



**DEĞİŞEN MADDE FONKSİYONU BELİRLEME YÖNTEMLERİNİN
MADDE VE YETENEK PARAMETRE KESTİRİMİNE ETKİSİ**

Fatma Gökçen Ayva Yörü

**DOKTORA TEZİ
EĞİTİM BİLİMLERİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

ARALIK, 2019

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

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

YAZARIN

Adı : Fatma Gökçen
Soyadı : AYVA YÖRÜ
Bölümü : : Eğitimde Ölçme ve Değerlendirme
İmza :
Teslim tarihi : ... / /

TEZİN

Türkçe Adı : Değişen Madde Fonksiyonu Belirleme Yöntemlerinin Madde ve Yetenek Parametre Kestirimine Etkisi

İngilizce Adı : The Effect of Diferantial Item Functioning Determination Methods on Item and Ability Parameters Estimates

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: Fatma Gökçen AYVA YÖRÜ

İmza:

JÜRİ ONAY SAYFASI

Fatma Gökçen Ayva Yörü tarafından hazırlanan “Değişen Madde Fonksiyonu Belirleme Yöntemlerinin Madde ve Yetenek Parametre Kestirimine Etkisi” adlı tez çalışması aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Gazi Üniversitesi Eğitim Bilimleri Anabilim Dalı’nda Doktora tezi olarak kabul edilmiştir.

Danışman: Doç. Dr. Hakan Yavuz ATAR

(Eğitimde Ölçme ve Değerlendirme Bilim Dalı, Gazi Üniversitesi)

Başkan: Prof. Dr. Nuri DOĞAN

(Eğitimde Ölçme ve Değerlendirme Bilim Dalı, Hacettepe Üniversitesi)

Üye: Prof. Dr. Şeref TAN

(Eğitimde Ölçme ve Değerlendirme Bilim Dalı, Gazi Üniversitesi)

Üye: Prof. Dr. Bayram ÇETİN

(Eğitimde Ölçme ve Değerlendirme Bilim Dalı, Gazi Üniversitesi)

Üye: Prof. Dr. Şener BÜYÜKÖZTÜRK

(Eğitimde Ölçme ve Değerlendirme Bilim Dalı, Hasan Kalyoncu Üniversitesi)

Tez Savunma Tarihi: 19/12/2019

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

Prof. Dr. Selma YEL

Eğitim Bilimleri Enstitüsü Müdürü

Yeryüzündeki mağdur çocuklara...

TEŞEKKÜR

En büyük teşekkürüm saygıdeğer danışman hocam Doç. Dr. Hakan Yavuz ATAR'a... Akademik gelişimime yapmış olduğu katkıları için sonsuz teşekkürler. Hayatımın en zor sürecinde verdiği tavsiyelerle yolumu aydınlatan, dik yokuşları çıkarken kendime olan inancımı kaybetmememi sağlayan, akademik ve kişisel hayatımda örnek aldığım nadir insanlardan birisi olan kıymetli hocama minnettarım.

Tez çalışmama yapmış oldukları değerli katkılarından dolayı kıymetli hocalarım Prof. Dr. Şener BÜYÜKÖZTÜRK, Prof. Dr. Şeref TAN, Prof. Dr. Nuri DOĞAN, Prof. Dr. Bayram ÇETİN ve Doç. Dr. Burcu ATAR'a,

Akademik ve manevi desteklerinden ötürü saygıdeğer hocalarım Prof. Dr. Mustafa SOLAK, Prof. Dr. Mehmet ŞİŞMAN, Prof. Dr. Engin KARADAĞ ve Prof. Dr. Mehtap ÇAKAN'a

Sağlığıma kavuşmamı sağlayan kıymetli doktorum Prof. Dr. Sezai LEVENTOĞLU'na ve manevi desteğini esirgemeyen değerli doktorum Prof. Dr. Bülent MENTEŞ'e,

Güler yüzüyle hep yanımda olan canım ablam Nükhet CANISAĞ DOĞDU'ya,

Sevgili arkadaşlarım Serap BÜYÜKKIDIK, Elif SEZER ve Fatıma Nur FİŞNE'ye,

Kan bağıımız olmasa bile benim için kız kardeşten öte olan manevi kardeşim Sibel'e,

Evlatları olmaktan gurur duyduğum babama ve anneme,

Hayatıma girdiği andan itibaren dünyamı güzelleştiren ve bana en güzel mutlulukları yaşatan sevgili eşim İsmail'e,

Ahde vefa gösteren tüm insanlara,

Teşekkürler...

DEĞİŞEN MADDE FONKSİYONU BELİRLEME YÖNTEMLERİNİN MADDE VE YETENEK PARAMETRE KESTİRİMİNE ETKİSİ

(Doktora Tezi)

Fatma Gökçen Ayva Yörü

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Aralık, 2019

ÖZ

Bu araştırma iki kategoride puanlanan test maddelerinde DMF belirleme yöntemlerinin performansları ile yöntemlerin madde ve yetenek parametre kestirimlerine olan etkisini incelemeyi amaçlamaktadır. Araştırmada farklı koşulları içeren simülasyon veri seti ele alınmıştır. Bu araştırma var olan durumun ortaya konmaya çalışıldığı betimsel bir çalışmadır. Simülasyon çalışması kapsamında örneklem büyüklüğü (600, 1200 ve 2400), odak ve referans grubun örneklem büyüklüğü oranı (1/1 ve 1/2), testte bulunan DMF’li madde oranı (%5, %10 ve %20) ve DMF büyüklüğü (0,5 ve 1,0) olmak üzere toplam 36 koşul manipüle edilmiştir. Araştırmada her bir koşul için 100 replikasyon yapılmıştır. Böylece 36 koşulun her biri için yapılan 100 replikasyon neticesinde analizler toplam 3600 veri seti üzerinden gerçekleştirilmiştir. WinGen yazılımı ile test uzunluğu 40, yanıt kategorisi iki ve 2PL modele uygun madde parametreleri üretilmiştir. Veri üretimi ise R programında yer alan “eRm” paketi aracılığıyla gerçekleştirilmiştir. DMF analizleri Mantel-Haenszel, Lord’un χ^2 , Lojistik Regresyon, SIBTEST ve Raju’nun alan ölçümü yöntemleri kullanılarak gerçekleştirilmiştir. Yöntemlerin analizleri R programında yer alan “difR” ve “mirt” paketleri kullanılarak yapılmıştır. DMF belirleme yöntemlerinin çeşitli koşullar altında performanslarının karşılaştırılmasında öncelikle I. Tip hata ve güç oranları hesaplanmıştır. Her bir DMF belirleme yönteminin sonuçlarını karşılaştırmak için kestirilen

madde ve yetenek parametre değerlerinin RMSE, Standart Hata (SH) ve BIAS (yanlılık) değerleri hesaplanmıştır. Yapılan hesaplamaların ardından madde ve yetenek parametre değerlerinin yöntemlere göre farklılaşma durumunu ve yöntemlerle simülasyon koşullarının birbiriyle olan etkileşimini test etmek amacıyla Faktöriyel ANOVA analizleri yapılmıştır. Simülasyon çalışması sonucunda örneklem büyüklüğü, DMF büyüklüğü ve testte bulunan DMF'li madde oranı arttıkça yöntemlere göre I. Tip hata ve güç oranlarının da yükseldiği görülmüştür. Odak ve referans grup örneklem büyüklükleri birbirine eşitken I. Tip hata ve güç oranlarının nispeten yüksek olduğu tespit edilmiştir. Ayırt edicilik parametresine ait RMSE değerlerinin örneklem büyüklüğü arttıkça azaldığı; örneklem büyüklüğü oranlarına, DMF büyüklüklerine ve DMF'li madde oranlarına göre değişmediği görülmüştür. Ayrıca ayırt edicilik parametresine ait RMSE değerlerinin DMF belirleme yöntemlerine göre farklılaşmadığı sonucuna ulaşılmıştır. Güçlük parametresine ait RMSE değerlerinin örneklem büyüklüğü arttıkça azaldığı ve yöntemlere göre farklılaştığı görülmüştür. Fakat güçlük parametresine ait RMSE değerlerinin örneklem büyüklüğü oranlarına, DMF büyüklüklerine ve DMF'li madde oranlarına göre ise değişmediği görülmüştür. Yetenek parametresine ait RMSE değerlerinin DMF yöntemlerine göre farklılaşmadığı fakat örneklem büyüklüklerine, DMF büyüklüklerine ve testte bulunan DMF'li madde oranlarına göre farklılaştığı tespit edilmiştir. Örneklem büyüklüğü, DMF büyüklüğü ve DMF'li madde oranları arttıkça yetenek parametresine ilişkin RMSE değerlerinin arttığı tespit edilmiştir. Odak ve referans grup örneklem büyüklüğü eşit (1/1) iken elde edilen yetenek parametresine ilişkin RMSE değerlerinin ise örneklem büyüklüğü oranı 1/2'ye göre nispeten yüksek olduğu görülmüştür.

Anahtar kelimeler: Değişen Madde Fonksiyonu, Madde ve Yetenek Parametre Kestirimleri, I. Tip Hata Oranı, Güç Oranı

Sayfa Adedi: xx + 136

Danışman: Doç. Dr. Hakan Yavuz ATAR

**THE EFFECT OF DIFFERANTIAL ITEM FUNCTIONING
DETERMINATION METHODS ON
ITEM AND ABILITY PARAMETERS ESTIMATES
(PhD Thesis)**

Fatma Gökçen Ayva Yörü

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

December, 2019

ABSTRACT

This research aims to investigate the performance of DIF determination methods and the effect of methods on item and ability parameter estimate in test items scored in two categories. In this study, simulation data set including different conditions were used. This research is a descriptive study since the current situation is tried to be revealed. In the simulation study, 36 conditions in total were manipulated: sample size (600, 1200 and 2400), sample size ratio of the focus and reference group (1/1 and 1/2), ratio of the items with DIF in the test (5%, 10% and 20%) and DIF magnitude (0.5 and 1.0). In this study, 100 replications were performed for each condition. Thus, as a result of 100 replications performed for each of the 36 conditions, analyzes were conducted on 3600 data sets in total. Through WinGen software, item parameters suitable for test length with 40 items, two response categories and 2PL model were produced. The data was generated through the “eRm” package included in the R program. DIF analyzes were performed through the use of Mantel-Haenszel, Logistics Regression, SIBTEST, Lord’s χ^2 and Raju’s area measurements methods. DIF methods were analyzed by using “difR” and “mirt” packages in R program.

In comparing the performance of DIF determination methods under various conditions, first Type I error and power ratios were calculated. RMSE, standard error and BIAS of item and ability parameters were calculated to compare the results of each DIF determination method. After the calculations, factorial ANOVA analyzes were performed to test the differentiation of the item and ability parameters according to the methods and the interaction of the simulation conditions with the methods. As a result of the simulation study, it was found that Type I error and power ratios increased according to the methods as the sample size, DIF magnitude and the ratio of the items with DIF in the test increased. It was found that Type I error and power ratios were relatively high when focus and reference group sample sizes were equal. It was found that the RMSE values of the discrimination parameter decreased as the sample size increased but these values did not change according to the sample size ratios, DIF magnitude and ratio of the items with DIF in the test. In addition, it was concluded that the RMSE values of the discrimination parameter did not differ according to DIF determination methods. It was determined that the RMSE values of the difficulty parameter decreased as the sample size increased, and these values differed according to the methods. However, it was found that the RMSE values of the difficulty parameter did not change according to the sample size ratios, DIF magnitude and ratio of the items with DIF in the test. It was found that the RMSE values of the ability parameter did not differ according to the DIF methods, but these values differed according to the sample sizes, DIF magnitude and ratio of the items with DIF in the test. As the sample size, DIF magnitude and ratio of the items with DIF in the test increased, the RMSE values of the ability parameter increased. While the sample size ratio of focus and reference groups was equal (1/1), the RMSE values of the ability parameter were found to be relatively high in comparison with the sample size ratio of 1/2.

Keywords : Differential Item Functioning, Item and Ability Parameters Estimates, Type I Error Ratio, Power Ratio

Page Number: xx + 136

Supervisor: Associate Professor 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.....	xv
ŞEKİLLER LİSTESİ	xviii
SİMGELER VE KISALTMALAR LİSTESİ.....	xix
BÖLÜM I.....	1
GİRİŞ	1
1.1. Problem Durumu	1
1.2. Araştırmanın Amacı	5
1.3. Araştırmanın Önemi.....	6
1.4. Sayıtlar.....	7
1.5. Sınırlılıklar.....	7

BÖLÜM II	9
KAVRAMSAL ÇERÇEVE	9
2.1. Değişen Madde Fonksiyonu	9
2.1.1. Madde Yanlılığı, Madde Etkisi ve Değişen Madde Fonksiyonu.....	11
2.1.2. Değişen Madde Fonksiyonu Çalışmalarının Gelişimi	12
2.1.3. Değişen Madde Fonksiyonu ile İlgili Temel Kavramlar	13
2.1.4. Değişen Madde Fonksiyonunun Belirlenmesinde Kullanılan Yöntemler...14	
2.1.4.1. Mantel-Haenszel Yöntemi	16
2.1.4.2. SIBTEST Yöntemi	18
2.1.4.3. Lojistik Regresyon Yöntemi.....	20
2.1.4.4. Lord'un χ^2 Testi Yöntemi	21
2.1.4.5. Raju'nun Alan Ölçümü Yöntemi	23
2.2. Madde Tepki Kuramı	24
2.2.1. Madde Tepki Kuramının Varsayımları.....	25
2.2.2. Madde Tepki Kuramı Modelleri	26
2.3. İlgili Araştırmalar	29
BÖLÜM III	39
YÖNTEM.....	39
3.1. Araştırmanın Türü	39
3.2. Veriler	39
3.2.1. Simülasyon Koşulları.....	40
3.2.1.1. Testte Bulunan Madde Sayısı	40
3.2.1.2. Kullanılan MTK Modeli.....	40
3.2.1.3. DMF Türü.....	40

3.2.1.4. Örneklem Büyüklüğü ile Odak ve Referans Grupların Örneklem Büyüklüğü Oranları.....	41
3.2.1.5. Testte Bulunan DMF’li Maddelerin Oranı.....	42
3.2.1.6. DMF Büyüklüğü	43
3.3. Veri Üretimi.....	45
3.4. Veri Analizi.....	46
BÖLÜM IV.....	52
BULGULAR VE YORUM.....	52
4.1. Farklı Koşullar Altında DMF Belirleme Yöntemlerinin I. Tip Hata Oranları.....	52
4.1.1. Farklı Örneklem Büyüklüklerine Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının Kıyaslanması	56
4.1.2. Farklı Örneklem Oranlarına Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının Kıyaslanması	57
4.1.3. Farklı DMF Büyüklüklerine Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının Kıyaslanması	58
4.1.4. Farklı DMF’li Madde Oranlarına Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının Kıyaslanması	60
4.2. Farklı Koşullar Altında DMF Belirleme Yöntemlerinin Güç Oranları	61
4.2.1. Farklı Örneklem Büyüklüklerine Göre DMF Belirleme Yöntemlerinin Güç Oranlarının Kıyaslanması.....	64
4.2.2. Farklı Örneklem Büyüklüğü Oranlarına Göre DMF Belirleme Yöntemlerinin Güç Oranlarının Kıyaslanması.....	66
4.2.3. Farklı DMF Büyüklüklerine Göre DMF Belirleme Yöntemlerinin Güç Oranlarının Kıyaslanması.....	67
4.2.4. Farklı DMF’li Madde Oranlarına Göre DMF Belirleme Yöntemlerinin Güç Oranlarının Kıyaslanması.....	68
4.3. Farklı Koşullar Altında DMF’li Maddelerin Testten Çıkarılmasında Madde Parametre Kestirimleri	70

4.3.1. Ayırt Edicilik Parametresine İlişkin RMSE Değerlerinin İncelenmesi.....	71
4.3.1.1. Farklı Örneklem Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin Kıyaslanması	73
4.3.1.2. Farklı Örneklem Oranları ve DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin Kıyaslanması.....	75
4.3.1.3. Farklı DMF Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin Kıyaslanması.....	75
4.3.1.4. Farklı DMF’li Madde Oranları ve DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin Kıyaslanması	76
4.3.2. Güçlük Parametresine İlişkin RMSE Değerleri.....	77
4.3.2.1. Farklı Örneklem Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin Kıyaslanması	80
4.3.2.2. Farklı Örneklem Büyüklüğü Oranları ve DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin Kıyaslanması.....	81
4.3.2.3. Farklı DMF Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin Kıyaslanması	82
4.3.2.4. Farklı DMF’li Madde Oranları ve DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin Kıyaslanması	83
4.4. Farklı Koşullar Altında DMF’li Maddelerin Testten Çıkarılmasında Yetenek Parametre Kestirimleri	83
4.4.1. Farklı Örneklem Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin Kıyaslanması.....	86
4.4.2. Farklı Örneklem Büyüklüğü Oranları ve DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin Kıyaslanması	88
4.4.3. Farklı DMF Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin Kıyaslanması.....	89
4.4.4. Farklı DMF’li Madde Oranları ve DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin Kıyaslanması.....	91
BÖLÜM V	93

