma

Çalışmanın birinci alt probleminde, PISA 2018 okuma becerileri alt testinde bulunan maddelerin, Türkiye - Birleşik Krallık, Türkiye – Fransa ve Fransa – Birleşik Krallık grupları arasında MH, LR ve SIBTEST yöntemleriyle analiz edildiğinde DMF gösterip göstermediği incelenmiştir.

Türkiye ve Birleşik Krallık örneklemlerinin karşılaştırılmasına yönelik yapılan MH analizinde, 35 maddenin 14'ünde DMF tespit edilmiştir. Tespit edilen maddelerin ikisi B düzeyinde, 12'si ise C düzeyinde DMF göstermektedir. LR yöntemi kullanılarak yapılan analizde ise 18 maddede DMF tespit edilmiştir. Tespit edilen maddelerin ikisinin C düzeyinde, altısının B, 10'unun ise A düzeyinde olduğu belirlenmiştir. SIBTEST

yöntemiyle yapılan analizde 11 maddede DMF olduğu görülmektedir. Tespit edilen maddelerin tamamının C düzeyinde olduğu belirlenmiştir.

Türkiye ve Fransa örneklemlerinin karşılaştırılmasına yönelik yapılan MH analizinde altı maddede DMF tespit edilmiştir. Tespit edilen maddelerin tamamı C düzeyinde DMF göstermektedir. LR yöntemi kullanılarak yapılan analizde ise sekiz maddede DMF tespit edilmiştir. Tespit edilen maddelerin birinin C düzeyinde, beşinin B düzeyinde, ikisinin ise A düzeyinde olduğu belirlenmiştir. SIBTEST yöntemiyle yapılan analizde beş maddede DMF olduğu görülmektedir. Tespit edilen maddelerin tamamının C düzeyinde olduğu belirlenmiştir.

Fransa ve Birleşik Krallık örneklemlerinin karşılaştırılmasına yönelik yapılan MH analizinde dokuz maddede DMF tespit edilmiştir. Tespit edilen maddelerin üçü B düzeyinde, altısı ise C düzeyinde DMF göstermektedir. LR yöntemi kullanılarak yapılan analizde ise 13 maddede DMF tespit edilmiştir. Tespit edilen maddelerin birinin C düzeyinde, üçünün B düzeyinde, dokuzunun ise A düzeyinde olduğu belirlenmiştir. SIBTEST yöntemiyle yapılan analizde sekiz maddede DMF olduğu görülmektedir. Tespit edilen maddelerin tamamının C düzeyinde olduğu belirlenmiştir.

Farklı kültür ve dil özelliklerine sahip Türkiye - Birleşik Krallık, Türkiye - Fransa ve Fransa - Birleşik Krallık grupları arasında MH, LR ve SIBTEST yöntemleriyle yapılan analizler sonucunda ihmal edilemeyecek kadar fazla sayıda DMF'li madde tespit edilmiştir. Bu sonuç, Demir (2013)'in çalışmasıyla benzerlik göstermektedir. Demir (2013) çalışmasında, PISA 2009 matematik okuryazarlığı alt testinde yer alan maddeleri Türkiye, Almanya, Finlandiya ve ABD örneklemlerinde farklı kültür değişkeni bakımından DMF gösterip göstermediğini MH, LR ve SIBTEST yöntemleriyle incelemiş ve Matematik okuryazarlığı alt testinde bulunan 21 maddenin yaklaşık olarak yarısında DMF içeren madde bulunduğunu tespit etmiştir. Benzer şekilde, Asil ve Gelbal (2012) da çalışmalarında, PISA 2006 kapsamında uygulanan öğrenci anketini kültür ve dil değişkenine göre DMF gösterip göstermediğini farklı örneklemler üzerinde karşılaştırmalı olarak incelemiş ve farklı kültür ve dil özellikleri arttıkça DMF gösteren madde sayısının arttığı sonucuna ulaşmıştır. Benzer şekilde Arslan (2020) çalışmasında 2016 yılında yapılan TEOG İngilizce alt testinde yer alan maddelerin, cinsiyet ve öğrenim görülen okul türü (devlet okulu – özel okul) değişkenleri bakımından DMF gösterip göstermediğini LR, SIBTEST ve MH yöntemlerinden faydalanarak incelemiştir. Analizler sonucunda 20 maddelik testin cinsiyet değişkenine göre dört, okul türü değişkenine göre ise bir maddenin

DMF'li olduğunu tespit etmiştir. Köse (2015) araştırmasında PISA 2009 matematik okuryazarlığı alt testinde yer alan maddeleri LR, MH ve SIBTEST yöntemlerini kullanarak cinsiyet ve farklı kültür değişkenleri bakımından DMF içerip içermediğini incelemiştir. Bunun sonucunda cinsiyet değişkenine göre beş, kültür değişkenine göre ise farklı gruplarda farklı yöntemlere göre çok sayıda DMF'li madde tespit etmiştir. Ayrıca, Ercikan ve Koh (2005), Grisay ve Monseur (2007) ve Atalay - Kabasakal (2010)'ın yaptığı çalışmalar da farklı kültür ve dil özellikleri arttıkça DMF gösteren madde sayısının arttığı sonucuyla bu çalışma ile benzerlik göstermektedir.

Araştırmada yer alan ülkeler, Türkiye-Fransa, Türkiye-Birleşik-Krallık, Fransa-Birleşik Krallık olmak üzere ikişerli olarak gruplandırılmıştır. Gruplandırılan ülkeler odak ve referans grup olarak ele alınmıştır. Bu durumda araştırmanın örneklem büyüklükleri sırasıyla "125-143", "125-326", "143-326" şeklinde olmuştur. PISA 2018 uygulamasında bireyselleştirilmiş test deseni uygulanması nedeniyle aynı madde paketlerini yanıtlayan kişi sayısı daha önceki uygulamalarda aynı kitapçık türünü yanıtlayan kişi sayısına göre nispeten daha azdır (MEB, 2019). DMF çalışmalarında, odak ve referans gruplarının örneklem büyüklüklerinin küçük olduğu durumlarda DMF belirleme yöntemlerinin gücünün belirgin bir şekilde düştüğü belirlenmiştir (Güler & Penfield, 2009). Ayrıca odak ve referans gruplarının örneklem büyüklüklerinin eşit olmadığı durumlarda DMF belirleme yöntemlerinin performanslarının olumsuz yönde etkilendiği belirlenmiştir (Kristjansson, Aylesworth, Mcdowell & Zumbo, 2005). Örneklem büyüklükleri eşit ya da birbirlerine yakın değerler almadığında DMF belirleme yöntemlerinin gücünün düştüğü tespit edilmiştir (Broer, Lee, Rizavi & Powers, 2005; Mazor, Clauser & Hambleton, 1992; Paek & Guo, 2011; Zieky, 1993). Finch ve French (2007) çalışmasında SIBTEST, LR, MTK Olabilirlik Oran ve Doğrulayıcı Faktör Analizi yöntemlerinin performanslarını farklı koşullar altında karşılaştırmışlardır. Örneklem büyüklüğü arttıkça SIBTEST, MTK Olabilirlik Oran ve Lojistik Regresyon yöntemlerinde güç oranlarının arttığını tespit etmiştir. Ayva Yörü (2019) çalışmasında MH, Lord'un $\chi 2$, LR, SIBTEST ve Raju'nun alan ölçümü yöntemlerini kullanılarak DMF belirleme yöntemlerinin performanslarını farklı koşullar altında incelemiştir. Gerçekleştirdiği simülasyon çalışması sonucunda örneklem büyüklüğü arttıkça yöntemlere göre I. Tip hata ve güç oranlarının da yükseldiğini tespit etmiştir. Odak ve referans grup örneklem büyüklükleri birbirine eşitken I. Tip hata ve güç oranlarının nispeten daha yüksek olduğunu gözlemlemiştir. Bu çalışmalardan farklı olarak Ö. Sünbül ve Sünbül (2016)çalışmasında MH, LR, Lord'un $\chi 2$ ve Raju'nun işaretsiz alan