SONUÇ VE TARTIŞMA	93
5.1. Sonuç ve Tartışma.....	93
5.2. Öneriler	100
KAYNAKLAR.....	102
EKLER	119
EK 1. 600 Örneklem Büyüklüğü için İlgili Koşullar Altında Ayırt Edicilik Parametresine İlişkin Standat Hata, RMSE ve Yanlılık Değerleri.....	120
EK 2. 1200 Örneklem Büyüklüğü için İlgili Koşullar Altında Ayırt Edicilik Parametresine İlişkin Standat Hata, RMSE ve Yanlılık Değerleri.....	121
EK 3. 2400 Örneklem Büyüklüğü için İlgili Koşullar Altında Ayırt Edicilik Parametresine İlişkin Standat Hata, RMSE ve Yanlılık Değerleri.....	122
EK 4. 600 Örneklem Büyüklüğü için İlgili Koşullar Altında Güçlük Parametresine İlişkin Standat Hata, RMSE ve Yanlılık Değerleri.....	123
EK 5. 1200 Örneklem Büyüklüğü için İlgili Koşullar Altında Güçlük Parametresine İlişkin Standat Hata, RMSE ve Yanlılık Değerleri.....	124
EK 6. 2400 Örneklem Büyüklüğü için İlgili Koşullar Altında Güçlük Parametresine İlişkin Standat Hata, RMSE ve Yanlılık Değerleri.....	125
EK 7. 600 Örneklem Büyüklüğü için İlgili Koşullar Altında Yetenek Parametresine İlişkin Standat Hata, RMSE ve Yanlılık Değerleri.....	126
EK 8. 1200 Örneklem Büyüklüğü için İlgili Koşullar Altında Yetenek Parametresine İlişkin Standat Hata, RMSE ve Yanlılık Değerleri.....	127
EK 9. 2400 Örneklem Büyüklüğü İçin İlgili Koşullar Altında Yetenek Parametresine İlişkin Standat Hata, RMSE ve Yanlılık Değerleri.....	128
EK 10. İlgili Koşullar Altında Yöntemlerin Belirlediği DMF’li Madde Oranları.	129
EK 11. Veri Üretmede Kullanılan Syntax	130
EK 12. Özgeçmiş	134
EK 13. Turnitin Orjinallik Raporu	135

TABLÖLAR LİSTESİ

Tablo 1. İki Kategorili Puanlanan Maddeler için DMF Belirlemede Kullanılan Yöntemler	15
Tablo 2. Odak ve Referans Grup için Oluşturulan 2x2'lik Dağılım	16
Tablo 3. Mantel-Haenszel Yönteminin Değerlendirilmesinde ETS Tarafından Önerilen Ölçütler	18
Tablo 4. SIBTEST Yönteminin Değerlendirilmesinde Önerilen Ölçütler	20
Tablo 5. Lojistik Regresyon Yönteminin Değerlendirilmesinde Önerilen Ölçütler	21
Tablo 6. Lord'un χ^2 Yönteminin Değerlendirilmesinde Önerilen Ölçütler	22
Tablo 7. NCDIF'in Değerlendirilmesinde Önerilen Ölçütler	24
Tablo 8. Simülasyon Deseni	44
Tablo 9. Simülasyon Çalışması İçin Üretilen Madde Parametre Değerleri	46
Tablo 10. I. Tip Hata ve Güç Oranlarının Hesaplanmasına İlişkin Örnek Tablo	49
Tablo 11. İlgili Koşullar Altında DMF Belirleme Yöntemlerinin I. Tip Hata Oranları	53
Tablo 12. Farklı Örneklem Büyüklüklerine Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının İncelenmesi	56
Tablo 13. Farklı Örneklem Oranlarına Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının İncelenmesi	58
Tablo 14. Farklı DMF Büyüklüklerine Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının İncelenmesi	58
Tablo 15. Farklı DMF'li Madde Oranlarına Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının İncelenmesi	60

Tablo 16. <i>İlgili Koşullar Altında DMF Belirleme Yöntemlerinin Güç Oranları</i>	62
Tablo 17. <i>Farklı Örneklem Büyüklüklerine Göre DMF Belirleme Yöntemlerinin Güç Oranlarının İncelenmesi</i>	64
Tablo 18. <i>Farklı Örneklem Büyüklüğü Oranlarına Göre DMF Belirleme Yöntemlerinin Güç Oranlarının İncelenmesi</i>	66
Tablo 19. <i>Farklı DMF Büyüklüklerine Göre DMF Belirleme Yöntemlerinin Güç Oranlarının İncelenmesi</i>	67
Tablo 20. <i>Farklı DMF'li Madde Oranlarına Göre DMF Belirleme Yöntemlerinin Güç Oranlarının İncelenmesi</i>	69
Tablo 21. <i>İlgili Koşullar Altında Ayırt Edicilik Parametresine İlişkin RMSE Değerleri</i>	72
Tablo 22. <i>Farklı Örneklem Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin İncelenmesi</i>	73
Tablo 23. <i>Farklı Örneklem Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin İncelenmesi</i>	75
Tablo 24. <i>Farklı DMF Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin İncelenmesi</i>	76
Tablo 25. <i>Farklı DMF'li Madde Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin İncelenmesi</i>	76
Tablo 26. <i>İlgili Koşullar Altında Güçlük Parametresine İlişkin RMSE Değerleri</i>	78
Tablo 27. <i>Farklı Örneklem Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin İncelenmesi</i>	80
Tablo 28. <i>Farklı Örneklem Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin İncelenmesi</i>	82
Tablo 29. <i>Farklı DMF Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin İncelenmesi</i>	82
Tablo 30. <i>Farklı DMF'li Madde Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin İncelenmesi</i>	83
Tablo 31. <i>İlgili Koşullar Altında Yetenek Parametresine İlişkin RMSE Değerleri</i>	85

Tablo 32. <i>Farklı Örneklem Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin İncelenmesi</i>	87
Tablo 33. <i>Farklı Örneklem Büyüklüğü Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin İncelenmesi</i>	89
Tablo 34. <i>Farklı DMF Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin İncelenmesi</i>	89
Tablo 35. <i>Farklı DMF'li Madde Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin İncelenmesi</i>	91

ŞEKİLLER LİSTESİ

<i>Şekil 1.</i> Tek biçimli ve tek biçimli olmayan DMF.....	14
<i>Şekil 2.</i> Farklı örneklem büyüklüklerinde DMF belirleme yöntemlerinin I. Tip hata oranları	57
<i>Şekil 3.</i> Farklı DMF büyüklüklerinde DMF belirleme yöntemlerinin I. Tip hata oranları ..	59
<i>Şekil 4.</i> Farklı DMF’li madde oranlarında DMF belirleme yöntemlerinin I. Tip hata oranları	61
<i>Şekil 5.</i> Farklı örneklem büyüklüklerinde DMF belirleme yöntemlerinin güç oranları	65
<i>Şekil 6.</i> Farklı DMF büyüklüklerinde DMF belirleme yöntemlerinin güç oranları	68
<i>Şekil 7.</i> Farklı DMF’li madde oranlarında DMF belirleme yöntemlerinin güç oranları.....	70
<i>Şekil 8.</i> Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre ayırt edicilik parametresinin RMSE değerleri.....	74
<i>Şekil 9.</i> Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre güçlük parametresinin RMSE değerleri.....	81
<i>Şekil 10.</i> Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre yetenek parametresinin RMSE değerleri.....	88
<i>Şekil 11.</i> Farklı DMF büyüklükleri ve farklı DMF belirleme yöntemlerine göre yetenek parametresinin RMSE değerleri.....	90
<i>Şekil 12.</i> Farklı DMF’li madde oranları ve farklı DMF belirleme yöntemlerine göre yetenek parametresinin RMSE değerleri.....	92

SİMGELER VE KISALTMALAR LİSTESİ

MTK	Madde Tepki Kuramı
KTK	Klasik Test Kuramı
DMF	Değişen Madde Fonksiyonu
DBF	Differential Bundle Functioning
LR	Lojistik Regresyon
M-H	Mantel-Haenszel
SIBTEST	Simultaneous Item Bias Test
ETS	Educational Testing Service
NCDIF	Noncompensatory Differential Item Functioning
RMSE	Root Mean Square Error
BIAS	İstatistiksel Yanlılık
SH	Standart Hata
1PLM	1 Parametrelili Lojistik Model
2PLM	2 Parametrelili Lojistik Model
3PLM	3 Parametrelili Lojistik Model
χ^2	Ki-kare
η^2	Eta-kare (Etki Büyüklüğü)
Δ	Delta
θ	Yetenek Parametresi
a	Ayırt Edicilik Parametresi

b	Güçlük Parametresi
c	Şans Parametresi

BÖLÜM I

GİRİŞ

Bu bölümde problem durumu ortaya konulmaya çalışılarak araştırmanın amacı, önemi, sayıltıları ve sınırlılıkları hakkında bilgilere yer verilmiştir.

1.1. Problem Durumu

Eğitim ve psikolojide eğitim kararlarının verilmesi, öğrencilerin güçlü ve zayıf oldukları alanların tespit edilmesi, öğrenme eksikliklerinin belirlenmesi, başarının saptanması ya da seçme ve yerleştirme gibi çeşitli amaçlar için testler kullanılır (Baykul, 2000). Farklı içerik alanlarını, becerileri ya da yetenekleri ölçmek için tasarlanan testler farklı hedef kitlelere uygulanır. Çeşitli amaçlar için kullanılan bu testler doğru şekilde uygulandıklarında testlerden elde edilen puanlar eğitimsel hedeflere ulaşmada, bireysel ve kurumsal düzeylerde alınacak kararlarda öğrenciler hakkında önemli bilgiler sağlar (Kolen & Brennan, 2014). Testlerden elde edilen puanlara dayalı olarak verilen kararların isabetli olması ise ölçme sonuçlarının geçerli ve güvenilir olmasıyla yakından ilişkilidir. Verilen kararların geçerliğini tehdit eden faktörlerden biri ise yanlılıktır (Clauser & Mazor, 1998). Testlerde yer alan maddelerin yanlı olmasının araştırmacılar, sınava girenler ve politikacılar için bir takım önemli etkileri vardır (Zumbo & Gelin, 2005). Örneğin yanlılık gösteren bir testin kullanılması, sınava giren gruptan birine avantaj sağlayarak o gruptaki bireylerin seçilme oranlarını artırırken, diğer grubun aleyhine işlev göstererek seçilme oranlarını azaltacaktır. Sınava girenlerin eşit fırsatlara sahip olmalarını sağlayamayan test ise insan haklarına uygun olmamasının yanı sıra toplumunda bu tür testlerin kullanımına karşı güven algılarını olumsuz yönde etkileyecektir.

Testi alan bireylerin eşit fırsatlara sahip olup olmadıklarını belirlemenin ilk adımı olan değişen madde fonksiyonu (DMF) alt gruplar arasındaki gerçek farklılık (madde etkisi) veya madde yanlılığı olmak üzere iki şekilde ortaya çıkmaktadır (Camilli & Shepard, 1994). Zumbo (1999)'ya göre madde etkisi olması halinde farklı gruplardaki bireylerin maddeyi doğru yanıtlama olasılıkları farklılaşmakta ve bu farklılık ölçülmesi amaçlanan özellik bakımından gruplar arasındaki gerçek farklılıktan kaynaklanmaktadır. Ayrıca bu gerçek farklılık gruplardan birinin geçmişte sahip olduğu tecrübe veya bilgi ile açıklanmaktadır (Zumbo, 2007). Madde yanlılığı ise aynı yetenek düzeyinde olan iki grubun testin amacına uygun olmayan ve testteki maddelerin özellikleri veya test koşullarından kaynaklı olarak bireylerin belirli bir maddeyi doğru yanıtlama olasılıklarının farklılaşmasıdır (Zumbo, 1999). Angoff (1993) ise yanlılığı "Farklı alt gruplarda yer alan ve yetenek bakımından eşit olan bireylerin ilgili test maddesini doğru yanıtlama olasılıklarının eşit olmaması durumudur." şeklinde tanımlamıştır (s. 4).

Yanlılık aslında istatistiksel ve sosyal olmak üzere en az iki farklı anlama gelmektedir (Angoff, 1993). Değişen madde fonksiyonu çalışmaları yanlılığın istatistiksel bir göstergesi olarak düşünülebilir. Çünkü DMF testi alan gruplar arasında farklılık gösteren madde yanıt örüntülerini tanımlamak için kullanılan ve bu sayede yanlı olan maddeleri belirlemeye yardımcı olan istatistiksel bir süreçtir (Zumbo & Gelin, 2005). Chaimongkol, Huffer ve Kamata (2007)'ya göre DMF, aynı yetenek düzeyine sahip olan ve farklı gruplarda yer alan bireylerin maddeyi doğru cevaplama olasılıklarının farklı olması durumudur. Test maddelerinde DMF olması durumunun sonucuna bağlı olarak ise farklı gruplardaki bireylerin toplam test puanları veya yetenek düzeylerinin kestirimleri farklılık gösterir (Chaimongkol vd., 2007). Belirli bir yetenek düzeyinde eşdeğer olan bireylerin maddeleri yanıtlamadaki grup farklılıklarının belirlenmesine olanak sağlayan DMF çalışmaları ise geçerliğin değerlendirilmesinde önemli bir bileşendir ve DMF varlığının bir testin geçerliğini azaltabileceği kabul edilmektedir (French & Finch, 2013, 2015; Li & Zumbo, 2009; Thissen, Steinberg & Wainer, 1988). Bu durum ise farklı gruplarda yer alan bireylerin test puanlarının hatalı yorumlanmasına neden olmaktadır. Özellikle eğitim ve psikoloji alanlarında testlerin değerlendirme araçları olarak kullanılması ve test sonuçlarına dayalı olarak alınan kararların önemide göz önüne alındığında testte yer alan bir maddenin DMF gösterip göstermeme durumu psikometristler için endişe kaynağı olmaktadır (Balluerka, Gorostiaga, Gómez-Benito & Hidalgo, 2010; Thissen vd., 1988, 1993). Karar verme amaçlı olarak testlerin kullanımındaki artış nedeniyle, test ile ölçülmek istenen yapıyla ilgili

olmayan deęişkenlerin incelenmesi ölçme araçlarının geçerliğini deęerlendirmede merkezi bir öneme sahiptir. (Balluerka vd., 2010).

DMF belirlemeye yönelik Mantel-Haenszel (Holland & Thayer, 1988), SIBTEST (Shealy & Stout, 1993), Lojistik Regresyon (Swaminathan & Rogers, 1990), Raju'nun alan ölçümü (Raju, 1988, 1990) ve Lord'un χ^2 (Lord, 1980) gibi birçok yöntem geliştirilmiştir. Bu yöntemler farklı eşleştirme kriterleriyle (örtük deęişken, gözlenen puan) farklı algoritmalar kullandıklarından ve maddeleri DMF'li olarak belirlemede kullandıkları kesme noktaları da farklılaştığından dolayı tam bir uyum içinde deęildirler (Doęan & Öğretmen, 2008; Gök, Kelecioęlu & Doęan, 2010). Bu nedenlerle DMF'yi belirlemek için güçlü bir istatistiksel yöntem üzerinde fikir birliğine varılamamıştır. Kullanılan bu yöntemlerle DMF belirlenirken “odak” ve “referans” olmak üzere iki grubun maddelere verdikleri yanıtlar karşılaştırılmakta ve yapılan analizlerin ardından her bir maddede DMF olup olmadığı ve varsa düzeyinin ne olduğuna ilişkin sınıflandırma yapılmaktadır (Swanson, Clauser, Case, Nungester & Featherman, 2002). DMF'nin neden ortaya çıktığına ilişkin önerilen açıklama ise çok boyutluluk ilkesine dayanmaktadır (Ackerman'den aktaran Balluerka vd., 2010). Maddenin DMF'li olması durumu çok boyutluluk açısından bakıldığında, maddenin ölçmeyi amaçladığı yetenek veya özelliğın yanı sıra istenmeyen bir yetenek veya özelliğı ölçtüęü anlamına gelmektedir (Balluerka vd., 2010; Roussos & Stout, 1996).

Eğitim alanında farklı amaçlarla ulusal ve uluslararası düzeylerde “Akademik Becerilerin İzlenmesi ve Deęerlendirilmesi (ABIDE)”, “Temel Eğitimden Ortaöğretime Geçiş (TEOG)”, “Yabancı Dil Sınavı (YDS)”, “Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES)”, “Uluslararası Öğrenci Deęerlendirme Programı (PISA)” ve “Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS)” gibi uygulamalar yapılmaktadır. Bu uygulamalarla kurumsal düzeyde kararlar alabilmek için ölçme ve deęerlendirme çalışmaları gerçekleştirilir. Geniş ölçekli bu deęerlendirmelerde DMF çalışmaları standart uygulamaların rutin bir parçasıdır (Taylor & Lee, 2012). Eğitimde ve Psikolojik Ölçmelerde Test Standartlarında da potansiyel olarak yanlış olan maddeleri araştırmanın ve DMF'ye neden olan kaynakları belirlemenin önemi vurgulanmıştır. Böylece DMF'nin kaynağının tespit edilmesi ile kaliteli test maddelerinin oluşturulması ve testi alan bireylerin gerçek yeteneklerinin daha iyi tahmin edilebileceğı ima edilmektedir.

DMF çalışmalarında sadece DMF'nin varlığını belirlemek ise testin geçerliğı hakkında sonuç çıkarmak için yeterli deęildir. DMF'li maddelerin belirlenmesinin ardından bu

maddelerle ilgili alınacak kararlar önemlidir. Çünkü sadece DMF’li maddelerin belirlenmesi, DMF’nin test istatistikleri üzerindeki etkileri konusunda sorunun çözülmesine yeterli olmayabilir. Daha açık ifadeyle bir testte bir ya da daha fazla maddede DMF belirlendiğinde bu maddelerin puanlamaya dahil olup olmamasıyla ilgili karar vermek gereklidir (Golia, 2014). Değişmezlik şartı ise bu maddelerin dahil edilmemesini önermektedir. Bununla birlikte bu maddeler, ölçülmek istenen özelliği ya da ölçülmesi amaçlanan yapıyı temsil eden maddeler olabilir. Böyle bir durumda ise maddelerin puanlara dahil edilmemesi uygun değildir ve maddelerin yeniden gözden geçirilmesinde DMF sorununu çözmekte yeterli olmaz (Golia, 2014).

Diğer taraftan yetenek kestirimleri DMF’nin etkisi hakkında önemli bilgiler sağlayabilir. Bu konunun önemine rağmen çalışmalar beklenmedik şekilde azdır (Wen, 2014). Ayrıca yetenek kestirimlerinin doğruluğu testin uzunluğuyla bağlantılıdır. Dolayısıyla testteki maddelerin bir ya da daha fazlasının testten çıkarılma kararı, yetenek kestirimlerinin doğruluğunu azaltabilir. Buna rağmen DMF’li maddelerin testten çıkarılmasının yetenek kestiriminin doğruluğu üzerinde önemli bir yanlılık (BIAS) etkisinin olup olmadığı kestirilememektedir (Golia, 2014). Testte sadece bir tane DMF’li madde bulunduğunda DMF türünden bağımsız olarak, yetenek kestirimi üzerinde yanlı (BIAS) bir etkinin olduğuna ilişkin kanıt bulunmamaktadır. Diğer taraftan testte birden fazla maddede DMF’nin olması ise daha ciddi etkiler yaratabilir (Golia, 2014).

Alanyazında maddelerin testten çıkarılmasının test istatistiklerini nasıl etkilediği konusunda çalışmalar yer almaktadır (Golia, 2010, 2014; Li & Zumbo, 2009; Roznowski & Reith, 1999; Rupp & Zumbo, 2003, 2006; Tennant & Pallant, 2007; Well, Subkoviak & Serlin, 2002). Maddelerin testten çıkarılması durumunu DMF kapsamında ele alarak parametre değişmezliğini inceleyen çalışmalarda mevcuttur (Golia, 2010, 2014; Lee & Zhang, 2017; Li & Zumbo, 2009; Tennant & Pallant, 2007). Fakat alanyazında yer alan bu çalışmalar incelendiğinde DMF belirleme yöntemlerine göre farklı koşullar altında DMF’li maddelerin testten çıkarılması durumunu hem madde hem de yetenek parametre kestirimlerine dayalı inceleyen çalışmaya rastlanılmamıştır. Bu nedenle bu araştırmada iki kategoride puanlanan test maddelerinde farklı test koşulları (örneklem büyüklüğü; odak ve referans grupların örneklem büyüklüğü oranları; DMF düzeyi ve testte yer alan DMF’li madde oranları) altında çeşitli DMF belirleme yöntemlerine göre madde ve yetenek parametre kestirimlerinin standart hata, RMSE ve yanlılık değerleri incelenmiştir. Bu sayede DMF’li maddelerin sayısı

ve DMF büyüklüğü bakımından yetenek kestirimi üzerindeki etkinin bilinmesiyle uygulayıcıların test geliştirme sürecinde daha iyi kararlar vermelerine yardımcı olabilir (Wen, 2014). Çünkü testte bulunan DMF'li maddelerin varlığın yetenek kestiriminden etkilenmiyorsa test doğrudan kullanılabilir. Fakat yetenek kestirimi DMF'li madde sayısı ve DMF düzeyinden etkileniyorsa maddelerin değiştirilmesi gerekmektedir (Wen, 2014).

1.2. Araştırmanın Amacı

Araştırmanın genel amacı farklı test koşulları (örneklem büyüklüğü; odak ve referans grupların örneklem büyüklüğü oranları; DMF düzeyi ve testte yer alan DMF'li madde oranları) altında DMF belirleme yöntemlerini madde ve yetenek parametre kestirimlerine dayalı olarak incelemektir. Bu doğrultuda aşağıda yer alan sorulara yanıt aranmıştır.

1. DMF belirleme yöntemlerinin (Mantel-Haenszel, Lord'un χ^2 , Lojistik Regresyon, SIBTEST ve Raju'nun alan ölçümü) I. Tip hata oranları;

- örneklem büyüklüğü (600, 1200 ve 2400),
- odak ve referans grup örneklem büyüklüğü oranları (1/1 ve 1/2),
- DMF büyüklüğü (0,5 ve 1,0) ve
- testte bulunan DMF'li madde oranı (%5, %10 ve %20)

koşullarında nasıl değişmektedir?

2. DMF belirleme yöntemlerinin (Mantel-Haenszel, Lord'un χ^2 , Lojistik Regresyon, SIBTEST ve Raju'nun alan ölçümü) güç oranları;

- örneklem büyüklüğü (600, 1200 ve 2400),
- odak ve referans grup örneklem büyüklüğü oranları (1/1 ve 1/2),
- DMF büyüklüğü (0,5 ve 1,0) ve
- testte bulunan DMF'li madde oranı (%5, %10 ve %20)

koşullarında nasıl değişmektedir?

3. DMF belirleme yöntemlerine göre (Mantel-Haenszel, Lord'un χ^2 , Lojistik Regresyon, SIBTEST ve Raju'nun alan ölçümü);

- örneklem büyüklüğü (600, 1200 ve 2400),

- b. odak ve referans grup örneklem büyüklüğü oranları (1/1 ve 1/2),
- c. DMF büyüklüğü (0,5 ve 1,0) ve
- d. testte bulunan DMF’li madde oranı (%5, %10 ve %20)

koşullarında madde parametre kestirimlerinin RMSE, standart hata ve yanlılık değerleri nasıl değişmektedir?

4. DMF belirleme yöntemlerine göre (Mantel-Haenszel, Lord’un χ^2 , Lojistik Regresyon, SIBTEST ve Raju’nun alan ölçümü);

- a. örneklem büyüklüğü (600, 1200 ve 2400),
- b. odak ve referans grup örneklem büyüklüğü oranları (1/1 ve 1/2),
- c. DMF büyüklüğü (0,5 ve 1,0) ve
- d. testte bulunan DMF’li madde oranı (%5, %10 ve %20)

koşullarında yetenek parametre kestirimlerinin RMSE, standart hata ve yanlılık değerleri nasıl değişmektedir?

1.3. Araştırmanın Önemi

Ölçme aracında bulunması gereken en önemli özelliklerden biri geçerliktir ve DMF varlığının bir testin geçerliğini azaltabileceği kabul edilmektedir (Baykul, 2000; French & Finch, 2013, 2015; Li & Zumbo, 2009; Thissen vd., 1988). Yapılan sınavlarda kullanılan test maddelerinde DMF’nin olması bu sınav sonuçlarının geçerliğini olumsuz yönde etkileyecek ve bireylerin puanlarının yanlış yorumlanmasına neden olacaktır. Bu nedenle test puanlarını yorumlamada geçerliği desteklemek için DMF çalışmalarının yapılması önemlidir.

DMF belirlemeye yönelik birçok yöntemin geliştirildiği görülmektedir. Fakat bu yöntemlerin kullandıkları eşleştirme kriterlerinin ve algoritmaların farklılaşmasından kaynaklı olarak yöntemlerin farklı sonuçlar ürettiği yapılan alanyazın incelemesinden anlaşılmaktadır. Bu araştırma ile farklı test koşulları altında çeşitli DMF yöntemlerine göre madde ve yetenek parametre kestirimleri incelenmiştir. Testte bulunan DMF’li madde oranları, örneklem büyüklüğü ve DMF büyüklüğünün yetenek kestirimine olan etkilerinin

bilinmesi uygulama yapanların test geliştirme sürecinde doğru karar vermelerine yardımcı olabilir (Wen, 2014).

Ayrıca bu araştırmada farklı DMF yöntemlerine göre madde ve yetenek parametre kestirimlerinin incelenmesinin yanı sıra yöntemlerin farklı koşullar altında I. Tip hata ve güç oranları da incelenmiştir. Değişen farklı koşullar altında bu yöntemlerin hangilerinin daha tutarlı sonuçlar verdiğinin bilinmesi testi geliştirenler ve uygulayanlar için testten elde edilen puanların daha geçerli ve adil olmasını sağlayacaktır. Alanyazında DMF belirleme yöntemlerini karşılaştıran çalışmalar yer almaktadır (Akin Arıkan, Uğurlu & Atar, 2015; Atalay Kabasakal, Arsan, Gök & Kelecioğlu, 2014; Awuor, 2008; Cohen, Kim & Wollack, 1996; Dainis, 2008; Demir & Köse, 2014; Erdem, 2015; Erdem Keklik, 2012; Finch & French, 2007; Jodoin & Gierl, 2001; Kim, 2010; Tunç, 2016; Özdemir, 2015; Özmen, 2013). Fakat araştırma kapsamında incelenen yöntemleri (Mantel-Haenszel, Lojistik Regresyon, Lord'un χ^2 , SIBTEST ve Raju'nun alan ölçümü) bir arada ele alan çalışmalara rastlanmamıştır. Ayrıca bu çalışma DMF belirlemede kullanılan yöntemler, test uzunluğu, yanıt kategorisi, örneklem büyüklüğü ve DMF büyüklüğü açılarından diğer simülasyon çalışmalarına göre de farklılaşmaktadır (Golia, 2014; Lee & Zhang, 2017; Li & Zumbo, 2009). Bu nedenle bu araştırma sonuçlarının; örneklem büyüklüğü, odak ve referans grup örneklem büyüklüğü oranları, DMF büyüklüğü ve testte bulunan DMF'li madde oranları farklılaştığında kullanılabilecek uygun DMF yöntemlerinin işe koşulmasında ulusal ve uluslararası geniş ölçekli test uygulayıcılarına, araştırmacılara ve test geliştiricilerine katkı sağlayacağı düşünülmektedir.

1.4. Sayıtlılar

Bu araştırmada simülasyon veri üretimi alanyazında yer alan çalışmalar ve manipüle edilmemiş gerçek veriler dikkate alınarak gerçekleştirilmiştir. Bu itibarla üretilen verilerin gerçek uygulama durumlarını yansıttığı varsayılmaktadır.

1.5. Sınırlılıklar

1. Araştırma sonuçları ele alınan simülasyon koşulları (örneklem büyüklüğü 600, 1200 ve 2400; odak ve referans grup örneklem büyüklüğü oranı 1/1 ve 1/2; DMF büyüklüğü 0,5 ve

1,0; testte bulunan DMF'li madde oranı %5, %10 ve %20), 2 kategoride puanlanan maddeler, 40 maddelik test uzunluđu ve 2PL model ile sınırlıdır.

2. Bu araştırma Mantel-Haenszel, Lord'un χ^2 , Lojistik Regresyon, SIBTEST ve Raju'nun alan ölçümü yöntemleri ile sınırlandırılmıştır.

3. Araştırmada beş farklı DMF belirleme yöntemine göre DMF'li olarak belirlenen maddeler testten çıkarılarak RMSE, standart hata ve yanlılık değeri hesaplanmıştır.

BÖLÜM II

KAVRAMSAL ÇERÇEVE

2.1. Değişen Madde Fonksiyonu

Eğitim alanında başarının saptanması, akademik performansın değerlendirilmesi, tanı koyma, seçme ve yerleştirme gibi çeşitli amaçlarla testler kullanılır. Test sonuçlarının geçerli olabilmesi için bireylerin belirli bir gruba olan üyeliğinin bu sonuçlar üzerinde etkisinin olup olmadığını belirlemeye ihtiyaç vardır. Bu grup üyeliği cinsiyet, etnik köken ya da sosyo-ekonomik durum gibi çeşitli demografik özellikler olarak tanımlanabilir. Grup üyeliğinin test sonuçları üzerinde etkisinin olup olmadığını anlamak için maddelerin DMF içerip içermediğini test etmek önemlidir. Bir maddede DMF'nin belirlenmesi ise gizil özellik dışındaki bazı özelliklerin (cinsiyet, sosyo-ekonomik durum vb.) madde puanını etkileyebileceğini gösterir (Ryan, 2008).

DMF çalışmalarında odak ve referans olmak üzere iki grubun performansları karşılaştırılır. Odak grup genelde ilgili madde tarafından dezavantajlı olup olmadığı araştırılan grup iken, referans grup ise karşılaştırma amacıyla kullanılan gruptur (French & Finch, 2013). DMF aynı yetenek düzeyine sahip olan odak ve referans gruptaki bireylerin maddeyi doğru yanıtlama olasılıklarının farklı olması durumudur (Chaimongkol vd., 2007). DMF genel olarak bir maddenin istatistiksel özelliklerinin alt gruplar arasında tutarsız olduğu anlamına gelmektedir. Burada maddenin istatistiksel özellikleri ile madde güçlüğü ve madde ayırt ediciliği kastedilmektedir. Eğitimde kullanılan ve maddenin doğru yanıtının bulunduğu testlerde (başarı testi gibi) madde güçlüğü, maddeye doğru yanıt verenlerin oranı olarak tanımlanmaktadır. Madde ayırt ediciliği ise bir maddenin testle ölçülmek istenen yeterliğe (yapı) sahip olan ve olmayanları ayırt etme derecesini ifade etmektedir. Dolayısıyla DMF,

madde ayırt ediciliği ve madde güçlüğü alt gruplarda farklılık gösterdiğinde ortaya çıkmaktadır (Sireci & Rios, 2013).

Özellikle eğitim ve psikoloji alanlarında testlerin değerlendirme araçları olarak kullanılması ve test sonuçlarına dayalı olarak önemli kararların alınması durumu göz önüne alındığında, testte yer alan bir maddenin DMF gösterip göstermeme durumu psikometristler için endişe kaynağıdır (Balluerka vd., 2010; Thissen vd., 1988, 1993). DMF çalışmalarında test puanlarının geçerliğini değerlendirmeye yardımcı olan önemli bir bileşen olarak karşımıza çıkmaktadır (French & Finch, 2013, 2015). Karar verme amaçlı olarak testlerin kullanımındaki artış nedeniyle test ile ölçülmek istenen yapıyla ilgili olmayan değişkenlerin incelenmesi, ölçme araçlarının geçerliğini değerlendirmede merkezi bir öneme sahiptir (Balluerka vd., 2010). Çünkü DMF varlığının bir testin geçerliğini azaltabileceği kabul edilmektedir (Finch & French, 2013, 2015; Li & Zumbo, 2009; Thissen vd., 1988, 1993). Bu nedenle DMF çalışmaları test puanlarının geçerliğini desteklemede önemli çalışmalardır (Sireci & Rios, 2013). Diğer bir ifadeyle DMF çalışmalarına ilişkin elde edilen sonuçlar bir geçerlik kanıtıdır (Clauser & Mazor, 1998). Ayrıca test maddelerinde DMF olmasıyla aynı yetenek düzeyinde olan ve farklı alt gruplarda yer alan bireylerin toplam test puanları veya yetenek düzeylerinin kestirimleri farklılık gösterir (Chaimongkol vd., 2007). Bu durum bireylerin test puanlarının hatalı yorumlanmasına neden olmaktadır. DMF çalışmaları aynı zamanda gruplar arasındaki değişmezlik varsayımının değerlendirilmesine de yardımcı olur ve bu sayede alınan puanların bütün bireyler için adil olmasını sağlar (French & Finch, 2013).

DMF çalışmaları potansiyel olarak yanlı olan maddeleri belirlemeyi amaçlamakta ve test geliştiricilere yanlı olmayan maddeleri oluşturmalarında rehberlik etmektedir. Fakat DMF'nin varlığı her zaman yanlılığın olduğu anlamına gelmeyebilir. Bir maddede DMF tespit edilmemişse madde yanlı değildir. Dolayısıyla madde yanlılığının araştırılabilmesi için DMF çalışmaları gerekli fakat yeterli değildir (Perrone, 2006). Ayrıca DMF'ye neden olan faktör ya da faktörler test tarafından ölçülen yapı ile ilgili değilse bu durum bir yanlılık kanıtıdır (Wen, 2014). Çünkü test veya madde yanlılığı birçok nedenle ortaya çıkabilir ve DMF analizleri sadece madde puanlarına ve grup göstergelerine dayanan istatistiksel kanıtlar sağlar (Wen, 2014). Özetle, testte yer alan madde veya maddelerde DMF tespit edilmesi halinde, ilgili madde/maddelerde yanlılığın olup olmadığının belirlenebilmesi için ampirik değerlendirme ya da içerik analizlerinin yapılması gerekir (Zumbo, 1999). Çünkü

maddenin farklı alt gruplar arasında neden farklı şekilde çalıştığını yani DMF'nin olası kaynaklarının belirlenmesi ileriye yönelik test hazırlayanlara yararlı bilgiler sağlar (Chaimongkol vd., 2007).

2.1.1. Madde Yanlılığı, Madde Etkisi ve Değişen Madde Fonksiyonu

Yanlılık, belirli bir amaçla uygulanan ölçme aracının farklı alt gruplar arasında oluşan puan farklılıklarının yorumlanmasını zorlaştıran bozucu faktörlere karşılık gelmektedir (Hambleton, Merenda & Spielberger, 2017). Birçok durumda testte yer alan maddeler ölçülmek istenen yapı ile ilgili olmayan özellikleri içerebilir ve bu özellikler bireylerin performansını etkileyebilir (Zumbo, 1999). Cinsiyet, kültür veya etnik köken vb. özelliklere bakılmaksızın test ile ölçülmek istenen özellik bakımından benzer bilgiye sahip olan bireylerin test maddelerinde de benzer performans göstermeleri beklenir (Subkoviak, Mack, Ironson ve Craig'den aktaran Perrone, 2006). Örneğin geniş ölçekli bir değerlendirmede kız ve erkek öğrencilerin eşit fırsatlara sahip olmaları ve bunun neticesinde testten aldıkları puanlar bakımından eşit muamele görmeleri beklenir (Zumbo & Gelin, 2005). Testi alan bireylerin eşit fırsatlara sahip olup olmadıklarını belirlemenin ilk adımını ise DMF çalışmaları oluşturmaktadır (Camilli & Shepard, 1994). Diğer bir ifadeyle DMF çalışmaları, alt gruplar arasında farklılık gösteren madde yanıt örüntülerini tanımlamak için kullanılan ve bu sayede yanlı olan maddeleri belirlemeye yardımcı olan istatistiksel bir süreçtir (Zumbo & Gelin, 2005).

DMF testi alan alt gruplar arasındaki gerçek farklılık (madde etkisi) veya madde yanlılığı olmak üzere iki şekilde ortaya çıkmaktadır. Zumbo (1999)'ya göre madde etkisi olması halinde maddeyle ölçülen özelliğe göre alt gruplar arasında gerçek bir farklılık vardır. Bu nedenle alt gruplarda yer alan bireylerin maddeyi doğru cevaplama olasılıkları farklılaşmaktadır. Bu gerçek farklılık ise alt gruplardan birinin geçmişte sahip olduğu tecrübe veya bilgi ile açıklanmaktadır (Zumbo, 1999). Madde yanlılığı ise Angoff (1993) tarafından "Farklı alt gruplarda yer alan ve yetenek bakımından eşit olan bireylerin ilgili test maddesini doğru yanıtlama olasılıklarının eşit olmaması durumudur." şeklinde tanımlanmıştır (s. 4). Zumbo (1999, 2007)'ya göre madde yanlılığı, testin amacına uygun olmayan ve testteki maddelerin özellikleri veya test koşullarından kaynaklı olarak, aynı yetenek düzeyinde olan iki alt grubun belirli bir maddeyi doğru yanıtlama olasılıklarının farklılaşmasıdır. Eğitimde ve Psikolojik Ölçmelerde Test Standartlarında (2014) ise yanlılık

yapıyla ilişkisiz olan faktörler nedeniyle katılımcı grup açısından sistematik olarak daha düşük veya daha yüksek puanlarla sonuçlanmasıdır.