ölçüleri yöntemlerini kullanarak farklı örneklem büyüklüğü (500, 1000, 1500, 2000 ve 3000), referans grubun tüm örnekleme oranı (0,50 ve 0,75) ve bazı belirli koşullar altında I. Tip hata ve güç oranlarını belirlemeyi amaçlamıştır. Araştırma sonucunda örneklem büyüklüğü ve örneklem büyüklüğü oranı koşullarında MH ve LR yöntemlerinde I. Tip hata oran ortalamalarında önemli değişiklik olmadığını gözlemlemişlerdir. Lord'un $\chi 2$ ve Raju'nun işaretsiz alan ölçülerinde ise örneklem büyüklüğü arttıkça I. Tip hata oranlarının azaldığını ve grupların örneklem büyüklüklerinin eşit olduğu durumda I. Tip hata ortalamasının daha düşük olduğunu saptamışlardır. Benzer şekilde Kim (2010) çalışmasında MH, LR, Test-DFIT ve Lord'un $\chi 2$ yöntemlerinin örneklem büyüklüğü ve bazı belirli koşullar altında I. Tip hata oranlarını araştırmıştır. Örneklem büyüklüğü oranlarının eşit ve büyük olmasının dört yöntemin hepsinde de I. Tip hata oranlarını şişirme eğiliminde olduğunu tespit etmiştir.

1.2. İkinci Alt Probleme İlişkin Sonuç ve Tartışma

Araştırmanın ikinci alt probleminde MH, LR ve SIBTEST yöntemleriyle yapılan DMF analiz sonuçlarının birbiriyle ne derece uyum gösterdiği incelenmiştir. Buna göre MH - LR yöntemleri arasındaki uyum incelendiğinde p=0,000 manidarlık düzeyinde sırasıyla Türkiye – Birleşik Krallık örnekleminde 0,793 korelasyon katsayısı ile yüksek düzeyde bir ilişki, Türkiye - Fransa örnekleminde 0,655 korelasyon katsayısı ile orta düzeyde bir ilişki, Fransa – Birleşik Krallık örnekleminde 0,765 korelasyon katsayısı ile yüksek düzeyde bir ilişki olduğu görülmüştür.

MH - SIBTEST yöntemleri arasındaki uyum incelendiğinde p=0,000 manidarlık düzeyinde sırasıyla TR-UK örnekleminde 0,704 korelasyon katsayısı ile yüksek düzeyde bir ilişki, TR-FR örnekleminde p=0,005 manidarlık düzeyinde 0,464 korelasyon katsayısı ile orta düzeyde bir ilişki, Fransa – Birleşik Krallık örnekleminde 0,765 korelasyon katsayısı ile yüksek düzeyde bir ilişki olduğu görülmüştür.

LR – SIBTEST yöntemleri arasındaki uyum incelendiğinde p=0,001 manidarlık düzeyinde sırasıyla Türkiye – Birleşik Krallık örnekleminde 0,535 korelasyon katsayısı ile orta düzeyde bir ilişki, Türkiye - Fransa örnekleminde p=0,033 manidarlık düzeyinde 0,361 korelasyon katsayısı ile düşük düzeyde bir ilişki, Fransa – Birleşik Krallık örnekleminde 0,708 korelasyon katsayısı ile yüksek düzeyde bir ilişki olduğu görülmüştür.

Genel olarak bakıldığında her ne kadar yöntemlerin DMF'li olarak tespit ettiği madde sayıları farklı olsa da DMF'li olarak belirledikleri maddelerin çoğunlukla birbiriyle benzer

olduğu tespit edilmiştir. Bunun için kullanılan Spearman Brown sıra farkları koreslasyon katsayısına bakıldığında yöntemler arasında bir paralellik olduğu söylenebilir. Doğan ve Öğretmen (2008) de 2003 yılında uygulanan Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı'nın (OKÖSYS) fen bilgisi alt testini Ki-kare, MH ve LR yöntemlerini kullanarak DMF açısından incelediği çalışmasında benzer sonuca ulaşmıştır. Yöntemlerin DMF'li olarak tespit ettiği madde sayısı farklı olsa da benzer sonuçlar verdiğini tespit etmişlerdir.

Yöntemler arasındaki en yüksek uyum MH – LR arasında görülürken en düşük uyum LR – SIBTEST arasında görülmektedir. Benzer şekilde Gierl, Jodoin ve Ackerman (2000) çalışmasında, farklı örneklem büyüklükleri (250, 500 ve 1000) üzerinde MH, LR ve SIBTEST yöntemlerini kullanarak DMF analizleri gerçekleştirmiş ve MH-LR yöntemlerinin birbirine yakın sonuçlar verdiğini tespit etmişlerdir. Yurdugül (2003) de 2001 yılı Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı′nı yanlılık bakımından ele aldığı çalışmasında, LR ve MH yöntemlerinin Türkçe, Fen Bilimleri ve Sosyal Bilgiler alt testleri için büyük ölçüde benzer sonuçlar verdiğini tespit etmiştir. Son olarak, Suna (2012) ve Demir (2013) çalışmalarında, MH, LR ve SIBTEST yöntemlerini kullanarak DMF analizleri gerçekleştirmiş ve bu çalışmayla benzer şekilde MH ve LR yöntemlerinin SIBTEST'e kıyasla daha uyumlu sonuçlar verdiğini tespit etmişlerdir.

Bu araştırmaya aksi yönde bir sonuca ise Gierl vd., (1999) tarafından gerçekleştirilen çalışmada ulaşılmıştır. Gierl vd., (1999) çalışmalarında Matematik ve Fen Bilgisi başarı testindeki maddelerin DMF gösterip göstermediğini MH ve LR yöntemleriyle analiz etmiş ve yöntemlerin farklı sayılarda DMF'li madde tespit ettiği sonucuna ulaşmışlardır. Ayan (2011) ise çalışmasında, PISA 2009 fen okuryazarlığı alt testinde yer alan maddelerinin DMF içerip içermediğini MH ve LR yöntemlerini kullanarak incelemiş ve yapılan analizlerde yöntemlerin farklı maddelerde DMF tespit ettiği, yöntemler arasında bir uyum bulunmadığını belirtmiştir. Benzer şekilde Gök vd., (2010) de çalışmasından elde ettiği bulgulardan yola çıkarak, MH ve LR yöntemleri arasında genel olarak düşük düzeyde bir uyum tespit etmiştir. Erdem (2015) ise çalışmasında MH, LR ve SIBTEST yöntemlerinin aynı maddelerde DMF belirleyebilme yönünden uyumlarını incelemiş ve MH-SIBTEST yöntemlerinin en uyumlu ikili olduğunu, en az uyumun ise MH-LR yöntemlerinde görüldüğünü tespit etmiştir.