Yanlılık seçme ve sınıflandırmada yapılan çıkarımları etkileyen sistematik hatalara neden olabilir (Zumbo, 1999). Diğer bir ifadeyle, farklı alt gruplarda yer alan bireylerin test performanslarının bulundukları alt gruplara bağlı olarak sistematik hata içermesidir. Bu sistematik hataların istatistikteki karşılığı olan yanlılık test sonuçlarının geçerliğini tehdit etmektedir (Camilli & Shepard, 1994; Zumbo, 1999). Testin tamamının ya da bir maddenin yanlı olması durumunda sınava giren bireylerin gerçek yeteneğiyle ilgili yapılacak olan çıkarımlar ise böyle bir test temelinde doğru bir şekilde yapılamayabilir (Camilli & Shepard, 1994). Ayrıca böyle bir durumda yetenek düzeyi benzer olan öğrencilerin başarı şansları da eşit olmayacaktır (Zumbo, 1999). Dolayısıyla testte yer alan yanlı maddelerin tespit edilmesi test sonuçlarına dayalı olarak alınacak kararların adil olarak değerlendirilmesini sağlar (Camilli & Shepard, 1994). Yanlı olduğu tespit edilen maddelerin değiştirilmesi ya da testten çıkarılması ise testin geçerliğini artıracak ve alt gruplar için testin adil olarak değerlendirilmesini sağlayacaktır (Camilli & Congdon, 1999). Testte yer alan yanlı maddeleri belirlemedeki ilk adım ise sınava giren bir alt grubun diğer alt gruba göre önemli ölçüde daha iyi performans gösterdiği test maddelerini belirlemektedir (Roever'den aktaran Perrone, 2006). Bu nedenle DMF çalışmaları yanlılığın istatistiksel bir göstergesidir (Zumbo & Gelin, 2005).

2.1.2. Değişen Madde Fonksiyonu Çalışmalarının Gelişimi

Zumbo (2007) DMF analizlerinin gelişimini üç kuşak olarak ele almaktadır. Birinci kuşak DMF çalışmalarında yaygın olarak madde yanlılığı teriminin kullanıldığını ve madde yanlılığı ile ilgili endişelerin ise başarı, yetenek, sertifika ve lisans testleri gibi testlere dayalı olarak alınan kararların adil olarak verilmesi bağlamında ortaya çıktığını aktarmıştır. Test yanlılığı ile ilgili yapılan çalışmalarda cinsiyete veya etnik kökene dayalı gruplara göre performansın farklılaşmasına odaklanılmıştır. Madde yanlılığını belirlemede ilk defa kullanılan yöntemlerin sadece iki gruptaki bireyleri karşılaştırmaya, odak/referans veya azınlık/çoğunluk gibi terminolojiye ve çoklu puanlanan maddeler dışındaki ikili puanlanan maddelere odaklanıldığını aktarmıştır.

DMF çalışmalarının ikinci kuşağa geçişinde ise “madde yanlılığı” yerine “değişen madde fonksiyonu” terimi yaygın olarak kabul görmüştür. DMF'nin tanımlanmasıyla birlikte

madde etkisi ile madde yanlılığı ayrımı yapılmış ve bu ayrımları şekillendiren yeni DMF yöntemlerinin kullanılmasının da sinyali verilmiştir. İkinci kuşağa geçişle öncelikli olarak DMF'in neden oluştuğuna ilişkin açıklamalar sunulmaya çalışılmış, diğer bir ifadeyle DMF'li maddelerin keşfedilmesine odaklanılmıştır. Ayrıca testlerde DMF ve madde yanlılığı araştırmalarına her ne kadar ilgi duyulsa da kullanılan istatistiksel terminoloji ve yazılımın çoğunlukla pek çok araştırmacı tarafından erişilebilir olmadığını belirtmiştir. Ayrıca ikinci kuşakta DMF'li maddelerin belirlenmesinin ardından uzman görüşüne başvurulmuş ve bu sayede madde etkisi ile madde yanlılığının olup olmadığına ilişkin çalışmalar yapılmıştır. Buna rağmen DMF'nin kaynağını ve neden oluştuğunu bulma çabasının bu dönemde zor olduğunu belirtmiştir. DMF'in neden meydana geldiğinin araştırılması ise üçüncü kuşak DMF çalışmalarının başlangıcı olarak kabul görmektedir. Zumbo (2007) üçüncü kuşak DMF çalışmalarının öğrencilerin sınıflarda, sınıflarında okullar içinde kümелendiğı karmaşık veri yapıları üzerine odaklandığını ve birçok kavramsal değişkenin potansiyel olarak DMF kaynağı olabileceğini belirtmiştir.

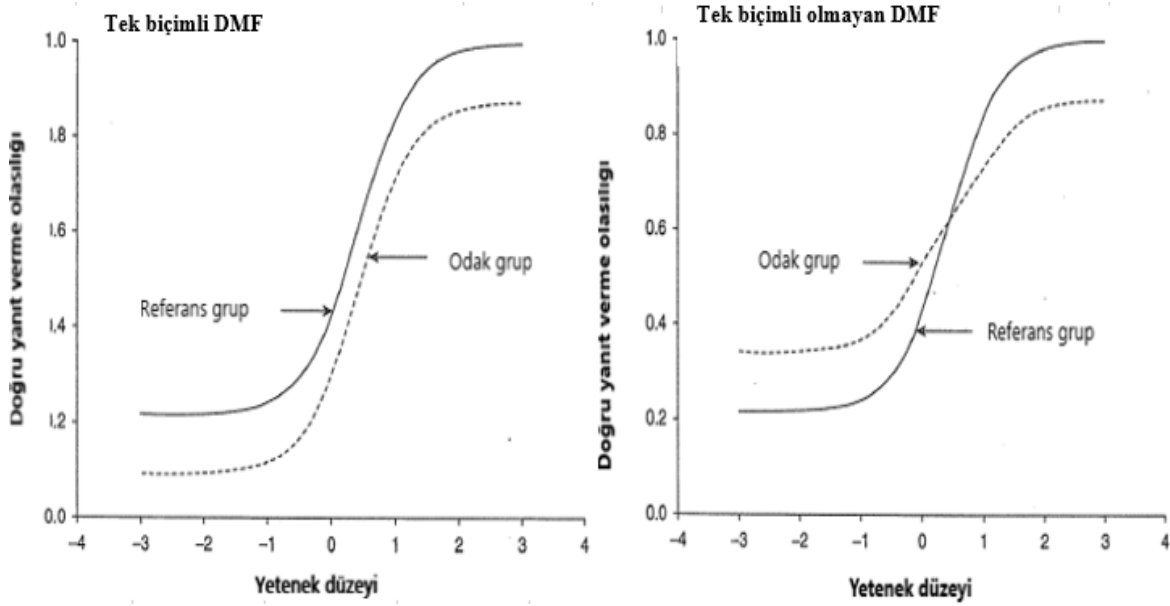
2.1.3. Değişen Madde Fonksiyonu ile İlgili Temel Kavramlar

Değişen madde fonksiyonu çalışmalarında geçen bazı temel kavramlar bu başlık altında açıklanmıştır.

DMF çalışmalarında bireyler ölçölmek istenen özelliğe göre gruplara ayrılırlar. “Referans” ve “odak” grup olarak tanımlanan bu gruplar ölçölen yetenek düzeyinde eşleştirilir ve daha sonra bu iki grubun performansları karşılaştırılır (Holland & Wainer, 1993; Osterlind, 1983). Literatürde bazen referans ve odak gruplarına sırasıyla çoğunluk ve azınlık grubu da denilmektedir (French & Finch, 2013).

DMF çalışmalarında yaygın olarak “tek biçimli” ve “tek biçimli olmayan” olmak üzere iki çeşit DMF türüne odaklanılmıştır. Tek biçimli DMF, belirli bir maddenin doğru yanıtlanma olasılığı tüm yetenek düzeylerinde bir grubun lehine işlediğinde ortaya çıkmaktadır. Belirli bir maddenin doğru yanıtlanma olasılığı farklı yetenek düzeylerinde farklı gruplar lehine işlediğinde ise tek biçimli olmayan DMF ortaya çıkmaktadır (Zumbo, 1999). Ayrıca sadece madde güçlük parametreleri odak ve referans grupları arasında farklılık gösterirse tek biçimli; madde ayırt edicilik ve madde güçlük parametreleri odak ve referans grupları için farklı ise tek biçimli olmayan DMF'nin varlığından söz edilir (Camilli & Shepard, 1994;

Gonzalez-Roma, Hernandez & Gomez-Benito, 2006; Zumbo, 1999). Tek biçimli ve tek biçimli olmayan DMF Şekil 1’de sunulmuştur.



Şekil 1. Tek biçimli ve tek biçimli olmayan DMF (Osterlind & Everson, 2009, s. 12).

2.1.4. Değişen Madde Fonksiyonunun Belirlenmesinde Kullanılan Yöntemler

Literatürde DMF belirlemeye yönelik Mantel-Haenszel (Holland & Thayer, 1988), SIBTEST (Shealy & Stout, 1993), Lojistik Regresyon (Swaminathan & Rogers, 1990), Lord'un χ^2 (Lord, 1980) ve Raju'nun alan ölçümü (Raju, 1988, 1990) gibi birçok yöntem geliştirilmiştir. Bu yöntemler farklı eşleştirme kriterleriyle farklı algoritmalar kullandıklarından ve DMF belirlemede kullandıkları kesme noktaları da farklılaştığından dolayı tam bir uyum içinde değildirler (Doğan & Öğretmen, 2008; Gök vd., 2010). Dolayısıyla yöntemlerin birbirine göre avantaj ve dezavantajlarının olması ve farklı yöntemlerin farklı sonuçlar vermesinden kaynaklı olarak DMF belirlemede birden fazla yöntemin kullanılması önerilmektedir (Camilli & Shepard, 1994; Holland & Wainer, 1993; Osterlind & Everson, 2009).

DMF belirlemede kullanılan yöntemler KTK ve MTK'ya dayalı olarak sınıflandırılabilir. Eğitim ve psikolojide MTK'nın uygulanması DMF'yi tanımlamada ve tespit etmede yeni bir yol sağlamıştır. Bu sayede a , b ve c parametrelerindeki farklılıklar ve bunların kombinasyonları incelenerek DMF ve DMF'nin büyüklüğü matematiksel olarak incelenebilir ve ölçülebilir (Lord, 1977, 1980). Çünkü aynı yetenek düzeyine sahip olan

bireylerin aynı maddeye doğru yanıt vermeleri durumunda daima madde karakteristik eğrileri aynı olacaktır. Buna rağmen iki grubun madde karakteristik eğrileri arasındaki fark DMF'nin varlığının bir göstergesidir (Hambleton, Swaminathan & Rogers, 1991). “SIBTEST, Lord’un χ^2 , Raju’nun alan testi, MTK olabilirlik oran, MTK-D², b farklarının testi, Sınırlandırılmış bilgi MTK-LR ve Log-doğrusal MTK-LR” yöntemleri ise MTK’ya dayalı yöntemlerdir. “Mantel-Haenszel, Ki-kare, Standartlaştırma, Delta grafiği, Log-Doğrusal modeller ve Lojistik regresyon” yöntemleri ise KTK’ya dayalıdır. Hem KTK hem de MTK’ya dayalı bu yöntemlerle DMF belirlenirken odak ve referans grupların maddelere verdikleri yanıtlar karşılaştırılmakta ve yapılan analizlerin ardından maddelerde DMF olup olmadığı ve varsa düzeyinin ne olduğuna ilişkin sınıflandırmalar yapılmaktadır (Swanson vd., 2002). İki kategoride puanlanan maddelerde DMF belirlemede kullanılan yöntemlerin sınıflandırılması (Potenza & Dorans, 1995) Tablo 1’de sunulmuştur.

Tablo 1.

İki Kategorili Puanlanan Maddeler için DMF Belirlemede Kullanılan Yöntemler

Eşleştirme değişkeni	Parametrik yöntemler	Parametrik olmayan yöntemler
Gözlenen puan	LR-DIF	Mantel-Haenszel
		Standartlaştırma
Örtük değişken	Genelleştirilmiş MTK-LR	SIBTEST
	Sınırlandırılmış bilgi MTK-LR	
	Loglineer MTK-LR	
	MTK-D ²	
	Lord’un χ^2	

Tablo 1’de yer alan sınıflandırmalara göre yöntemler eşleştirme değişkeninin gözlenen puan ve örtük değişken olup olmasına ve yöntemin parametrik olup olmasına göre farklılık göstermektedir (Potenza & Dorans, 1995). Çoğu geleneksel DMF belirleme yöntemi MTK’ya dayalı değildir ve tek biçimli DMF’yi tespit etmek için tasarlanmıştır. Mantel-Haenszel, Standartlaştırma ve SIBTEST yöntemleri olasılık tabloları için istatistiklere dayanmaktadır. Lojistik Regresyon yöntemi ise MTK’ya dayalı olan ve olmayan yöntemler arasında köprü görevi görmektedir (Magis, Beland, Teurlinckx & Boeck, 2010).

2.1.4.1. Mantel-Haenszel Yöntemi

Holland ve Thayer (1988) tarafından önerilen Mantel-Haenszel yöntemi DMF'yi tespit etmede yaygın olarak kullanılan yöntemlerden biridir. Grup üyeliği ve madde yanıtı arasında, toplam test puanına bağlı olarak bir ilişki olup olmadığını test etmeyi amaçlamaktadır (Magis vd., 2010). Mantel-Haenszel yönteminin hesaplanması ve yorumlanması kolay olmakla birlikte, anlamlılık testi ile DMF'nin büyüklüğü de kestirilmektedir (Millsap, 2011). Yöntem oldukça küçük örneklem (örneğin her grup için 200) üzerinde uygulanabilmektedir (Clauser & Mazor, 1998). Mantel-Haenszel yönteminin uygulanmasında ETS test programlarında örneklem büyüklüğünün 500 ve ayrıca odak grubun örneklem büyüklüğünün en az 100 (Kubiak ve Colwell'dan aktaran Awuor, 2008); Hills ise odak ve referans grubun örneklem büyüklükleri için 100'ün yeterli olduğunu belirtmiştir (Hills'den aktaran Awuor, 2008).

Mantel-Haenszel yönteminde gruplar (odak ve referans) toplam test puanı (gözlenen puan) üzerinden eşleştirilir ve her bir grup için $2 \times 2 \times A$ (gruplar - madde puanı - puan düzeyi) olasılık tablosu oluşturulur (Clauser & Mazor, 1998). Odak ve referans grup için oluşturulan 2×2 'lik dağılım Tablo 2'de sunulmuştur.

Tablo 2.

Odak ve Referans Grup için Oluşturulan 2×2 'lik Dağılım

Grup	1	0	Toplam
Referans	A_j	B_j	N_{Rj}
Odak	C_j	D_j	N_{Oj}
Toplam	M_{1j}	M_{0j}	T_j

Mantel-Haenszel χ^2 istatistiği ve test edilen hipotez Eşitlik 1, 2, 3 ve 4'te verilmiştir.

$$H_1: \frac{P_{Rj}}{Q_{Rj}} = a \frac{P_{Oj}}{Q_{Oj}}; \quad j=1, 2, \dots, k \quad (1)$$

$$\hat{a}_{MH} = \frac{\sum_j A_j D_j / T_j}{\sum_j B_j C_j / T_j} \quad (2)$$

Eşitlik 2'ye göre;

A_j ; j. maddeyi doğru yanıtlayan referans gruptaki bireylerin sayısı,

C_j ; j. maddeyi doğru yanıtlayan odak gruptaki bireylerin sayısı,

B_j ; j. maddeyi yanlış yanıtlayan referans gruptaki bireylerin sayısı,

D_j ; j. maddeyi yanlış yanıtlayan odak gruptaki bireylerin sayısı,

T_j ; Toplam birey sayısıdır.

Eşitlik 2’de elde edilen $a_{MH} < 1$ ise odak grup avantajlıdır, $a_{MH} = 1$ ise maddede DMF yoktur ve $a_{MH} > 1$ ise referans grup avantajlıdır (Holland ve Thayer, 1988).

$$MH\chi^2 = \frac{(|\sum_j A_j - \sum_j E(A_j)| - \frac{1}{2})^2}{\sum_j \text{var}(A_j)} \quad (3)$$

$$\text{var}(A_j) = \frac{N_{Rj}N_{Oj}M_{1j}M_{0j}}{T_j^2(T_j-1)} \quad (4)$$

Eşitlik 3 ve 4’e göre;

A_j ; j. maddeyi doğru yanıtlayan referans gruptaki bireylerin sayısı,

$E(A_j)$; A_j değerinin beklenen değeri,

$\text{var}(A_j)$; A_j değerinin varyansı,

N_{Rj} ; Referans gruptaki bireylerin sayısı,

N_{Oj} ; Odak gruptaki bireylerin sayısı,

M_{1j} ; j. maddeyi doğru yanıtlayan bireylerin sayısı,

M_{0j} ; j. maddeyi yanlış yanıtlayan bireylerin sayısı,

T_j ; Toplam bireylerin sayısıdır.

Mantel-Haenszel yöntemiyle elde edilen a_{MH} değerinin doğal logaritması alındıktan sonra “-2,35” ile çarpılarak delta katsayısı (Δ_{MH}) hesaplanır [$\Delta_{MH} = -2,35 * \ln(a_{MH})$] (Holland ve Thayer, 1988). Elde edilen Δ_{MH} katsayısına göre maddenin odak ya da referans grubuna göre DMF’li olup olmadığı tespit edilir (Millsap, 2011). Δ_{MH} katsayısının 0’dan küçük olması referans gruptaki bireylerin maddeyi doğru yanıtlama eğilimlerinin odak gruptakilere göre daha fazla olduğunu göstermektedir. Δ_{MH} katsayısının 0’a eşit olması her iki grupta yer alan bireylerin madde yanıtlama performanslarının aynı olduğu anlamına gelmektedir. Δ_{MH} katsayısının 0’dan büyük olması ise odak gruptaki bireylerin maddeyi doğru yanıtlama eğilimlerinin referans gruptakilere göre daha fazla olduğunu göstermektedir (De Ayala, 2009). Education Testing Service (ETS) tarafından önerilen ve Δ_{MH} katsayısının değerlendirilmesinde kullanılan ölçütler Tablo 3’te sunulmuştur (Wiberg, 2007, s. 8-9).