Bu çalışmada, yöntemler arasında DMF'li olarak tespit ettiği madde sayısı tüm gruplarda en yüksek olan yöntem LR olurken en az sayıda maddeyi DMF'li olarak tespit eden yöntem ise SIBTEST olmuştur. Gomez Benito ve Navas Ara (2000) araştırmasında, alan

indeksleri Ki-kare temelli MH, lojit model ve lojistik regresyon yöntemleriyle DMF analizleri gerçekleştirmiş ve sonuçları karşılaştırmayı amaçlamıştır. Gerçekleştirdiği çalışmada, diğer yöntemlere kıyasla MH yönteminin daha fazla DMF'li madde tespit ettiği sonucuna ulaşmıştır. Benzer şekilde, Başusta (2013) da çalışmasında PISA 2006 fen okuryazarlığı alt testinde yer alan maddeleri farklı kültür ve dil değişkenine göre DMF içerip içermediğini MH, LR ve Alan İndeksleri yöntemleriyle incelemiş ve DMF gösteren madde sayısının en fazla MH yöntemiyle tespit edildiğini belirtmiştir. Bekci (2007) tarafından yapılan çalışmada ise Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'nın cinsiyet ve okul türü değişkenlerine göre DMF incelemesi yapılmıştır. Bu çalışmanın sonucunda da MH ve LR yöntemlerinin uyum göstermediği, alt test ve gruba göre farklı sonuçlar verdiği belirlenmiştir. Gierl vd., (2000) ise farklı koşullarda MH, LR ve SIBTEST yöntemlerini karşılaştırdığı çalışmasında SIBTEST'i DMF içeren maddeleri belirleyebilme gücü bakımından en güçlü yöntem olarak belirlemiştir.

Bu çalışmanın bulgularından yola çıkarak bir testte yer alan maddelerin DMF içerip içermediğini belirlemede bir yöntemin en doğru sonucu verdiğini söylemek doğru olmayacaktır. Bunun nedeni ise yöntemlerin maddeleri DMF'li olarak belirlerken izledikleri farklı matematiksel yöntemler, kullandıkları farklı algoritmalar ve kesme noktaları nedeniyle tam bir uyum içinde olmamalarıdır (Doğan & Öğretmen, 2008; Gök vd., 2010; Alatlı & Şenel, 2020; Bakan-Kalaycıoğlu & Berberoğlu, 2010). Bundan dolayı, DMF'yi belirlemek için güçlü bir istatistiksel yöntem üzerinde ortak bir fikre varılamamıştır (Gierl vd., 1999). Bu nedenle yanlılık incelemelerinde birden fazla DMF belirleme yönteminin bir arada kullanılması ve elde edilen sonuçların karşılaştırılarak karar verilmesi önerilmektedir (Hambleton, 2006). İlgili alanyazın tarandığında bu çalışmayla benzer ve çelişkili sonuçlar veren çalışmalara rastlanması da bu görüşü desteklemektedir.

2. Öneriler

Bu araştırma sonucunda yapılabilecek öneriler aşağıda sıralanmıştır.

Bu çalışmada PISA 2018 okuma becerileri alt testindeki maddelere sadece DMF analizi yapılmıştır. DMF analizleri yanlılık çalışmalarının ilk adımı olarak değerlendirilir. DMF analizlerinin yanında DMF gösteren maddeler uzmanlar tarafından incelemeye alınarak DMF gösterme nedenleri belirlenmeli ve yanlılık içerip içermedikleri tespit edilmelidir.

Geçmiş yıllarda yapılan PISA uygulamalarına yönelik DMF analizleri gerçekleştirilmelidir. DMF'ye neden olan noktalar belirlenerek belli türde soruların

yanlılık eğilimi taşıyıp taşımadığı üzerine çalışmalar yapılarak sonraki yıllarda yapılacak olan uygulamalarda bu hususlara dikkat edilerek madde hazırlanması önerilmektedir.

Bu çalışmada DMF analizleri yürütülürken MH, LR ve SIBTEST yöntemlerinden faydalanılmıştır. Bir testte yer alan maddelerin DMF içerip içermediğini ortaya çıkarmayı sağlayan birçok yöntem bulunmaktadır. Ancak bu yöntemler, maddeleri DMF'li olarak belirlerken izledikleri farklı matematiksel süreçler, kullandıkları farklı algoritmalar ve kesme noktaları nedeniyle tam bir uyum içinde değillerdir. Bu nedenle yanlılık incelemelerinde birden fazla DMF belirleme yönteminden faydalanılması ve ortaya çıkan sonuçların kıyaslanarak bir karara varılması tavsiye edilmektedir. Araştırmacılar bu çalışmada kullanılan yöntemlerin dışında MIMIC, Alan İndeksleri vb. yöntemlerle de analizleri yürütebilirler.

Bu çalışmada yalnızca PISA 2018 okuma becerileri alt testinde yer alan madde paketlerinden Temel bölüm için 1. madde paketi (Core RC1), 1.Aşama için 1. madde paketi (Stage 1 - R11H), 2.Aşama için 1. madde paketi (Stage 2 - R21H) seçilmiştir. Araştırmacılar tarafından farklı madde paketleri ele alınarak DMF analizleri yürütülebilir.

Bu çalışmanın örneklemi için yalnızca PISA uygulamasını kendi ana dillerinde alan Birleşik Krallık ve Fransa ile uygulamayı kendi ulusal dilinde bir çeviriye dönüştürerek kullanan Türkiye örneklemleri seçilmiştir. Araştırmacılar farklı kültür ve dil gruplarından örneklemler ekleyerek çalışma grubunu genişletebilirler.

Bu çalışmada, DMF analizleri farklı kültür ve dil değişkeni ile sınırlı tutulmuştur. Çalışmaya cinsiyet, okul türü vb. değişkenler eklenerek farklı açılardan DMF analizlerinin gerçekleştirilmesi sağlanabilir.

Bu çalışmada DMF analizleri yürütülürken yalnızca 1-0 şeklinde kodlanabilen iki kategorili maddeler analize dâhil edilmiştir. PISA uygulamasında farklı türde birçok madde bulunmaktadır. Araştırmacılara farklı madde türlerini kapsayan çalışmalar yapmaları önerilebilir.

KAYNAKLAR

Abbot, M. L. (2007). A confirmatory approach to differential item functioning on an ESL reading assessment. *Language Testing, 2*(1), 7-36.

Akın Arıkan, Ç., Uğurlu, S. & Atar, B. (2016). MIMIC, SIBTEST, Lojistik Regresyon ve Mantel-Haenszel yöntemleriyle gerçekleştirilen DMF ve yanlılık çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 31*(1), 34-52.

Alatlı, B. A., & Şenel, S. (2020). Değişen madde fonksiyonunun belirlenmesinde "difR" r paketinin kullanımı: ortaöğretime geçiş sınavı fen alt testi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 53*(3),865-901.

Angoff, W. (1993). Perspective on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 3–24). Hillsdale, NJ: Lawrence Erlbaum Associates.

Arslan, S. (2014). *Çoklu göstergeler çoklu nedenler ve lojistik regresyon yöntemlerinin değişen madde fonksiyonunu belirleme performansları.* Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Arslan, M. (2020). *TEOG sınavının yabancı dil alt testine ait maddelerin yanlılığının incelenmesi.* Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Asil, M. (2010). *Uluslararası öğrenci değerlendirme programı (PISA) 2006 öğrenci anketinin kültürler arası eşdeğerliğinin incelenmesi.* Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Asil, M., & Brown, G. T. (2016). Comparing OECD PISA reading in English to other languages: Identifying potential sources of non-invariance. *International Journal of Testing, 16*(1), 71-93.

Asil, M., & Gelbal, S. (2012). PISA öğrenci anketinin kültürler arası eşdeğerliği. *Eğitim ve Bilim, 37*(166), 236-249.