Tablo 3.

Mantel-Haenszel Yönteminin Değerlendirilmesinde ETS Tarafından Önerilen Ölçütler

DMF Düzeyi	Ölçütler	Açıklama
A	$ \Delta_{MH} < 1$	Düşük veya ihmal edilebilir düzey DMF
B	$1 \leq \Delta_{MH} < 1,5$	Orta düzey DMF
C	$ \Delta_{MH} \geq 1,5$	Yüksek düzey DMF

2.1.4.2. SIBTEST Yöntemi

SIBTEST yöntemi tek biçimli ve tek biçimli olmayan DMF belirlemede kullanılan ve Shealy ve Stout (1993) tarafından geliştirilen istatistiksel bir yöntemdir (French & Finch, 2015). Bu yöntemde DMF analizi yapılacak olan madde(ler) birinci gruba, diğer madde(ler) de ikinci gruba olmak üzere testte yer alan maddeler iki gruba ayrılırlar. Gözlenen puanlara dayalı olarak kestirilen gerçek puanlar grup karşılaştırması için eşleştirme değişkeni olarak kullanılırlar (Gierl, 2005). Diğer bir ifadeyle gruplar ölçülmek istenen örtük özelliğe göre eşleştirilir ve maddeler üzerinde performans farklılıklarının olup olmadığı incelenir. Bu eşleştirme puanının kestirimiyle, I. Tip hatanın kontrol edilmesinde yararlı olduğu gösterilen regresyon temelli bir düzeltmeyi içerir. Bu düzeltmeyle odak ve referans grubun yetenek dağılımlarındaki farklılıklar için maddeye ilişkin düzeltilmiş ortalama elde edilir (Roussos & Stout, 1996; Shealy ve Stout'dan aktaran Clauser & Mazor, 1998).

SIBTEST yöntemi nominal düzeyde I. Tip hatalar üretir, makul güce sahiptir ve nispeten küçük örneklerde de iyi performans göstermektedir. Tek biçimli DMF'nin tanımlanmasında Mantel-Haenszel istatistiğine benzer şekilde çalıştığı gösterilmiştir (Narayanan & Swaminathan, 1994; Roussos & Stout, 1996). Mantel-Haenszel yöntemine benzer performans sergilemesine rağmen SIBTEST yöntemi ayırt ediciliği yüksek olan maddelerin analizinde I. Tip hata oranı kontrolünde daha iyi sonuçlar verir (Roussos & Stout, 1996). Ayrıca SIBTEST yöntemi daha küçük örneklerde etkili olabilirken, odak ve referans grupların örneklem büyüklükleri çok farklı olduğunda Mantel-Haenszel yöntemi daha iyi performans gösterir.

SIBTEST yöntemi grup performanslarının eşit olup olmadığına karar vermede her bir maddenin marjinal yanıt fonksiyonunu karşılaştırır. Eğer eşit değilse maddede DMF olduğu kararına varılır. Her bir yetenek düzeyi için bir indeks değeri hesaplanır ve ardından toplam bir indeks değeri olan $\hat{\beta}$ istatistiği elde edilir (Cappaert, 2014, s. 42). Referans ve odak grup

arasında yetenek farklılığı olmadığında ise kestirilen β 'nın yansız bir tahmini olduğu söylenebilir (Awuor, 2008).

SIBTEST yöntemiyle DMF analizi yapılırken test edilen hipotez ve hipotezi test etmede kullanılan oran Eşitlik 5 ve 6'da verilmiştir (Shealy ve Stout'dan aktaran French & Finch, 2015).

$$H_0: \beta_{UNI} = 0 \text{ ve } H_0: \beta_{UNI} \neq 0 \quad (5)$$

$$SIB = \frac{\beta_{UNI}}{\sigma(\beta_{UNI})} \quad (6)$$

İki kategoride puanlanan maddeler için Shealy ve Stout (1993) tarafından tanımlanan indeks ise Eşitlik 7'de verilmiştir.

$$\beta_{UNI} = \int B(\theta) f_F(\theta) d(\theta) \quad (7)$$

Eşitlik 7'ye göre;

β_{UNI} ; belirli bir madde için DMF miktarını belirleyen parametre,

$B(\theta)$; [$B(\theta) = P(\theta, R) - P(\theta, O)$], aynı θ yetenek düzeyine sahip referans ve odak gruptaki bireylerin maddeyi doğru yanıtlama olasılıkları arasındaki fark,

$f_F(\theta)$; odak grubunun θ yetenek düzeyindeki yoğunluk fonksiyonu,

$d(\theta)$; theta'nın farkıdır.

Belirli bir test maddesi için referans ve odak grupları arasındaki ağırlıklı ortalama farkını test etmede ise Eşitlik 8'den yararlanılır (Shealy & Stout, 1993).

$$\hat{\beta}_{UNI} = \sum_{k=0}^k p_k d_k \quad (8)$$

Eşitlik 8'e göre,

p_k : odak gruptaki bireylerin oranı,

d_k ise ($d_k = p_{Rk} - p_{Ok}$) referans ve odak grubun düzeltilmiş ortalamaları arasındaki farktır.

Referans ve odak grubun yetenek dağılımları farklı olduğunda kestirilen β değeri yanlı olacaktır. β kestirimini daha doğru yapabilmek için SIBTEST yöntemi regresyona dayalı bir düzeltme kullanır (Shealy & Stout, 1993). Regresyona dayalı bu düzeltmeyi uygulamak için referans ve odak grupların puan ortalamaları, varyans ve güvenilirliklerine ilişkin tahminler, gözlenen test puanları kullanılarak hesaplanır (Gierl, Gotzmann & Boughton, 2004).

SIBTEST yöntemiyle elde edilen sonuçların değerlendirilmesinde Roussos ve Stout (1996) tarafından önerilen ölçütler Tablo 4’te sunulmuştur.

Tablo 4.

SIBTEST Yönteminin Değerlendirilmesinde Önerilen Ölçütler

DMF Düzeyi	Ölçütler	Açıklama
A	$\beta_U < 0,059$	Düşük veya ihmal edilebilir düzey DMF
B	$0,059 \leq \beta_U < 0,088$	Orta düzey DMF
C	$\beta_U \geq 0,088$	Yüksek düzey DMF

2.1.4.3. Lojistik Regresyon Yöntemi

Lojistik Regresyon yöntemi ilk olarak Swaminathan ve Rogers (1990) tarafından DMF belirlemek için tanıtılmıştır. Bağımsız değişkenin sürekli, bağımlı değişkenin ise süreksiz olduğu durumlarda kullanılan çok değişkenli istatistiksel bir modeldir. Bu yöntemle hem tek biçimli hem de tek biçimli olmayan DMF belirlenebilmektedir (Swaminathan & Rogers, 1990). Ayrıca tek biçimli ve tek biçimli olmayan DMF belirlemede Mantel-Haenszel ve SIBTEST yöntemlerine göre daha güçlü sonuçlar vermektedir (Li & Stout, 1996; Rogers & Swaminathan, 1993; Swaminathan & Rogers, 1990). Fakat DMF’li maddeleri belirlemede ilişkili bir etki büyüklüğü ölçüsü (DMF’nin büyüklüğünü ölçen tanımlayıcı bir istatistik) olmadan istatistiksel bir test kullanılması yöntemin dezavantajıdır (Jodoin & Gierl, 2001).

Lojistik Regresyon yöntemiyle yapılan DMF analizinde eşleştirme değişkeni olarak toplam test puanı kullanılır. Belirli bir bireyin her bir maddeye doğru cevap verme olasılıklarını tahmin etmek için toplam puan, grup üyelikleri ve etkileşim katsayısı regresyon analizinde kullanılan değişkenlerdir. Lojistik Regresyon modelinde kullanılan bağımlı değişken maddeden alınan puandır. Bağımsız değişkenler ise eşleştirme değişkeni olarak bireyin gözlenen yeteneği (toplam test puanı), gruplama değişkeni ile grup ve yetenek arasındaki etkileşimdir (toplam puan x grup). Analizde dört adet parametre kestirilmektedir. Bunlar: β_0 (kesişim), β_1 (regresyon katsayısı), β_2 (madde performansındaki grup farklılığı) ve β_3 (grup ve yetenek arasındaki etkileşim katsayısı)’tür (Zumbo, 1999).

Kavramsal olarak DMF analizi için kullanılan Lojistik Regresyon yöntemi bir dizi iç içe geçmiş modelin karşılaştırmasına dayanmaktadır. DMF’nin istatistiksel anlamlılığının test edilmesinde bağımsız değişkenler modele sırayla eklenir ve sonrasında χ^2 testi yapılarak her bir modelin uyumu istatistiksel olarak değerlendirilir (Zumbo, 1999). Tek biçimli DMF

olması halinde $\beta_2 \neq 0$ ve $\beta_3 = 0$ iken, tek biçimli olmayan DMF olması halinde ise $\beta_3 \neq 0$ olur. Ayrıca tek biçimli ve tek biçimli olmayan DMF'nin belirlenmesinde serbestlik derecesi iki olan $\chi^2_{sd:2}$ testi kullanılır (Swaminathan & Rogers, 1990).

DMF analizinde kullanılan Lojistik Regresyon modeli Eşitlik 9'da yer almaktadır (Zumbo, 1999).

$$Y = \beta_0 + \beta_1 \text{tot} + \beta_2 \text{grup} + \beta_3 \text{tot} \times \text{grup} \quad (9)$$

Y odds oranının doğal logaritmasıdır ve maddeyi doğru cevaplamanın log-odds'una (ya da logit) ilişkin modelleme ise Eşitlik 10'da verilmiştir.

$$\ln \left[\frac{p_i}{(1-p_i)} \right] = \beta_0 + \beta_1 x + \beta_2 g + \beta_3 xg \quad (10)$$

Eşitlik 10'da yer alan “x” bireyin toplam test puanını, “g” grup üyeliğini ve “xg” ise toplam test puanı ve grup üyeliği etkileşimini göstermektedir (Swaminathan & Rogers, 1990).

Lojistik Regresyon yönteminin değerlendirilmesinde Zumbo ve Thomas (1996) ile Jodoin ve Gierl (2001) tarafından önerilen ölçütler Tablo 5'te sunulmuştur.

Tablo 5.

Lojistik Regresyon Yönteminin Değerlendirilmesinde Önerilen Ölçütler

Zumbo ve Thomas (1996) tarafından önerilen ölçütler		
DMF Düzeyi	Ölçütler	Açıklama
A	$\Delta R^2 < 0,13$	Düşük veya ihmal edilebilir düzey DMF
B	$0,13 \leq \Delta R^2 < 0,26$	Orta düzey DMF
C	$\Delta R^2 \geq 0,26$	Yüksek düzey DMF
Jodoin ve Gierl (2001) tarafından önerilen ölçütler		
DMF Düzeyi	Ölçütler	Açıklama
A	$\Delta R^2 < 0,035$	Düşük veya ihmal edilebilir düzey DMF
B	$0,035 \leq \Delta R^2 < 0,070$	Orta düzey DMF
C	$\Delta R^2 \geq 0,070$	Yüksek düzey DMF

2.1.4.4. Lord'un χ^2 Testi Yöntemi

Lord'un χ^2 yönteminde (Lord, 1980) odak ve referans grubun madde parametreleri arasındaki farkın istatistiksel olarak anlamlı olup olmadığı test edilir (Magis vd., 2010).

Yöntem hem tek biçimli hem de tek biçimli olmayan DMF belirlemede kullanılır. Lord'un χ^2 yöntemi güçlük ve ayırt edicilik parametrelerinin varyans-kovaryans matrisindeki farklılıkların incelenmesine dayanır ve grupların madde karakteristik eğrileri arasında kalan alan hesaplanır (Hambleton vd., 1991).

Bu yöntem χ^2 dağılımına sahip bir istatistiğe dayanır (Magis vd., 2010). Ki-kare dağılımı 2PL modelde iki, 3PL modelde ise üç serbestlik derecesinde χ^2 dağılımına yakın olmalıdır (Lord, 1980). Lord'un χ^2 yönteminde 2PL ve 3PL modelleri 1PL'ye göre daha kullanışlıdır. Çünkü gruplar arasındaki farkın sadece 1PL model kullanılarak incelenmesi DMF'nin yanlış ve yetersiz şekilde hesaplanmasına neden olabilir (Kim, 2010). Her türlü madde tepki modeli ile hesaplanabilen bu yöntemde, istatistiksel testlerden önce madde parametreleri ortak bir metrik ile ölçeklendirilmelidir. Bu yöntemde kullanılan Q_j istatistiği Eşitlik 11'de verilmiştir (Magis vd., 2010).

$$Q_j = (v_{jR} - v_{jO})'(\Sigma_{jR} - \Sigma_{jO})^{-1}(v_{jR} - v_{jO}) \quad (11)$$

Eşitlik 11'e göre;

v_{jR} ; referans grup için j. maddenin madde ayırt edicilik, güçlük ve şans parametre vektörleri
 $\{v_{jR} = (a_{jR}, b_{jR}, c_{jR})\}$,

v_{jO} ; odak grup için j. maddenin madde ayırt edicilik, güçlük ve şans parametre vektörleri
 $\{v_{jO} = (a_{jO}, b_{jO}, c_{jO})\}$,

Σ_{jR} ; referans grubun varyans-kovaryans matrisi,

Σ_{jO} ; odak grubun varyans-kovaryans matrisidir.

Lord'un χ^2 yönteminin değerlendirmesinde Mantel-Haenszel yönteminde elde edilen Δ_{MH} katsayısının değerlendirilmesi için önerilen ölçütler dikkate alınır ve bu ölçütler Tablo 6'da sunulmuştur (Holland & Thayer, 1985).

Tablo 6.

Lord'un χ^2 Yönteminin Değerlendirilmesinde Önerilen Ölçütler

DMF Düzeyi	Ölçütler	Açıklama
A	$\Delta_{Lord} < 1$	Düşük veya ihmal edilebilir düzey DMF
B	$1 \leq \Delta_{Lord} < 1,5$	Orta düzey DMF
C	$\Delta_{Lord} \geq 1,5$	Yüksek düzey DMF

2.1.4.5. Raju'nun Alan Ölçümü Yöntemi

Raju'nun alan ölçümü yönteminde (Raju, 1988, 1990) odak ve referans grubun madde karakteristik eğrileri arasında kalan alan hesaplanır. Karşılık gelen Z istatistiği ise gerçek alanın sıfır olduğu varsayımına dayanır (Magis vd., 2010). Madde karakteristik eğrileri arasında kalan alanın sıfır olması maddede yanlılığın olmadığı göstergesidir. Alan sıfırdan uzaklaştıkça madde için yanlılıkta artmaktadır (Lord, 1980 ve Raju, 1988'den aktaran Ateşok Deveci, 2008). 1PL modelde bir maddenin karakteristik eğrileri arasında kalan alan (odak ve referans grubunda) madde güçlüğü kestirimlerindeki farktan kaynaklanmaktadır. Eşitlik 12'de Z istatistiği verilmiştir (Raju, 1988'den aktaran Magis vd., 2010).

$$Z = \frac{b_{jR} - b_{jO}}{\sqrt{(\hat{\sigma}_{jR}^2 + \hat{\sigma}_{jO}^2)}} \quad (12)$$

Eşitlik 12'ye göre;

b_{jR} ; referans grup için j. maddenin madde güçlük parametresi,

b_{jO} ; odak grup için j. maddenin madde güçlük parametresi,

$\hat{\sigma}_{jR}^2$; referans grup için j. maddenin varyansı,

$\hat{\sigma}_{jO}^2$; odak grup için j. maddenin varyansıdır.

Bu yöntemde DMF düzeyinin sınıflandırılmasında Mantel-Haenszel yönteminin analizinde elde edilen Δ_{MH} değeri dikkate alınır. Δ_{MH} değerinin “k= -15” sabitine oranı ile β_u değeri elde edilir (Shealy & Stout, 1993). Elde edilen β_u değerinin karesi alınarak NCDIF katsayısı hesaplanır. NCDIF, iki farklı grubun bir test maddesine doğru yanıt verme olasılığının farkıdır ve Eşitlik 13 ve 14'te verilmiştir (Wright & Oshima, 2015).

$$d_i(\theta_s) = P_{iO}(\theta_s) - P_{iR}(\theta_s) \quad (13)$$

$$NCDIF_i = E_O(d_i(\theta_s)^2) \quad (14)$$

Eşitlik 13 ve 14'e göre;

$P_{iO}(\theta_s)$; theta (θ) yetenek düzeyinde odak grubun i. maddeyi doğru yanıtlama olasılığı,

$P_{iR}(\theta_s)$; theta (θ) yetenek düzeyinde referans grubun i. maddeyi doğru yanıtlama olasılığıdır.

NCDIF diğer DMF istatistiklerine benzer şekilde çalışır. Bu değer hesaplanırken madde karakteristik fonksiyonları arasındaki farkın karesi hem tek biçimli hem de tek biçimli olmayan DMF'nin tespit edilmesini sağlar. NCDIF'in değerlendirilmesinde Wright ve Oshima (2015) tarafından önerilen ölçütler Tablo 7'de sunulmuştur.

Tablo 7.

NCDIF'in Değerlendirilmesinde Önerilen Ölçütler

DMF Düzeyi	Ölçütler	Açıklama
A	$NCDIF < 0,003$	Düşük veya ihmal edilebilir düzey DMF
B	$0,003 \leq NCDIF < 0,008$	Orta düzey DMF
C	$NCDIF \geq 0,008$	Yüksek düzey DMF

2.2. Madde Tepki Kuramı

Örtük Özellik Teorisi (Latent Trait Theory) olarak bilinen Madde Tepki Kuramı (MTK), Klasik Test Kuramının (KTK) bazı sınırlılıklarına alternatif olarak 20. yy ortalarında ortaya çıkmıştır (Crocker & Algina, 1986; Embretson & Reise, 2000). Test eşitleme, bilgisayarda ölçme uygulamaları, test yapılandırma, madde ve seçenekleri ağırlıklandırma veya değişen madde fonksiyonunu belirlemede MTK'dan yararlanılmaktadır (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985).