Atalay, K.(2010). *PISA 2006 öğrenci anketinde yer alan tutum maddelerinin değişen madde fonksiyonu açısından incelenmes.* Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Atalay, K., Gök, B., Kelecioğlu, H. & Arsan, N. (2012). Değişen madde fonksiyonunun belirlenmesinde kullanılan farklı yöntemlerin karşılaştırılması: bir simülasyon çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 2012*(43), 270-281.

Ayan, C. (2011). *PISA 2009 fen okuryazarlığı alt testinin değişen madde fonksiyonu açısından incelenmesi,* Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Balluerka, N., Gorostiaga, A., Gómez-Benito, J., & Hidalgo, M. D. (2010). Use of multilevel logistic regression to identify the causes of differential item functioning. *Psicothema, 22*(4), 1018-1025.

Bakan Kalaycıoğlu, D. & Berberoğlu, G. (2010). Differential item functioning analysis of the science and mathematics items in the university entrance examinations in Turkey. *Journal of Psychoeducational Assessment, 29*(5), 467-478.

Başusta, N.B. (2013), *PISA 2006 fen başarı testinin madde yanlılığının kültür ve dil açısından incelenmesi*. Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

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

Bauman, Z. (2020). *Küreselleşme: Toplumsal sonuçları*. İstanbul: Ayrıntı.

Bekçi, B. (2007). *Ortaöğretim kurumları öğrenci seçme sınavının değişen madde fonksiyonlarının cinsiyete ve okul türüne göre incelenmesi,* Yüksek Lisans Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.

Berberoglu, G., & Sireci, S. G. (1996). Evaluating translation fidelity using bilingual examinees. So What's the Big Dif ?

Bertrand, R., & Boiteau, N. (2003). Comparing the Stability of IRT- Based and non IRT Based DIF Methods in Different Cultural Contexts Using TIMSS Data. *Educational Resources Information Center (ERIC).*

Bolt, D. (2000). A SIBTEST approach to testing DIF hypothesis using experimentally designed test items. *Journal of Educational Measurement, 37*(4), 307-327.

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221-256). Westport: American Council on Education & Praeger Publishers.

Camilli, G., & Shepard, L.A. (1994). *Methods for identifying biased test items*. California: Sage.

Chan, D. (2000). Detection of differential item functioning on the kirton adaptation-innovation inventory using multiple-group mean and covariance structure analyses. *Multivariate Behavioral Research, 35*(2), 169-199.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice, 17*(1), 31-47.

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Austin: Holt, Rinehart and Winston.

Çet, S. (2006), *PISA 2003 matematik maddeleri kullanılarak yanlı çalışan maddelerin tespitinde çok boyutlu eşleştirme analizi.* Doktora Tezi, Orta Doğu Teknik Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Demir, S. (2013). *PISA 2009 matematik okuryazarlığı alt testinde bulunan maddelerin Mantel-Haenszel, SIBTEST ve lojistik regresyon yöntemleri ile değişen madde fonksiyonunun incelenmesi*. Yüksek Lisans Tezi, Abant İzzet Baysal Üniversitesi Eğitim Bilimleri Estitüsü, Bolu.

Demir, S., & Köse, İ. A. (2014). Mantel-Haenszel, SIBTEST ve Lojistik Regresyon Yöntemleri ile değişen madde fonksiyonunu analizi. *International Journal of Human Sciences, 11*(1), 700-714.

Doğan, N. & Öğretmen, T. (2008). Değişen madde fonksiyonunu belirlemede Mantel - Haenszel, ki-kare ve Lojistik Regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim, 33*(148), 100-112.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of educational measurement, 23*(4), 355-368.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology, 70*(4), 662-680.

Dzul-Garcia, C., & Atar, B. (2020). Investigation of possible item bias on PISA 2015 science items across Chile, Costa Rica and Mexico. *Culture and Education, 32*(3), 470-505.

Ellis, R. (1989). *Understanding second language acquisition* (Vol. 31). Oxford: Oxford University Press.

Eells, K., Allison, D., Havighurst, R. J., Herrick, V. E., & Tyler, R. (1951). *Intelligence and cultural differences: A study of cultural learning and problem-solving*. Chicago: University of Chicago.

Ercikan, K. (1998). Translation effects in international assessments. *International Journal of Educational Research, 29*(1998), 543-553

Ercikan, K., Gierl, M. J., McCreith, T., Puhan, G., & Koh, K. (2004). Comparability of bilingual versions of assessments: sources of incomparability of English and French versions of Canada's national achievement tests. *Applied Measurement in Education, 17*(3), 301-321.

Ercikan, K., & Koh, K. (2005). Examining the construct comparability of the English and French versions of TIMSS. *International Journal of Testing, 5*(1), 23-35.

Ercikan, K., & McCreith, T. (2002). Effects of adaptations on comparability of test items and test scores. *In Secondary analysis of the TIMSS data* (pp. 391-405). Springer, Dordrecht.

Erdem, B. (2015). *Ortaöğretime geçişte kullanılan ortak sınavların değişen madde fonksiyonu açısından kitapçık türlerine göre farklı yöntemlerle incelenmesi.* Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Finch, W. H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel- Haenszel, SIBTEST and the IRT likelihood ratio. *Applied Psychological Measurement, 29*(4), 278-295.

Finch, W. H., & French, B. F. (2011). Estimation of MIMIC model parameters with multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(2), 229-252.

Finch, H., & French, B. (2013). A Monte Carlo comparison of robust MANOVA test statistics. *Journal of Modern Applied Statistical Methods, 12*(2), 35-81.

French, B. F., & Finch, W. H. (2015). Transforming SIBTEST to account for multilevel data structures. *Journal of Educational Measurement, 52*(2), 159-180.

Frey, A. & Bernhardt R. (2012). On the importance of using balanced booklet designs in PISA. *Psychological Test and Assessment Modeling, 54*(4), 397-417.

Gierl, M. J. (2000). Construct Equivalence on Translated Achievement Tests. *Canadian*

Journal of Education, 25(4), 280-296.

Gierl, M. J., Gotzmann, A., & Boughton, K. A. (2004). Performance of SIBTEST when the percentage of DIF items is large. *Applied Measurement in Education, 17*(3), *241-264.*

Gierl, M. H., Khaliq, S. N. & Boughton, K. (1999, June). *Gender differential item functioning in mathematics and science: prevalence and policy implications.* Paper presented at the Annual Meeting of the Canadian Society for the Study of Education, Canada.

Gierl, M. J., Jodoin, M. G., & Ackerman, T. A. (2000). *Performance of Mantel-Haenszel Simultaneous Item Bias Test, and Logistic Regression when the proportion of DIF items is large*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Gierl, M., Rogers, W. T., & Klinger, D. A. (1999). Using statistical and judgmental reviews to identify and interpret translation differential item functioning. *Alberta Journal of Educational Research*, *45*(4), 353-376.

Gomez-Benito, J., & Navas-Ara, M. J. (2000). A comparison of χ2, RFA and IRT based procedures in the detection of DIF. *Quality and Quantity*, *34*(1), 17-31.

Gök, B., Atalay Kabasakal, K., & Kelecioğlu, H. (2014). PISA 2009 öğrenci anketi tutum maddelerinin kültüre göre değişen madde fonksiyonu açısından incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *5*(1), 72-87.

Gök, B., Kelecioğlu, H., & Doğan, N. (2010). Değişen madde fonksiyonunu belirlemede Mantel-Haenzsel ve lojistik regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim, 35*(156), 3-16.