Eğitim ve psikolojide yapılan ölçme çalışmalarının çoğunda ölçme işlemine konu olan birçok değişken bulunmaktadır. Bu değişkenlerin bazıları boy uzunluğu veya ağırlık gibi fiziksel özellikler olmayan kavramsal yapılardır. Doğrudan ölçülmesi mümkün olmayan bu değişkenler araştırmacılar tarafından örtük veya gözlenemeyen değişken olarak tanımlanır. Bu ölçme çalışmalarının amacı ise bireylerin bu tür örtük değişkenlere ne derece sahip olduklarını belirlemektir. MTK'da bu örtük değişkenler genellikle “yetenek” terimi ile tanımlanır (Baker, 2001). Bu kuramda bireyin yetenek düzeyi maddeye verilen tepkilere ve maddelerin özelliğine bağlıdır. Dolayısıyla MTK, yetenek düzeyi ile maddeye verilen tepkiler arasındaki ilişkinin matematiksel bir fonksiyonla açıklandığı bir olasılık modeli önermektedir. Bu sayede testle ölçülmek istenen yetenek düzeyi maddelere verilen tepkilerle kestirilmeye çalışılır. Ayrıca kuramda kullanılan modeller güçlü varsayımlara dayanmaktadır ve özellikle birçok yetenek testi MTK'ya göre geliştirilmiştir (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985).

MTK bireyin her bir maddeye verdiği cevabın doğruluğu ya da yanlışlığıyla ilgilenirken, KTK’da ise bireyin ham puanı testin her bir maddesinden alınan puanların toplanmasıyla elde edilir (Baker, 2001). Test ve madde istatistikleri de toplam puanlar üzerinden hesaplanır ve toplam puanlar testin güçlük düzeyine göre değişkenlik gösterir (Lord & Novick, 1968). Dolayısıyla bireyler zor testlerde daha düşük, kolay testlerde ise daha yüksek puana sahip olurlar. Fakat MTK’ya göre bireylerin yetenekleri değişmemektedir (Hambleton & Jones, 1993). Daha açık ifadeyle gözlenen puanlar teste bağımlı iken yetenek puanları testten bağımsızdır ve bu nedenle yetenek puanlarının daha temel olduğu vurgulanmıştır (Lord’dan aktaran Hambleton & Jones, 1993).

MTK’nın önemli özelliklerinden biri ise değişmezlik özelliğidir. Bu özellikte madde parametrelerinin kestirimi yanıtlayıcı örneklemeden, yetenek parametre kestirimi ise madde örneklemeden bağımsızdır. Dolayısıyla madde parametreleri maddeyi tanımlayan gruba bağlı değil maddeye ilişkin bir özelliktir. Bu sayede madde parametreleri ile bireyin karakteristik özellikleri birbirinden ayrılabilir. Model veri uyumunun sağlanmasıyla yetenek ve madde parametreleri birbirinden bağımsız bir şekilde kestirilir aksi halde değişmezlik özelliği sağlanamaz. MTK’nın değişmezlik özelliğinden dolayı örneklemin teorik olarak ilgili evrenin tesadüfi örnekleme olmasına gerek yoktur. Böylece tesadüfi olmayan örneklem kullanılabilir ve bu örneklemden kestirilen parametreler evren metriğine doğrusal dönüşümler yapılarak yerleştirilebilir. Bu özellik kuramın diğer kuramlara tercih edilmesinde de önemli rol oynamaktadır (Baker, 2001; DeMars, 2010; Hambleton vd., 1991).

2.2.1. Madde Tepki Kuramının Varsayımları

Tek boyutlu MTK’da kullanılan modellerin temel varsayımları tek boyutluluk ve yerel bağımsızlıktır.

Tek boyutluluk varsayımı, test ile tek bir örtük özelliğin ölçülmesi ve elde edilen puanların sadece o örtük özelliğe bağlı olmasıdır. Bir test için sadece tek bir puan raporlandığında maddelerin ortak bir yapıyı paylaştığına ilişkin üstü kapalı bir varsayım vardır (DeMars, 2010). Tek boyutluluk, modelde her bir birey için tek bir theta (yetenek düzeyi) olduğu anlamına gelir. Fakat test uygulamalarında karşılaşılan bazı durumlar (bilişsel veya kişisel faktörler gibi) bireylerin test performansını etkilediğinden dolayı bu varsayımın karşılanması zorlaşabilir. Tek boyutluluk varsayımının ihlali ise parametrelerin ve standart

hatanın yanlış kestirimine yol açabildiği için test puanlarının güvenilirlik ve geçerliğini tehlikeye atar (DeMars, 2010; Embretson & Reise, 2000; Hambleton & Swaminathan, 1985; Hambleton vd., 1991).

Yerel bağımsızlık varsayımı, belirli bir yetenek düzeyindeki bireylerin belirli bir maddeyi cevaplama olasılıklarının diğer maddelerden etkilenmemesidir. Yani aynı theta (yetenek düzeyi) değerine sahip bireylerden oluşan bir yanıtlayıcı grubunun maddeleri doğru yanıtlama olasılıklarının birbirinden bağımsız olması durumudur (Lord & Novick, 1968). Hambleton vd. (1991)'e göre ise yerel bağımsızlık, bireylerin herhangi bir madde çiftine verdiği yanıtların istatistiksel olarak birbirinden bağımsız olmasıdır. Hambleton ve Swaminathan (1985) maddelere yanıt verme davranışı altında yatan baskın yeteneğin başka bir faktörden olumlu ya da olumsuz etkilenmemesi gerektiğini ve bu durumda tek boyutluluk varsayımının sağlanmasıyla yerel bağımsızlık varsayımının da karşılanacağını belirtmişlerdir. Crocker ve Algina (1986)'ya göre ise belirli bir yetenek düzeyindeki bireylerin madde puanları arasındaki madde-madde korelasyonları incelenerek yerel bağımsızlık varsayımı test edilebilir. Eğer tek boyutlu bir modelde madde cevapları yerel bağımsız değil ise diğer boyut bağımlılığa neden olmaktadır. Bu bağımlılık maddelerin büyük bir kısmını etkilemediği sürece ayrı boyutlar olarak ortaya çıkmayabilir ve tek boyutluluk testi ile de belirlenemeyebilir (DeMars, 2010).

2.2.2. Madde Tepki Kuramı Modelleri

MTK'da kullanılan modeller bir ölçme aracıyla ölçülen yetenek düzeyi ile maddeyi doğru yanıtlama olasılığı arasındaki ilişkiye dair matematiksel bir eşitlik sunmaktadır. Her bir modelde madde karakteristik eğrisini tanımlayan bir ya da daha fazla parametre bulunur. Bu modeller verinin parametrik ya da parametrik olmamasına, ölçülen yeteneğin boyut sayısına ve tepkilerin iki ya da ikiden fazla kategoride puanlanmasına göre farklılaşmaktadır (Baker, 2001; DeMars, 2010; Embretson & Reise, 2000).

İki kategorili madde tepki kuramı modelleri maddelerin doğru/yanlış (1/0) şeklinde ikili puanlandığı durumlarda kullanılmaktadır. Bireyin maddeye verdiği yanıtlar başarı testlerinde doğru/yanlış, tutum ölçeklerinde katılıyorum/katılmıyorum veya anketlerde evet/hayır şeklinde puanlanmaktadır. İki kategoride puanlanan maddeler için genel olarak kullanılan modeller; bir parametrelili lojistik (1PL), iki parametrelili lojistik (2PL) ve üç

parametrelili lojistik (3PL) modellerdir. Bu modeller madde parametrelerinin sayılarına göre adlandırılırlar (DeMars, 2010).

Bir parametrelili lojistik modelde (1PLM) bütün maddelerin ayırt edicilik parametrelerinin eşit olduğu ve şans başarısının olmadığı varsayılmaktadır. Modelde sadece madde güçlük parametresi kestirilmektedir. 1PL model için bir yanıtlayıcının i maddesini yanıtlama olasılığı Eşitlik 15’te verilmiştir (Hambleton & Swaminathan, 1985, s. 49).

$$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}} \quad (15)$$

Eşitlik 15’e göre;

$P_i(\theta)$; θ yetenek düzeyinde rastgele seçilen bir yanıtlayıcının i maddesine cevap verme olasılığı,

b_i ; i . maddenin güçlük parametre değeri,

e ; 2,718 sabiti ve

D ; ölçek sabiti (1,7)’dir.

İki parametrelili lojistik modelde madde güçlük parametresinin yanı sıra madde ayırt edicilik parametresi de kestirilmektedir. 2PL modelde kestirilen güçlük parametresi 1PL modelle kestirilen güçlük parametresinden farklıdır. Çünkü 1PL modelde ayırt edicilik parametresi bütün maddeler için eşit (sabit değer) olduğu kabul edilmekte iken 2PL modelde her madde için ayırt edicilik parametresi hesaplanmaktadır. 2PL model için bir yanıtlayıcının i maddesini yanıtlama olasılığı Eşitlik 16’da verilmiştir (Hambleton & Swaminathan, 1985, s. 36).

$$P_i(\theta) = \frac{e^{D_{ai}(\theta-b_i)}}{1 + e^{D_{ai}(\theta-b_i)}} \quad (16)$$

Eşitlik 16’ya göre;

$P_i(\theta)$; θ yetenek düzeyinde rastgele seçilen bir yanıtlayıcının i maddesine cevap verme olasılığı,

a_i ; i . maddenin ayırt edicilik parametre değeri,

b_i ; i . maddenin güçlük parametre değeri,

e ; 2,718 sabiti ve

D; ölçek sabiti (1,7)'dir.

3PL modelde ise güçlük ve ayırt edicilik parametrelerine ek olarak şans parametresi de kestirilmektedir. Bu sayede alt yetenek seviyesindeki bireylerin maddeyi şans başarısıyla doğru cevaplayabilme olasılıkları da hesaplanmış olur. 3PL model için bir yanıtlayıcının i maddesini yanıtlama olasılığı Eşitlik 17'de verilmiştir (Hambleton & Swaminathan, 1985, s. 37).

$$P_i(\theta) = c_i + (1-c_i) \frac{e^{D_{ai}(\theta-b_i)}}{1 + e^{D_{ai}(\theta-b_i)}} \quad (17)$$

Eşitlik 17'ye göre;

$P_i(\theta)$; θ yetenek düzeyinde rastgele seçilen bir yanıtlayıcının i maddesine cevap verme olasılığı,

a_i ; i. maddenin ayırt edicilik parametre değeri,

b_i ; i. maddenin güçlük parametre değeri,

c_i ; i. maddenin şans başarısı parametre değeri,

e; 2,718 sabiti ve

D; ölçek sabiti (1,7)'dir.

Modellerde yer alan ayırt edicilik parametresi (a_i) kuramsal olarak “ $-\infty$ ve $+\infty$ ” arasında değer almasına rağmen uygulamada “0 ve 2” arasındadır. Bu parametreyle maddenin alt ve üst yetenek seviyesine sahip olan bireylerin birbirinden ne derece ayırt edebildiği ile ilgili bilgi sağlanmış olur. Güçlük parametresi (b_i) genelde “-2 ve +2” arasında değer almaktadır. Parametre değeri -2'ye yakın olan maddeler çok kolay iken +2'ye yakın olanlar çok zor maddelerdir. Şans başarısı parametresi (c_i) ise düşük yetenek düzeyindeki bireylerin i. maddeye doğru cevap verme olasılıklarıdır (DeMars, 2010; Embretson & Reise, 2000; Hambleton & Swaminathan, 1985).

2.3. İlgili Araştırmalar

Golia (2014) DMF'li maddelerin testten çıkarılmasının yetenek kestirimlerini nasıl etkilediğine yönelik çalışma yapmıştır. Çalışmanın simülasyon aşamasında DMF'li madde sayısı, DMF büyüklüğü ve DMF türü farklı olacak şekilde ikiden fazla kategoride puanlanan maddelere yönelik veri üretmiştir. Ayrıca örneklem büyüklüğünü (1000) ve testin uzunluğunu (15) sabit koşul olarak ele almıştır. Çalışma sonucunda sadece DMF'nin türünün değil aynı zamanda DMF'li madde sayısının da yetenek kestirimlerini etkilediğini tespit etmiştir. Bu durumun da DMF'li madde sayısının odak grubun yanıtlarını etkilemesinden kaynaklandığını belirtmiştir. DMF'li maddelerin türünün, büyüklüğünün ve sayısının yetenek kestiriminin doğruluğunu etkilediği ve anketin çok kısa olmadığı durumlarda, bir veya daha fazla DMF'li maddelerin çıkarılmasıyla DMF'li maddelerin yanlılık (BIAS) etkisini ihmal edilebilir hale getirdiği sonucuna ulaşmıştır. Ayrıca Golia (2014) DMF'li maddelerin varlığının yetenek kestirimleri üzerinde bir etkisinin olmadığı durumlarda bu maddeleri ankette kullanmanın uygun olacağını belirtmiştir.

Lee ve Zhang (2017) DMF'nin yetenek kestirimleri ve güvenilirlik üzerindeki etkisini incelemek için simülasyon çalışması yapmışlardır. Çalışmada test uzunluğu (30 ve 60), testteki DMF'li madde oranı (%5, %10 ve %20), örneklem büyüklüğü ($N_R/N_0=500/250$, $1000/500$ ve $1500/1500$), grupların yetenek düzey farklılığı ve madde türü olmak üzere beş faktör manipüle edilmiştir. Çalışmada 72 simülasyon koşulu oluşturulmuş ve 100 replikasyon gerçekleştirilmiştir. Çalışma sonucunda 2PL model için DMF'nin toplam puanlar üzerindeki etkisinin 3PL modele göre daha önemli olduğu sonucuna ulaşmışlardır. Yetenek kestirimlerine ilişkin standart hatanın yetenek aralığında eşit olmadığını ve belirli yetenek değerleri için etki büyüklüğünün belirgin olduğunu tespit etmişlerdir. Ayrıca genel olarak DMF'nin yetenek kestirimi ve güvenilirlik tahmini üzerindeki etkilerinin standart sapmasının büyük örneklerde daha az olduğunu belirtmişlerdir. Bunlara ek olarak DMF'nin gerçek puan varyansının gözlemlenen puan varyansına oranı üzerinde sınırlı etkisinin olduğunu, fakat yetenek parametresinin kestirim hatası üzerinde belirli etkilerinin olduğu sonucuna ulaşmışlardır.

Cappaert (2014) tek biçimli ve tek biçimli olmayan DMF ve DBF'nin çeşitli koşullar altında yetenek kestirimi üzerindeki etkilerini araştırmak için simülasyon çalışması yapmıştır. Simülasyon koşullarını demet (bundle) büyüklüğü, bir demet içindeki tek biçimli DMF'li madde sayısı, bir demetteki tek biçimli olmayan DMF'li madde sayısı, referans ve odak

grubun sayısı ve test uzunluğu oluşturmaktadır. Çalışma sonucunda tek biçimli DMF ve DBF'deki artışın referans grubun yetenek parametresinin yanlış kestirimine neden olduğunu belirtmiştir. Tek biçimli olmayan DMF ve DBF büyüklüğünün ise yetenek kestirimleri ile kestirimlerin RMSE ve standart hatasını etkilediğini saptamıştır. Gerçek ve kestirilen yetenek parametrelerine ilişkin korelasyon değerleri ise demetteki tek biçimli ve tek biçimli olmayan DMF ve DBF'nin büyüklüğüne bakılmaksızın tutarlı sonuçlar verdiğini belirtmiştir. Crossing SIBTEST yönteminin ise tek biçimli DBF için kabul edilebilir düzeyde I. Tip hata ve güç oranları sağladığını tespit etmiştir.

Ö. Sünbül ve Sünbül (2016) KTK'ya dayalı Mantel-Haenszel ve Lojistik Regresyon ile MTK'ya dayalı Lord'un χ^2 ve Raju'nun işaretli alan ölçüleri yöntemlerinin belirli koşullar altında I. Tip hata ve güç oranlarını belirlemek amacıyla çalışma yapmışlardır. Çalışmada örneklem büyüklüğü (500, 1000, 1500, 2000 ve 3000), madde sayısı (40 ve 60), referans grubun tüm örnekleme oranı (0,50 ve 0,75), testteki DMF'li madde oranı (%5 ve %10) ve DMF büyüklüğü (0,50 ve 0,75) koşulları altında simülatif veri üretmişlerdir. Araştırma sonucunda örneklem büyüklüğü, madde sayısı ve örneklem büyüklüğü oranı koşullarında KTK'ya dayalı yöntemlerin (Mantel-Haenszel ve Lojistik Regresyon) I. Tip hata oran ortalamalarında önemli değişiklik olmadığını gözlemlemişlerdir. MTK'ya dayalı yöntemlerde (Lord'un χ^2 ve Raju'nun işaretli alan ölçüleri) ise örneklem büyüklüğü arttıkça I. Tip hata oranlarının azaldığını; madde sayısının artmasının I. Tip hata oranların ortalamasını değiştirmediklerini ve grupların örneklem büyüklüklerinin eşit olduğu durumda I. Tip hata ortalamasının daha düşük olduğunu saptamışlardır. Kullanılan dört yöntemde de örneklem büyüklüğünün ve DMF sayısının artmasıyla güç oranlarının ortalamasının arttığını belirlemişlerdir. Tüm yöntemlerde madde sayısındaki farklılaşmanın güç oranlarının ortalamasını çok fazla değiştirmediklerini, DMF'li madde sayısının artmasının ve örneklem büyüklüğü oranlarının da farklı olmasının güç oranlarını azalttığını tespit etmişlerdir.