Grisay, A., De Jong, J. H. A. L., Gebhardt, E., Berezner, A., & Halleux-Monseur, B. (2007). Translation equivalence across PISA countries. *Journal of Applied Measurement*, *8*(3), 249-266.

Grisay, A. , & Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation, 33*(1), 69-86.

Guilford, J. P. (1956). *Fundamental statistic in psychology and education* (3.th ed.). USA:McGraw-Hill

Gür, E. (2019). *PISA 2015 Uygulamasındaki maddelerin kültüre göre değişen madde fonksiyonu açısından incelenmesi.* Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Hambleton, R. K. (1994). The rise and fall of criterion-referenced measurement?. *Educational Measurement: Issues and Practice*.

Hambleton, R. K. (1996, April). *Guidelines for adapting educational and psychological tests*. Paper presented at the Annual Meeting of the National Council on Measurement in Education , New York.

Hambleton, R.K. (2002). Adapting achievement tests into multiple languages for ınternational assessments. In A. Porter y A. Gamoran (Ed.), *Methodological Advances in Large-scale Cross-national Education Surveys* (pp. 58-79) Washington: National Academy of Sciences.

Hambleton, R. K. (2006). Good practices for identifying differential item functioning. *Medical Care*, *44*(3), 182-188.

Hambleton, R.K., & de Jong, J. H. A. L. (2003). Advances in translating and adapting educational and psychological tests. *Language Testing*, *20*(1), 127-134.

Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (Eds.) (2005). *Adapting educational and psychological tests for cross-cultural assessment.* Mahwah, NJ: Lawrence Erlbaum.

Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research, 45*(1), 153-171.

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology, 1*(1), 1-30.

Hambleton, R., & Rogers, J. (1994). Item bias review. *Practical Assessment, Research, and Evaluation*, *4*(1), 6.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. CA: Sage.

Hidalgo, M. D., & Lopez-Pina, J. A. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement, 64*(6), 903-915.

Holland, P. W.& Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer ve H.I. Braun (Eds.), *Test validity* (pp.129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. London: Lawrence Erlbaum Associates.

Huang, X., Wilson, M., & Wang, L. (2016). Exploring plausible causes of differential item functioning in the PISA science assessment: language, curriculum or culture. *Educational Psychology, 36*(2), 378-390.

Hui, C. H., & Triandis, H. C. (1985). Measurement in Cross-Cultural Psychology: A Review and Comparison of strategies. *Journal of Cross-Cultural Psychology, 16*(2), 131-152.

Jodoin, M. G. & Gierl, M. J. (2001).Evaluating type _ error and power using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*(4), 329-349.

Kabasakal, K. A., & Kelecioğlu, H. (2012). Evaluation of attitude items in PISA 2006 student questionnaire in terms of differential item functioning. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, *45*(2), 77-96.

Köse, İ. A. (2015). PISA 2009 öğrenci anketi alt ölçeklerinde (q32-q33) bulunan maddelerin değişen madde fonksiyonu açısından incelenmesi. *Kastamonu Eğitim Dergisi*, *23*(1), 227-240.

Le, L. T. (2006, April). *Analysis of differential item functioning*. Paper presented at the meeting of the American Educational Research Association, San Francisco CA.

Le, L. T. (2009). Investigating gender differential item functioning across countries and test languages for PISA science items. *International Journal of Testing, 9*(2), 122–133.

Li, H., & Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika, 61*(4), 647-677.

Li, Z., & Zumbo, B. D. (2009). Impact of differential item functioning on subsequent statistical conclusions based on observed test score data. *Psicológica*, *30*(2), 343-370.

Lord, F. M. (1980). *Aplications of item response theory to practical testing problems.* New Jersey: Lawrence Erlbaum.

Magis, D., Beland, S., Teurlinckx, F., & Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*(3), 847-862.

Mellenberg, G. J. (1983). Conditional item bias methods. In S. H. Irvine and W. J. Barry (Eds), *Human assesment and cultural factors* (pp. 293–302). New York: Plenum Pres.

Meredith, W. & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement invariance. *Psychometrica, 57*(2*)*, 289-311.

Milli Eğitim Bakanlığı (2011). *MEB 21. yüzyıl öğrenci profili*. https://www.meb.gov.tr/earged/earged/21.%20yy_og_pro.pdf sayfasından erişilmiştir.

Milli Eğitim Bakanlığı (2019). *PISA 2018 Türkiye ön raporu*. http://www.meb.gov.tr/meb_iys_dosyalar/2019_12/03105347_PISA_2018_Turkiye_On_Raporu.pdf sayfasından erişilmiştir.

Millsap, R. E. (2005). Four unresolved problems in studies of factorial invariance. In A. MaydeuOlivares & J. J. McArdle (Eds.), *Contemporary psychometrics* (pp. 153–172). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

Narayanan, P., & Swaminathan, H. (1994). Performance of the mantel-haenszel and simultaneous item bias procedures for detecting differential item functioning. *Applied Psychological Measurement, 18*(4), 315-328.

Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement, 20*(3), 257-274.

OECD (2007). *PISA 2006: Science Competencies for Tomorrow's World: Volume 1and Volume 2*. Paris: OECD.

OECD (2019). *PISA 2018 results volume I: What students know and can do*. Paris: OECD.

Osterlind, S. J. (1983). *Test item bias* (No. 30). London: Sage.

Osterlind, S. & Everson H. (2009). *Differential item functioning*. London: Sage.

Özmen, D.T. (2013). *PISA 2009 okuma becerileri testi maddelerinin yanlılık açısından Türkiye, Amerika Birleşik Devletleri ve Birleşik Krallık uygulamalarında*

karşılaştırılması. Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Psychometrics* (pp. 125-167). Amsterdam: Elsevier.

Potenza, M. T., & Dorans, N. J. (1995). DIF assessment for polytomously scored items: a framework for classification and evaluation. *Applied Psychological Measurement*, *19*(1), 23–37.

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, 53(4), 495-502

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, *14*(2), 197-207.

Raju, N. S., & Ellis, B. B. (2002). Differential item and test functioning. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 156–188). Jossey-Bass..

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychological Bulletin, 114*(3), 552-566

Rogers, H. J., & Swaminathan, H. (1993). A comparison of the logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement, 17*(2), 105-116.

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.

RStudio (2020). *Take control of your R code.* Retrieved from https://rstudio.com/products/rstudio/

Samuelsen, K. (2005). *Examining differential item functioning from a latent class perspective*. Doctoral Dissertation, University of Maryland, College Park.

Schulz, W. (2003, April). *Validating questionnaire constructs in international studies: Two examples from PISA 2000*. Paper presented at the Annual Meetings of the American Educational Research Association (AERA), Chicago, IL.

Schulz, W. (2005, April). *Testing parameter invariance for questionnaire indices using confirmatory factor analysis and item response theory*. Paper presented at the Annual Meetings of the American Educational Research Association (AERA), San Francisco, CA.

Shealy, R., & Stout, W. F. (1993). A model-based standardization approach that separates true bias/DIF from group differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika, 58*(2), 159-194.

Smith, L. L., ve Reise, S. P. (1998). Gender differences on negative affectivity: an IRT study of differential item functioning on the multidimensional personality questionnaire stress reaction scale. *Journal of Personality and Social Psychology, 75*(5), 1350-1362.