Tunç (2016) doktora tez çalışmasında iki ve çok kategorili puanlanan maddelerde DMF yöntemlerinin karşılaştırmasını yapmıştır. Bu kapsamda simülasyon veri seti oluşturmuş ve I. Tip hata ve güç oranlarını incelemiştir. Simülasyon koşullarında DMF biçimi (tek biçimli DMF) ve test uzunluğu (20) sabit koşullar olarak ele alırken; örneklem büyüklüğü (600, 1200 ve 2400), örneklem büyüklüğü oranları (1/1 ve 1/2), DMF'li madde oranı (%10, %30 ve %50) ve DMF büyüklüğü (0,25, 0,50, 1 ve 1,5) koşullarını manipüle etmiştir. İki kategorili veriler için Mantel-Haenszel ve Lordif; çok kategorili veriler için ise Lordif ve

ANOVA yöntemlerini kullanmıştır. Araştırma sonucunda örneklem büyüklüğü arttıkça I. Tip hata oranlarının arttığını fakat yöntemlere göre I. Tip hata oranlarında manidar farklılık olmadığını ($p>,05$) bulmuştur. Örneklem büyüklüğünün artmasıyla güç oranlarının da bu doğrultuda arttığını ve ANOVA yöntemi dışındaki yöntemlerde manidar bir farklılık olmadığını tespit etmiştir. Örneklem büyüklüğü oranlarına göre I. Tip hata ve güç oranlarında manidar farklılık bulmamıştır. DMF’li madde sayısı arttıkça I. Tip hata oranlarının arttığını, güç oranının azaldığını ve I. Tip hata ile güç oranlarının DMF’li madde oranlarına göre manidar farklılık gösterdiğini belirlemiştir. DMF’li madde sayısının %10 olduğu durumda en düşük I. Tip hata ve en yüksek güç oranının gerçekleştiğini saptamıştır. DMF büyüklüğü 0,25’ten 1,5’e yükseldikçe I. Tip hata ve güç oranlarının arttığını ve DMF büyüklüğüne göre her ikisinin de manidar farklılık gösterdiğini bulmuştur.

Atalay Kabasakal vd. (2014) belirli koşullar altında SIBTEST, Mantel-Haenszel ve MTK Olabilirlik Oran yöntemlerinin performansını (I. Tip hata ve güç oranları) karşılaştırmışlardır. Çalışmada örneklem büyüklüğü oranları, yetenek dağılımı farklılıkları, test uzunluğu, DMF’li madde sayısı ve model türü koşullarına göre simülatif veri üretmişlerdir. Araştırma sonuçlarına göre kullanılan üç DMF yönteminin de I. Tip hatalarının nominal hata düzeyinden daha yüksek ($,05$) olduğunu tespit etmişlerdir. SIBTEST, Mantel-Haenszel ve MTK Olabilirlik Oran yöntemlerinden I. Tip hata oranı en yüksek olan yöntemin SIBTEST; güç oranı en yüksek olan yöntemin ise Mantel-Haenszel olduğunu tespit etmişlerdir. Test uzunluğu koşuluna göre yapılan incelemede ise; SIBTEST yönteminin I. Tip hata oranının daha fazla etkilediğini bulmuşlardır. Yetenek dağılım farklılıkları koşulu altında SIBTEST yönteminin I. Tip hata oranının diğer yöntemlerden daha yüksek, güç oranının ise diğer yöntemlere göre daha düşük olduğunu belirlemişlerdir. Ayrıca odak ve referans grubun örneklem büyüklük farklılıklarından en fazla Mantel-Haenszel yönteminin etkilendiğini; eşit olmayan örneklem büyüklüklerinin eşit olan örneklem büyüklüklerine göre I. Tip hata oranının daha düşük, testin gücünün de görece düşük olduğunu bulmuşlardır.

Erdem Keklik (2012) çalışmasında iki kategorili verilerde DMF belirlemeye yönelik kullanılan Mantel-Haenszel, Lojistik Regresyon ve MTK Olabilirlik Oran yöntemlerinin I. Tip hata ve güç oranlarını incelemiştir. Çalışma kapsamında örneklem büyüklüğü (600, 1200 ve 2400), örneklem büyüklüğü oranı (1/1 ve 1/2) ve grupların yetenek dağılımı ($N(0, 1)$, $N(-0,5, 1)$ ve $N(-0,75, 1)$) koşullarını manipüle etmiştir. Test uzunluğu (20), DMF’li madde

sayısı (3) ve DMF türü (tek biçimli) koşullarını ise sabit tutarak toplam 18 simülasyon koşulu oluşturmuştur. Araştırma sonucunda referans ve odak grubun yetenek dağılımları normal dağılım gösterdiğinde Mantel-Haenszel ve Lojistik Regresyon yöntemlerinin benzer fakat MTK Olabilirlik Oran yöntemine kıyasla daha düşük I. Tip hata oranına sahip olduğunu belirlemiştir. Referans ve odak grubun yetenek dağılımları farklılaştığında ise MTK Olabilirlik Oran yönteminin Mantel-Haenszel ve Lojistik Regresyon yöntemlerine kıyasla I. Tip hata kontrolünü daha iyi sağladığını tespit etmiştir. Ayrıca ele alınan koşullar altında Mantel-Haenszel ve Lojistik Regresyon yöntemlerinde şişirilmiş I. Tip hataların ortaya çıktığını rapor etmiştir.

Kim (2010) çalışmasında Mantel Haenszel, Lojistik Regresyon, Test-DFIT ve Lord'un χ^2 yöntemlerinin I. Tip hata oranlarını incelemiştir. Çalışmada örneklem büyüklüğü, test uzunluğu, grupların ortalama ve standart sapmalarındaki farklılıklar olmak üzere toplam 36 farklı test koşulunu ele almıştır. I. Tip hata oranı kontrollerinde Bonferroni, Holm's ve B-H (Benjamini-Hochberg) olmak üzere üç farklı düzeltme yöntemlerini uygulayarak karşılaştırma yapmıştır. Elde edilen sonuçlar neticesinde uygulanan düzeltme tekniklerinin I. Tip hatanın kontrolünde etkili olduğunu tespit etmiştir. Ayrıca örneklem büyüklüğü oranlarının eşit ve büyük olmasının dört yöntemin hepsinde de I. Tip hata oranlarını şişirme eğiliminde olduğunu tespit etmiştir.

Awuor (2008) örneklem büyüklüğü (250, 500 ve 1000), örneklem büyüklüğü oranları (1,00, 0,50 ve 0,10) ve DMF büyüklüğü (0,35 ve 0,65) koşulları altında SIBTEST ve Mantel-Haenszel yöntemlerinin sonuçlarını karşılaştırmıştır. Elde edilen sonuçlara göre referans ve odak gruplar arasındaki örneklem büyüklüğü oranlarının SIBTEST ve Mantel-Haenszel yöntemlerinin gücü üzerinde önemli etkiye sahip olduğunu tespit etmiştir. Grupların örneklem büyüklüğü oranları 1/0,5 iken SIBTEST yöntemi Mantel-Haenszel'e göre daha yüksek I. Tip hata oranına sahip olmuştur. Sonuçlar sadece referans ve odak grupları oranlarının değil aynı zamanda örneklem büyüklüğü ve DMF düzeyinin de SIBTEST ve Mantel-Haenszel yöntemlerinin I. Tip hata oranlarını değiştirdiğini göstermiştir. Küçük örneklemelerde ve DMF büyüklüğü 0,35 olduğunda her iki yöntemde gücü önemli ölçüde düşerek II. Tip hataya neden olduğunu belirtmiştir. Yapılan bu çalışmada örneklem büyüklükleri eşit olmadığında Mantel-Haenszel yönteminin SIBTEST'e göre I. Tip hata oranlarının daha az ve güç oranının ise daha yüksek sonuçlandığını belirlemiştir. Bu nedenle örneklem büyüklükleri eşit olmadığında DMF belirlemede Mantel-Haenszel yönteminin

kullanılmasını önermiştir. Ayrıca orta ve yüksek düzey DMF büyüklüklerinin belirlenmesinde Mantel-Haenszel'in SIBTEST'e göre yeterli oranda güce sahip olduğunu tespit etmiştir.

Dainis (2008) Olabilirlik Oran, DFIT, Lojistik Regresyon ve Mantel-Haenszel yöntemlerinin I. Tip hata ve güç oranlarını incelemiştir. Çalışmada örneklem büyüklüğü ($N_R = 630$ ve $N_O = 360$; $N_R = 2520$ ve $N_O = 1440$), DMF büyüklüğü (0,3, 0,5 ve 1,0), tek biçimli ve tek biçimli olmayan DMF türü koşullarına göre simülatif veri üretmiştir. Elde edilen sonuçlar DFIT yönteminin tüm simülasyon koşullarında en düşük I. Tip hata oranına sahip olduğunu göstermiştir. Mantel-Haenszel yönteminin I. Tip hata oranı tüm koşullarda düşük iken, tek biçimli olmayan DMF belirlemede gücün de düşük olduğunu tespit etmiştir. Olabilirlik Oran ve DFIT yöntemlerinin güç oranlarının yüksek, Lojistik Regresyon yönteminin ise güç oranının yetersiz olduğunu saptamıştır.

Finch ve French (2007) SIBTEST, Lojistik Regresyon, MTK Olabilirlik Oran ve Doğrulayıcı Faktör Analizi yöntemlerinin performanslarını karşılaştırmışlardır. Çalışmada 60 maddelik lise matematik testinden alınan 30 maddenin parametreleri kullanılarak simülatif veri üretilmiştir. Örneklem büyüklüğü ($N_R = 250$ ve $N_O = 250$; $N_R = 500$ ve $N_O = 250$; $N_R = 500$ ve $N_O = 500$; $N_R = 1000$ ve $N_O = 500$), grupların yetenek dağılımları ve testteki DMF oranı (%0, %10 ve %20) manipüle edilen koşullardır. Ayrıca yöntemlerin performanslarını değerlendirmede 2PL ve 3PL modelleri kullanılmıştır. Sonuçlar ele alınan yöntemlerin I. Tip hata oranlarında belirgin bir fark olmadığını, I. Tip hata oranlarının kullanılan modelden, örneklem büyüklüğünden ve grupların yetenek dağılımından etkilenmediğini göstermiştir. Güç oranları incelendiğinde ise SIBTEST yönteminin en yüksek güce sahip olduğunu, MTK Olabilirlik Oran ve DFA'nın ise Lojistik Regresyon ve SIBTEST'ten çok daha düşük güce sahip olduğunu tespit etmişlerdir. Örneklem büyüklüğü arttıkça SIBTEST, MTK Olabilirlik Oran ve Lojistik Regresyon yöntemlerinde güç oranlarının artmasına rağmen, manipüle edilen değişkenlerin hiçbirinin DFA yönteminin gücünü etkilemediğini belirlemişlerdir.

Jodoin ve Gierl (2001) Lojistik Regresyon yönteminin hem tek biçimli hem de tek biçimli olmayan DMF belirlemede etki büyüklüğünü dikkate alarak, I. Tip hata ve güç oranlarının etkilenip etkilenmediğini belirlemek için simülasyon çalışması yapmışlardır. Örneklem büyüklüğü ($N_R = 250$, $N_R = 500$ ve $N_R = 1000$; $N_O = 250$, $N_O = 500$ ve $N_O = 1000$), yetenek dağılımı ve DMF'li madde oranı (%10 ve %20) manipüle edilen değişkenlerdir. Elde edilen

sonuçlara göre hem tek biçimli hem de tek biçimli olmayan DMF belirlemede Lojistik Regresyon yönteminin gücünün $N_R = 1000$ ve $N_O = 250$ olduğu durumda beklenmedik şekilde düştüğünü tespit etmişlerdir.

Cohen vd., (1996) DMF belirlemede kullanılan Olabilirlik Oran yönteminin performansını monte carlo simülasyon çalışmasıyla incelemişlerdir. Çalışmada 2PL ve 3PL modele göre küçük ve büyük örneklem büyüklüklerine (250 ve 1000) göre ve test uzunluğu 50 olacak şekilde simülatif veri üretilmiş ve MTK Olabilirlik Oran yönteminin I. Tip hata oranları hesaplanmıştır. Elde edilen sonuçlara göre 2PLM için I. Tip hata oranlarının her iki örneklem büyüklüklerinde de beklenen düzeyde olduğunu tespit etmişlerdir. 3PLM ve 3PLM-c için ise “,0005 ile ,005” seviyelerinde hata oranlarının nominal seviyelerden daha yüksek olduğunu fakat “,01 ile 0,10” seviyelerinde ise hata oranlarının beklenen düzeyde olduğunu tespit etmişlerdir. Ayrıca sonuçlar I. Tip hata kontrolünün 2PLM için çok iyi, 3PLM için ise makul düzeyde olduğunu göstermiştir.

Özmen (2013) doktora tez çalışması kapsamında PISA 2009 okuma becerileri testinde yer alan maddelerin Türkiye ile ABD, Türkiye ile Birleşik Krallık ve ABD ile Birleşik Krallık grupları arasında DMF analizlerini gerçekleştirmiş ve ardından yanlışlık çalışması yapmıştır. Araştırma PISA 2009 okuma becerileri testine ilişkin dördüncü ve altıncı kitapçıkları yanıtlayan Türkiye örnekleminde 767, Birleşik Krallık örnekleminde ise 765 ve ABD örnekleminde 796 öğrenci verisi üzerinden yürütülmüştür. Ayrıca DMF belirleme yöntemlerinden Mantel-Haenszel, SIBTEST, MTK Olabilirlik Oran ve b parametreleri ele alınmıştır. Çalışma sonucunda altıncı kitapçıkta DMF’li olarak belirlenen madde sayısının dördüncü kitapçığa göre daha az olduğunu belirlemiştir. DMF’li maddelerin gruplara göre eşit dağılım gösterdiğini; referans grubun lehine DMF gösteren maddelerin ise çoğunlukla açık uçlu maddeler olduğunu tespit etmiştir. Ayrıca SIBTEST ve MTK temelli yöntemlerden elde edilen sonuçların birbirleriyle tutarlılık gösterdiğini, Mantel-Haenszel yönteminin ise diğer yöntemlere kıyasla daha az sayıda maddeyi DMF’li olarak belirlediğini saptamıştır.

Selvi (2013) doktora tez çalışması kapsamında farklı puanlama durumlarında KTK ve MTK’ya dayalı DMF belirleme yöntemlerini incelemiştir. Çalışmayı 1593 bireyin sözel akıl yürütme yeteneği testine verdikleri yanıtlar üzerinden yürütmüştür. Mantel-Haenszel/Mantel, Standardizasyon, Lojistik Regresyon ve Olabilirlik Oran Testi yöntemlerini kullanarak yaptığı DMF analizleri sonucunda, KTK ve MTK’ya dayalı yöntemlerin kendi içinde tutarlı sonuçlar verdiğini bulmuştur. Fakat ikili puanlama

durumundan ağırlıklı puanlama ve deneysel ağırlıklı puanlama durumuna gidildikçe uyumun bozulduğunu tespit etmiştir. Aynı zamanda farklı puanlama durumlarında KTK ve MTK'ya dayalı yöntemler arasındaki uyumun düşük olduğu ve sonuçların da birbirinden anlamlı derecede farklılaştığı sonucuna ulaşmıştır.

Atalay, Gök, Kelecioğlu ve Arsan (2012) çalışmalarında DMF belirlemede kullanılan Mantel-Haenszel, Lojistik Regresyon, MTK Olabilirlik Oran ve SIBTEST yöntemlerini karşılaştırmışlardır. Yöntemlerin karşılaştırmasında örneklem büyüklüğü, yetenek dağılımı ve DMF'li madde oranı koşullarına göre simülatif veri üretmişlerdir. Çalışma sonucunda MTK Olabilirlik Oran ve SIBTEST yöntemlerinin Mantel-Haenszel, Lojistik Regresyon yöntemlerine göre daha duyarlı ve etkili olduklarını saptamışlardır. Lojistik Regresyon yönteminin ise DMF'li madde belirleme oranının diğer üç yönteme göre daha düşük olduğunu; MTK Olabilirlik Oran yönteminin ise daha yüksek olduğunu tespit etmişlerdir. Tek biçimli olmayan DMF belirlemede Lojistik Regresyon, SIBTEST ve MTK Olabilirlik Oran; tek biçimli DMF belirlemede ise Mantel-Haenszel, SIBTEST ve MTK Olabilirlik Oran yöntemlerinin birbirleriyle uyumlu sonuçlar verdiği sonucuna ulaşmışlardır.

Demir ve Köse (2014) PISA 2009 uygulamasının matematik okuryazarlığı alt testinde yer alan maddelerin cinsiyet ve kültür değişkenlerine göre DMF analizlerini yapmışlardır. Çalışma kapsamında DMF belirleme yöntemlerinden Mantel-Haenszel, SIBTEST ve Lojistik Regresyon yöntemlerini ele almışlardır. Çalışmanın veri setini kültür değişkeni için Türkiye, Finlandiya, Almanya ve ABD'deki öğrencilerin beşinci kitapçığa verdikleri yanıtlar oluştururken; cinsiyet değişkeni için Türkiye'de beşinci kitapçığı alan 332 öğrenci yanıtları oluşturmuştur. Elde edilen sonuçlara göre kullanılan yöntemlerin DMF'li madde belirlemede aynı sonuçları verdiklerini, fakat DMF düzeylerini belirleme bakımından kullanılan sınıflama sisteminden kaynaklı olarak farklılıklar olduğunu belirtmişlerdir.

Akın Arıkan vd. (2015) çalışmalarında MIMIC, SIBTEST, Mantel-Haenszel ve Lojistik Regresyon yöntemlerini karşılaştırmışlardır. Çalışmada uzman görüşleri de alınarak sonuçların tutarlılıkları incelenmiştir. Çalışmada 340000 kişilik veri setinden 300, 600, 1000, 1200 ve 2000'lik alt örneklem oluşturmuşlardır. Çalışma sonucunda örneklem sayısı büyük olduğunda kullanılan DMF belirleme yöntemlerinin birbirleriyle uyumlu sonuçlar verdiğini tespit etmişlerdir. Uzman görüşleri neticesinde ise yanlı olduğuna karar verilen dört maddenin farklı örneklem büyüklükleri ve farklı yöntemlerde de DMF'li olarak belirlendiğini tespit etmişlerdir.