Sırgancı, G. (2012). *PISA 2006 öğrenci anketi madde yanlılığının sıralı lojistik regresyon ve poly-SIBTEST teknikleri ile test edilmesi*. Yüksek Lisans Tezi, Abant İzzet Baysal Üniversitesi Eğitim Bilimleri Enstitüsü, Bolu.

Sireci, S. G., & Berberoğlu, G. (2000). Using bilingual respondents to evaluate translated – adapted items. *Applied Measurement in Education, 13*(3), 229-248.

Sireci, S. G., Fitzgerald, C., & Xing, D. (1998). *Adapting credentialing examinations for international uses.* Paper presented at the meeting of the American Educational Reaserch Association, San Diego, CA.

Somer, O., Korkmaz, M., Dural, S., & Can, S. (2009). Ölçme eşdeğerliğinin yapısal eşitlik modellemesi ve madde cevap kuramı kapsamında incelenmesi. *Türk Psikoloji Dergisi, 24*(64), 61-75.

Suna, H.E. (2012). *TIMSS 2007 fen bilimleri testindeki maddelerin dil ve cinsiyet yanlılığı açısından incelenmesi*. Yüksek Lisans Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*(4), 361-370.

Swanson, D. B., Clauser, B. E., Case, S. M., Nungester, R. J., & Featherman, C. (2002). Analysis of differential item functioning (DIF) using hierarchical logistic regression models. *Journal of Educational and Behavioral Statistics, 27*(1), 53-75.

Tabachnick, B. G., & Fidell, L. S. (2015). *Çok değişkenli istatistiklerin kullanımı* (M. Baloğlu, Çev.). Ankara: Nobel.

Tan, Ş. (2016). *SPSS ve Excel uygulamalı temel istatistik-1*. Ankara:Pegem Akademi

Thissen, D., & Steinberg, L. (1988). Data analysis using item response theory. *Psychological Bulletin, 104*(3), 385.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67–113). Lawrence Erlbaum Associates, Inc.

Ulutaş, S. (2012). *PISA 2006 fen okuryazarlığı testindeki maddelerin yanlılık bakımından araştırılması.* Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Using Rtools4.0 on Windows (2020). Retrieved from https://cran.rstudio.com/bin/windows/Rtools/

Uzun, N. B., & Gelbal, S. (2017). PISA fen başarı testinin madde yanlılığının kültür ve dil açısından incelenmesi. *Kastamonu Eğitim Dergisi, 25*(6), 2427-2446.

Van de Vijver, F., & Tanzer, N. K. (1998). Bias and equivalence in cross-cultural assessment. *European review of applied psychology, 47*(4), 263-279.

Van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 54*(2), 119-135.

Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods, 5*(2), 139-158.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the MI literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-69.

Wiberg, M. (2007). *Measuring and detecting differential item functioning in criterionreferenced licensing test: A theoretic comparison of methods.* EM No. 60. Umea University Department of Educational Measurement, Umea.

Wu, A. D., & Ercikan, K. (2006). Using multiple-variable matching to identify cultural sources of differential item functioning. *International Journal of Testing, 6*(3), 287-300.

Yayan, B., & Berberoğlu, G. (2004). A re-analysis of the TIMSS 1999 mathematics assessment data of the Turkish students. *Studies in Educational Evaluation, 30*(1), 87-104.

Yıldırım, H. H., & Berberoğlu, G. (2009). Judgmental and statistical DIF analyses of the PISA-2003 Mathematics Literacy Items. *International Journal of Testing, 9*(2), 108-121.

Yurdugül, H. (2003). *Ortaöğretim kurumları seçme ve yerleştirme sınavının madde yanlılığı açısından incelenmesi.* Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Zhou, J., & Gierl, M. J. (2006, April). *Evaluating the performance of MULTISIB using different multidimensional matching criteria.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Zheng, Y., Gierl, M. J., & Cui, Y. (2007). *Using real data to compare DIF detection and effect size measures among Mantel-Haenszel, SIBTEST, and Logistic Regression procedures.* Thesis in Educational Psychology. Retrieved from http://etda.libraries.psu.edu/theses/approved/WorldWideFiles/ETD-312 /hglmdif.pdf

Zieky, M. (1993) Practical questions in the use of DIF statistics in test development. In P.W. Holland ve H. Wainer (Eds.), *Differential item functioning* (pp. 337-347). Hillsdale NJ: Erlbaum.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF).* Ottawa: National Defense Headquarters.

Zumbo, B. D., (2003). Does item-level DIF manifest itself in scale-level analyses? implications for translating language tests. *Language Testing, 20* (2), 136- 147.

Zumbo, B. D. (2005). Manifestation of differences in item-level characteristics in scale-level measurement invariance tests of multi-group confirmatory factor analyses. *Journal of Modern Applied Statistical Methods, 4* (1), 275-282.

Zumbo, B. D. (2007). Validity: Foundational Issues and Statistical Methodology. In C. R. Rao and S. Sinharay (Eds.), *Handbook of Statistics, Vol.26: Psychometrics* (pp. 45-79). Elsevier Science B. V.: The Netherlands.

Zumbo, B. D., & Gelin, M. N. (2005). A matter of test bias in educational policy research: Bringing the context into picture by investigating sociological/community moderated (or mediated) test and item bias. *Journal of Educational Research and Policy* Studies, *5*(1), 1-23.

Zumbo, B. D., & Thomas, D. R. (1996, October). *A measure of DIF effect size using logistic regression procedures.* Paper presented at the National Board of Medical Examiners, Philadelphia.

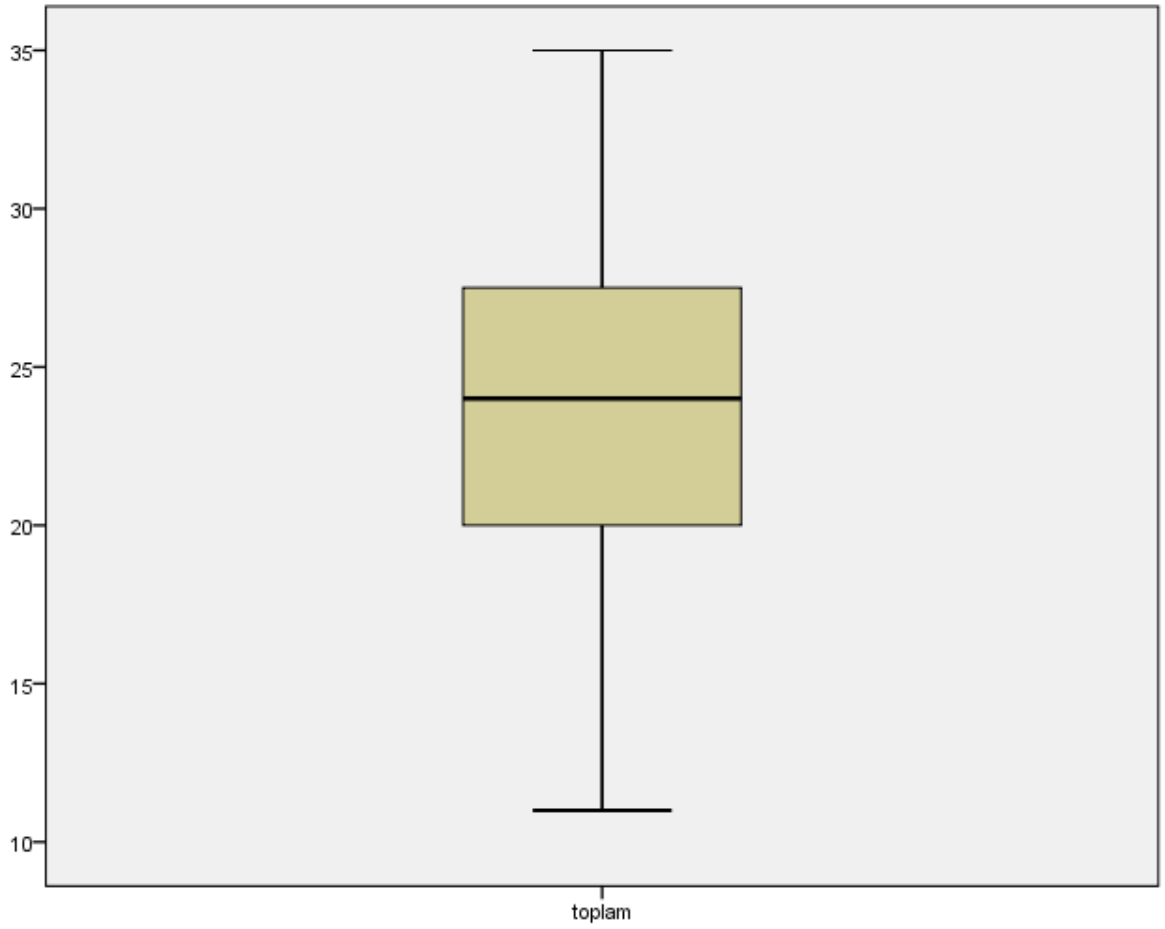
Zumbo, B. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. Lissitz (Ed.), *The concept of validity* (pp. 65–82). Charlotte, NC: Information Age.

EKLER

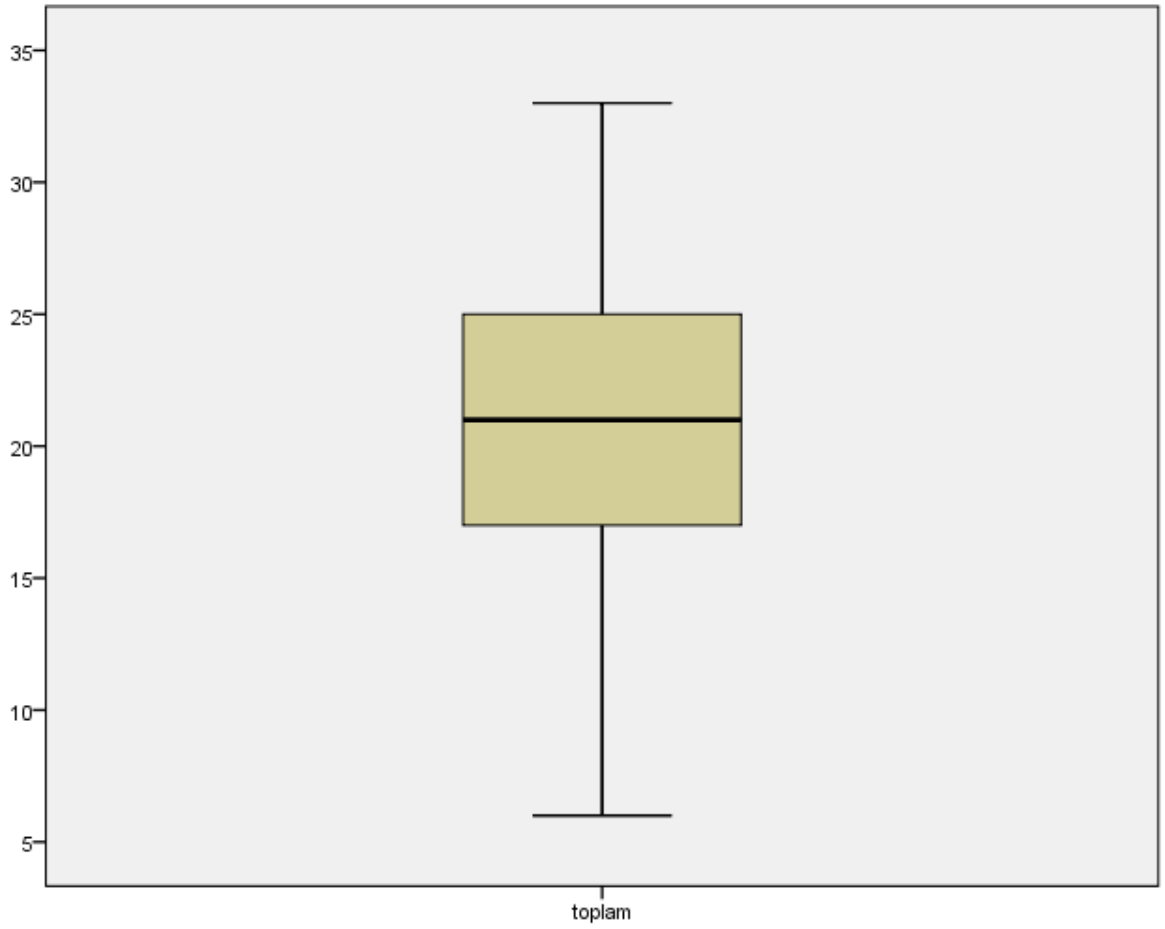
EK 1. Kayıp Veri Oranları

Madde No	Fransa			Türkiye			Birleşik Krallık		
	Geçerli Veri	Kayıp Veri		Geçerli Veri	Kayıp Veri		Geçerli Veri	Kayıp Veri	
		Sayısı	Yüzdesi		Sayısı	Yüzdesi		Sayısı	Yüzdesi
1	143	0	0	125	0	0,0	325	1	0,3
2	134	9	6,3	114	11	8,8	323	3	0,9
3	143	0	0	124	1	0,8	325	1	0,3
4	143	0	0	125	0	0,0	326	0	0,0
5	143	0	0	118	7	5,6	314	12	3,7
6	143	0	0	123	2	1,6	322	4	1,2
7	143	0	0	125	0	0,0	325	1	0,3
8	143	0	0	124	1	0,8	326	0	0,0
9	143	0	0	125	0	0,0	325	1	0,3
10	143	0	0	125	0	0,0	324	2	0,6
11	142	1	0,7	125	0	0,0	323	3	0,9
12	143	0	0	125	0	0,0	324	2	0,6
13	133	10	7	118	7	5,6	302	24	7,4
14	133	10	7	118	7	5,6	299	27	8,3
15	141	2	1,4	121	4	3,2	323	3	0,9
16	138	5	3,5	117	8	6,4	312	14	4,3
17	141	2	1,4	121	4	3,2	323	3	0,9
18	141	2	1,4	120	5	4,0	321	5	1,5
19	140	3	2,1	120	5	4,0	317	9	2,8
20	140	3	2,1	121	4	3,2	316	10	3,1
21	105	38	26,6	98	27	21,6	276	50	15,3
22	139	4	2,8	121	4	3,2	313	13	4,0
23	138	5	3,5	124	1	0,8	322	4	1,2
24	139	4	2,8	121	4	3,2	321	5	1,5
25	139	4	2,8	118	7	5,6	322	4	1,2
26	137	6	4,2	115	10	8,0	319	7	2,1
27	135	8	5,6	114	11	8,8	320	6	1,8
28	116	27	18,9	101	24	19,2	291	35	10,7
29	127	16	11,2	112	13	10,4	314	12	3,7
30	122	21	14,7	108	17	13,6	310	16	4,9
31	104	39	27,3	99	26	20,8	290	36	11,0
32	91	52	36,4	93	32	25,6	283	43	13,2
33	65	78	54,5	81	44	35,2	242	84	25,8
34	94	49	34,3	87	38	30,4	287	39	12,0
35	108	35	24,5	101	24	19,2	294	32	9,8

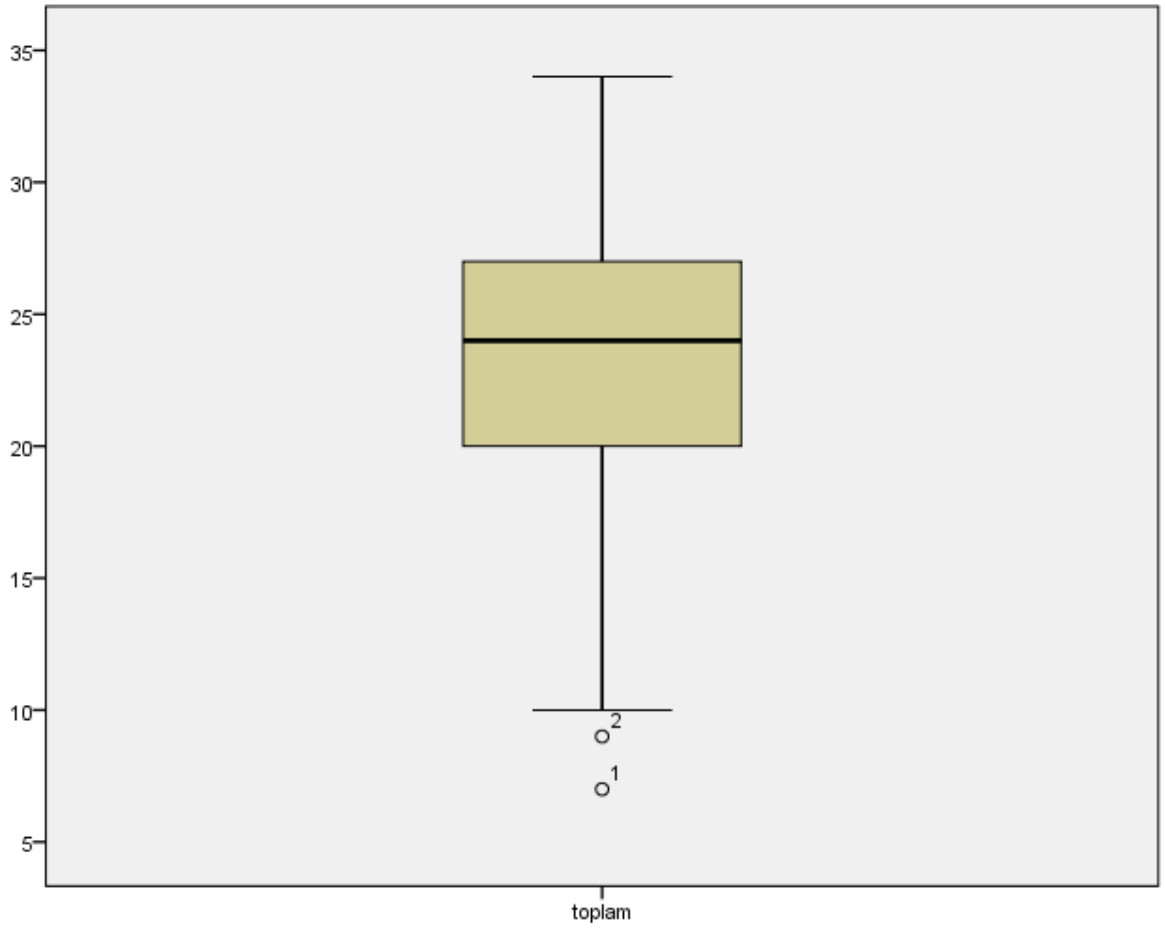
EK 2. Fransa Örneklemi İçin Oluşturulan Uç Değer Analizi Kutu Grafiği



EK 3. Türkiye Örneklemi İçin Oluşturulan Uç Değer Analizi Kutu Grafiği



EK 4. Birleşik Krallık Örneklemi İçin Oluşturulan Uç Değer Analizi Kutu Grafiği



EK 5. Fransa ve Birleşik Krallık Örneklemi İçin DMF Belirleme Yöntemlerinin Spearman Brown Sıra Farkları Korelasyon Katsayısı

Correlations

			M-H	Logistic
Spearman's rho	M-H	Correlation Coefficient	1,000	,765**
		Sig. (2-tailed)	.	,000
		N	35	35
	Logistic	Correlation Coefficient	,765**	1,000
		Sig. (2-tailed)	,000	.
		N	35	35

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

			M-H	SIBTEST
Spearman's rho	M-H	Correlation Coefficient	1,000	,614**
		Sig. (2-tailed)	.	,000
		N	35	35
	SIBTEST	Correlation Coefficient	,614**	1,000
		Sig. (2-tailed)	,000	.
		N	35	35

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

			Logistic	SIBTEST
Spearman's rho	Logistic	Correlation Coefficient	1,000	,708**
		Sig. (2-tailed)	.	,000
		N	35	35
	SIBTEST	Correlation Coefficient	,708**	1,000
		Sig. (2-tailed)	,000	.
		N	35	35

**. Correlation is significant at the 0.01 level (2-tailed).

EK 6. Türkiye ve Fransa Örneklemi İçin DMF Belirleme Yöntemlerinin Spearman Brown Sıra Farkları Korelasyon Katsayısı

Correlations

			M-H	Logistic
Spearman's rho	M-H	Correlation Coefficient	1,000	,655**
		Sig. (2-tailed)	.	,000
		N	35	35
	Logistic	Correlation Coefficient	,655**	1,000
		Sig. (2-tailed)	,000	.
		N	35	35

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

			M-H	SIBTEST
Spearman's rho	M-H	Correlation Coefficient	1,000	,464**
		Sig. (2-tailed)	.	,005
		N	35	35
	SIBTEST	Correlation Coefficient	,464**	1,000
		Sig. (2-tailed)	,005	.
		N	35	35

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

			Logistic	SIBTEST
Spearman's rho	Logistic	Correlation Coefficient	1,000	,361*
		Sig. (2-tailed)	.	,033
		N	35	35
	SIBTEST	Correlation Coefficient	,361*	1,000
		Sig. (2-tailed)	,033	.
		N	35	35

*. Correlation is significant at the 0.05 level (2-tailed).

EK 7. Türkiye ve Birleşik Krallık Örneklemi İçin DMF Belirleme Yöntemlerinin Spearman Brown Sıra Farkları Korelasyon Katsayısı

Correlations

			M-H	Logistic
Spearman's rho	M-H	Correlation Coefficient	1,000	,793**
		Sig. (2-tailed)	.	,000
		N	35	35
	Logistic	Correlation Coefficient	,793**	1,000
		Sig. (2-tailed)	,000	.
		N	35	35

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

			M-H	SIBTEST
Spearman's rho	M-H	Correlation Coefficient	1,000	,704**
		Sig. (2-tailed)	.	,000
		N	35	35
	SIBTEST	Correlation Coefficient	,704**	1,000
		Sig. (2-tailed)	,000	.
		N	35	35

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

			Logistic	SIBTEST
Spearman's rho	Logistic	Correlation Coefficient	1,000	,535**
		Sig. (2-tailed)	.	,001
		N	35	35
	SIBTEST	Correlation Coefficient	,535**	1,000
		Sig. (2-tailed)	,001	.
		N	35	35

**. Correlation is significant at the 0.01 level (2-tailed).



GAZİLİ OLMAK AYRICALIKTIR...