Erdem (2015) yüksek lisans tez çalışmasında TEOG ortak sınavında bulunan maddelerin kitapçık türüne göre DMF analizlerini yapmıştır. Çalışmada Mantel-Haenszel, Lojistik Regresyon ve SIBTEST yöntemlerini kullanmıştır. Yöntemlere göre DMF'li olarak belirlenen madde sayılarının birbirleriyle uyumlu olmadığı; aynı maddelerin DMF'li olarak belirlenmesi yönünden ise en uyumlu Mantel-Haenszel ve SIBTEST olduğu; en az uyumlu olan yöntemlerin ise Mantel-Haenszel ve Lojistik Regresyon olduğu sonucuna ulaşmıştır.

Özdemir (2015) TIMMS 2011 matematik testinde yer alan maddelerin cinsiyete göre DMF'li olup olmadığını belirlemeyi amaçlamıştır. Bu amaçla MTK çerçevesinde Lord'un χ^2 , Raju'nun alan testi ve MTK Olabilirlik Oran yöntemlerini kullanmıştır. Aynı zamanda her bir yöntem için madde saflaştırma gerçekleştirmiş ve madde saflaştırmanın DMF'li madde sayısını nasıl etkilediğini belirleyebilmek için yöntemlerin sonuçları karşılaştırmıştır. Elde edilen sonuçlar neticesinde 2PL modelin Lord'un χ^2 ve Raju'nun alan testi yöntemleri için uygun model olduğunu tespit etmiştir. Lord'un χ^2 ve Raju'nun alan testi yöntemleriyle yapılan madde saflaştırmanın DMF'li madde sayısını etkilediğini fakat MTK Olabilirlik Oran yönteminin DMF'li madde sayısını etkilemediğini bulmuştur. Lord'un χ^2 yönteminin DMF'li maddelerin belirlenmesinde diğer iki yönteme göre daha duyarlı olduğu sonucuna ulaşmıştır. MTK'ya dayalı olan bu üç yöntemin uygulamasında kullanılan farklı varsayımlar ve kriterler nedeniyle DMF'li olarak belirlenen madde sayısının değiştiğini de tespit etmiştir. Bu nedenle Özdemir (2015) bir testteki DMF'li maddelerin belirlenmesinde sadece bir yöntemin kullanılmamasını, MTK'ya dayalı olan yöntemlerin yanı sıra MTK'ya dayalı olmayan yöntemlerin de kullanılmasını önermektedir.

Akalın (2014) çalışmasında Kamu Personeli Seçme Sınavı (KPSS) genel yetenek testinde yer alan maddelerin cinsiyet ve mezun olunan bölümlere göre DMF gösterip göstermediğini ve olası yanlışlık nedenlerini incelemiştir. Çalışma 2007 KPSS genel yetenek testini yanıtlayan toplam 4864 kişi üzerinden yürütülmüştür. DMF belirleme yöntemlerinden Mantel-Haenszel, MTK Olabilirlik Oran ve BILOG-MG kullanılmıştır. Elde edilen sonuçlarda Mantel-Haenszel, MTK Olabilirlik Oran yöntemlerinin DMF belirlemede benzer sonuçlar verdiğini, BILOG-MG yönteminin ise diğer iki yönteme kıyasla daha az sayıda maddeyi DMF'li olarak tespit ettiğini bulmuştur.

Ayan (2011) yaptığı yüksek lisans tez çalışmasında, PISA fen okuryazarlığı alt testinde yer alan maddelerin cinsiyet değişkenine göre DMF analizini gerçekleştirmiştir. Çalışma PISA 2009 uygulaması Türkiye örnekleminde 12. kitapçığı alan 338 öğrenci verisi üzerinden

yürütmüştür. Ayrıca DMF belirleme yöntemlerinden Lojistik Regresyon ve Mantel-Haenszel yöntemleri ele alınmıştır. Yapılan analizler neticesinde DMF büyüklükleri sıralaması bakımından her iki yöntemde birbirine yakın sonuçlar vermiştir. Ayrıca tek biçimli DMF belirlemede Mantel-Haenszel yönteminin daha duyarlı olduğunu belirlemiştir.

Çepni (2011) doktora tez çalışmasında Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) sayısal yetenek testlerinde cinsiyete ve mezun olunan/olunacak bölüme göre DMF ve değişen madde grubu fonksiyonu (DMGF) analizleri yapmıştır. Çalışma kapsamında DMF belirleme yöntemlerinden Lojistik Regresyon ve Mantel-Haenszel, SIBTEST, MTK Olabilirlik Oran ve BILOG-MG algoritması yöntemlerini; DMGF analizleri için ise SIBTEST yöntemini ele almıştır. Yapılan analizler sonucunda, sayısal-1 testide üç maddenin erkekler, dört maddenin ise kızlar lehine olduğunu; sayısal-2 testinde ise bir maddenin erkekler, üç maddenin ise kızlar lehine olduğunu bulmuştur.

Gök vd. (2010) yaptıkları çalışmalarında Mantel-Haenszel ve Lojistik Regresyon yöntemlerini karşılaştırmışlardır. OKS 2005 matematik ve fen bilgisi alt testlerinde yer alan maddelerin cinsiyet ve okul türü değişkenlerine göre DMF analizleri sonucunda, Mantel-Haenszel ve Lojistik Regresyon yöntemleri arasında genel olarak düşük düzeyde uyum olduğunu belirlemişlerdir.

Le (2009) PISA fen uygulamasında maddelerin cinsiyete göre DMF gösterip göstermediğini incelemiştir. Bu bağlamda 50 katılımcı ülke ile 60 test dili grubundan toplam 83000 öğrenci yanıtları çalışma kapsamında ele alınmıştır. Çalışma sonucunda maddelerin cinsiyete göre DMF göstermesinin madde biçimi ve içerik alanlarına bağlı olduğunu göstermiştir. Test dili gruplarının her birinde maddelerin %5,6'sının erkekler ve %2,8'inin ise kadınlar lehine olduğunu belirlemiştir. Ayrıca çoktan seçmeli ve kapalı uçlu maddelerin de erkekler lehine olduğu tespit edilmiştir.

Doğan ve Öğretmen (2008) çalışmalarında Mantel-Haenszel, Lojistik Regresyon ve χ^2 yöntemlerini karşılaştırmışlardır. 2003 yılı OKÖSYS sınavına katılan 600000 öğrenci arasından yansız olarak seçilen 3345 öğrencinin fen bilgisi alt testine verdikleri yanıtlardan oluşan veri seti üzerinden çalışmayı gerçekleştirmişlerdir. Cinsiyet grubu üzerinden yapılan analizler neticesinde kullanılan yöntemlerden elde edilen korelasyonlara dayanarak, χ^2 temelli olan bu yöntemlerin aralarında önemli farklılıklar olmadığını ve birbirine benzer sonuçlar verdiklerini dolayısıyla hepsinin de kullanılabileceğini belirtmişlerdir. DMF'li madde sayısı dikkate alındığında ise Lojistik Regresyon yönteminin daha duyarlı sonuçlar

verdiğini, Mantel-Haenszel ve işaretli χ^2 yöntemlerinin ise birbiri yerine kullanılabileceğini saptamışlardır.

Alanyazında yer alan farklı DMF belirleme yöntemlerinin I. Tip hata ve güç oranlarını karşılaştıran araştırmalar ve DMF'nin yetenek kestirimi üzerindeki etkilerini inceleyen çalışmalar yukarıda özetlenmiştir.

Yurt içinde ve yurt dışında DMF belirleme yöntemlerini kıyaslayan araştırmalar mevcuttur. Fakat bu çalışmalar incelendiğinde çoğunlukla DMF belirleme yöntemlerinin I. Tip hata ve güç oranlarının karşılaştırıldığı görülmektedir. Ayrıca DMF'nin yetenek kestirimi üzerindeki etkilerini inceleyen sınırlı sayıda araştırma ile karşılaşmıştır. Buna rağmen alanyazında DMF belirleme yöntemlerini madde ve yetenek parametre kestirimlerine dayalı olarak inceleyen çalışmalara rastlanmamıştır. Alanyazında tamamen benzer bir çalışmaya rastlanmadığından dolayı yapılan bu araştırma diğer çalışmalardan farklılık göstermektedir.

BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın türü, verinin üretilmesi, simülasyon koşulları ve veri analizi ile ilgili açıklamalara yer verilmiştir.

3.1. Araştırmanın Türü

Bu araştırma iki kategoride puanlanan test maddelerinde DMF belirleme yöntemlerinin performansları ile yöntemlerin madde ve yetenek parametre kestirimlerine olan etkisini incelemeyi amaçlamaktadır. Araştırma var olan durumun ortaya konmaya çalışıldığı betimsel bir çalışmadır. Betimsel araştırmalarda incelenen olayların daha önceki durumları dikkate alınarak değişkenler arasındaki ilişkiler açıklanarak betimlenmeye çalışılır (Brown vd.,1999; Karasar, 1998).

3.2. Veriler

Araştırmada farklı koşullar altında DMF belirleme yöntemlerinin I. Tip hata ve güç oranları ile yöntemlerin madde ve yetenek parametre kestirimlerine olan etkisinin incelenmesinde simülatif veri üretilmiştir. Verilerin üretilmesinde ele alınan simülasyon koşulları aşağıda açıklanmaktadır.

3.2.1. Simülasyon Koşulları

Bu başlık altında verilerin hangi koşullar altında üretildiği ayrıntılarıyla sunulmuştur. Sabit ve değişen koşullar açıklamalarıyla aşağıda yer almaktadır.

3.2.1.1. Testte Bulunan Madde Sayısı

Genellikle uzun testler MTK uygulamalarında daha doğru ve kararlı parametre kestirimleri sağlar ve güvenilirliği daha fazladır (Baker, 2001; Hambleton & Cook, 1983; Hambleton, 1989; Lord, 1980; Swaminathan & Gifford, 1983). Ayrıca eşleştirme kriteri gözlenen puana dayalı olan DMF belirleme yöntemleri için test uzunluğunda etkisi vardır (Swaminathan & Rogers, 1990; Zwick, Donoghue & Grima, 1993). DMF kapsamında yürütülen simülasyon çalışmalarında araştırmacıların test uzunluğunu farklı sayılarda (20 ile 80 aralığında) ele aldıkları görülmüştür (Clauser vd., 1993; Güler & Penfield, 2009; Jodoin & Gierl, 2001; Narayanan & Swaminathan, 1994, 1996; Rogers & Swaminathan, 1993; Wells, Subkoviak & Serlin, 2002). Jodoin ve Gierl (2001); Narayanan ve Swaminathan (1994); Raju, Linden ve Fler (1995); Rogers ve Swaminathan (1993) ise DMF kapsamında 40 maddelik test uzunluğu üzerinden çalışmalarını yürütmüşlerdir. Literatürde yer alan DMF çalışmaları dikkate alınarak bu çalışmada testte bulunan madde sayısı 40 olarak belirlenmiştir.

3.2.1.2. Kullanılan MTK Modeli

Finch ve French (2007), DMF kapsamında yapılan çalışmaların ayırt edicilik (*a*) ve güçlük (*b*) parametrelerine dayalı olarak yapıldığını belirtmişlerdir. Hambleton vd. (1991) yaptıkları çalışmalarında 3PL modelin kullanılmasıyla DMF belirlemede güçlü bir test istatistiğinin elde edilemediğini belirtmişlerdir. Ayrıca Kim, Cohen ve Kim (1994) ise şans (*c*) parametresinin yaptıkları çalışma sonuçları üzerinde bir etkisini bulamamışlardır. Bütün bunlar dikkate alınarak bu çalışmada referans ve odak gruplara ilişkin madde yanıtları ile madde ve yetenek parametreleri 2PL modele uygun olarak üretilmiştir.

3.2.1.3. DMF Türü

Araştırma kapsamında ele alınan Mantel-Haenszel, Lord'un χ^2 , Lojistik Regresyon, SIBTEST ve Raju'nun alan ölçümü yöntemleri tek biçimli DMF koşulu altında

incelenmiştir. Tek biçimli DMF genellikle b parametresi değiştirilerek simüle edilir; tek biçimli olmayan DMF’de ise genellikle b parametre değerleri sabit tutulurken a parametre değerleri değiştirilerek simüle edilir (Swaminathan & Rogers, 1990).

Bu çalışmada tek biçimli DMF koşulunun sağlanabilmesi için ayırt edicilik parametre değerleri sabit tutulurken, referans grubuna ait güçlük parametre değerlerine DMF gösterecek olan maddeler için belirlenen DMF büyüklükleri tanımlanmış ve odak grubun madde güçlük parametreleri elde edilmiştir.

3.2.1.4. Örneklem Büyüklüğü ile Odak ve Referans Grupların Örneklem Büyüklüğü Oranları

Örneklem büyüklüğü istatistiksel yöntemlerin gücünü etkiler. Örneklem büyüklüğünün artırılması özellikle DMF analizinde sıklıkla kullanılan MTK parametre kestirimlerinin doğruluğunu artıracaktır. Yeterince büyük örneklem büyüklüğü ile istatistiksel modeller önemsiz bir etkiyi bile tespit etmede yeterli güce sahip olacaktır (Cohen, 1988). Bu sebeple yeterli güce sahip olan uygun bir örneklem büyüklüğünü belirlemek istatistiksel ve psikometrik araştırmaların temel sorunlarından biridir (Li, 2009).

DMF çalışmalarında odak ve referans grubun örneklem büyüklükleri küçük olduğunda DMF belirleme yöntemlerinin gücü çarpıcı şekilde düşmektedir (Güler & Penfield, 2009). Odak ve referans grubun örneklem büyüklüklerinin eşit olmadığı koşullarda ise DMF belirleme yöntemlerinin performansları etkilenmektedir (Kristjansson, Aylesworth, Mcdowell & Zumbo, 2005). Ayrıca eşit olmayan örneklem büyüklükleri DMF analizlerinde gücü tehdit etmektedir (Broer, Lee, Rizavi & Powers, 2005; Mazor, Clauser & Hambleton, 1992; Paek & Guo, 2011; Zieky, 1993).

Alanyazında DMF kapsamında yapılan simülasyon çalışmalarında en küçük örneklem büyüklüğünün odak ve referans grup için 200 ile 250 arasında belirlendiği görülmüştür (Narayanan & Swaminathan, 1996; Rogers & Swaminathan, 1993). Orta büyüklükteki örneklem büyüklüğünün odak ve referans grup için 400, 500 ile 600 ve büyük örneklemelerde ise 750, 800, 1000, 1200, 2000 ile 3000 olarak belirlendiği görülmüştür. (Atar, 2007; Atar & Kamata, 2011; Bolt, 2002; Carter, 2011; Duncan, 2006; Finch & French, 2008; French & Maller, 2007; Johnson Frotman, 2007; Jodoin & Gierl, 2001; Kristjansson vd., 2005; Li, Brooks & Johanson, 2012; Wang, Tay & Drasgow, 2013; Wyse & Mapuranga, 2009).

Ayrıca yapılan çalışmalar Mantel-Haenszel ve Lojistik Regresyon yöntemlerini uygularken her bir grup için en az 200 ile 250 örneklem büyüklüğünün gerekli olduğunu göstermektedir (Narayanan & Swaminathan, 1996; Rogers & Swaminathan, 1993). Gierl vd. (2004) ise örneklem büyüklüğü 1000 olduğunda SIBTEST yönteminin tek biçimli DMF belirlemede yeterli olduğunu belirtmişlerdir. Bu doğrultuda alanyazında yer alan DMF çalışmaları dikkate alınarak bu çalışmada örneklem büyüklüğü küçük, orta ve büyük olmak üzere üç düzeyde ele alınmıştır ($N_{Küçük}=600$, $N_{Orta}=1200$ ve $N_{Büyük}=2400$).

DMF çalışmalarında odak grup sayısı referans grup sayısına göre daha az ya da eşit olduğu görülmüştür (Ankenmann, Witt & Dunbar, 1999; Garrett, 2009; French & Maller, 2007; Wood, 2011). Bu çalışmada odak ve referans grupların arasındaki oran hem dengeli (1/1) hem de dengesiz (1/2) olmak üzere iki düzey olarak belirlenmiştir. Dolayısıyla örneklem büyüklüğü $N=600$ iken 300/300 ve 200/400; $N=1200$ iken 600/600 ve 400/800; $N=2400$ iken 1200/1200 ve 800/1600 olmak üzere 6 koşul olarak ele alınmıştır.

3.2.1.5. Testte Bulunan DMF'li Maddelerin Oranı

DMF kapsamında yapılan araştırmalar incelendiğinde testte yer alan DMF'li maddenin bir veya daha fazla sayıda ele alındığı çalışmalara rastlanmıştır (Bolt, 2002; Chang, Mazzeo & Roussos, 1996; Garrett, 2009; Su & Wang, 2005; Wang & Su, 2004; Zwick, Thayer & Mazzeo, 1997). Finch ve French (2007) ile French ve Maller (2007) ise çalışmalarında DMF'li madde sayısını %0, %10 ve %20 oranlarında belirlemişlerdir. Alanyazındaki çalışmalar testte yer alan DMF'li madde oranının yetenek parametre kestirimi üzerinde etkili olduğunu göstermektedir (Walker, Zhang, Banks & Cappaert, 2012; Wells vd., 2002; Zumbo, 2003). Ayrıca testteki DMF'li madde oranının fazla olmasının da yetenek parametre kestirimlerinin hassaslığının azalmasına neden olduğu da belirtilmektedir (Finch, 2005; Wang & Yeh, 2003). Jodoin ve Gierl (2001)'e göre testteki DMF'li madde sayısının fazla olmasının DMF belirleme yöntemlerinin gücünün azalmasına ve geçerliğin de düşmesine neden olmaktadır. Ek olarak bazı araştırmacılar ise sonuçların güvenilirliğini sağlamak için büyük örneklem ve uzun testlerin kullanılmasını önermektedir (Chan, 2000; Millsap & Meredith, 1992). Alanyazındaki çalışmalar dikkate alınarak bu çalışmada her bir simülasyon koşulu için DMF'li madde oranı %5, %10 ve %20 olarak belirlenmiştir.

3.2.1.6. DMF Büyüklüğü

Alanyazındaki çalışmalar incelendiğinde DMF büyüklüğü için farklı değerlerin (0.10, 0.15, 0.25, 0.30, 0.32, 0,4, 0,43, 0.53, 0,6 ve 0,8) ele alındığı görülmüştür (Atar, 2007; Fidalgo, Mellenberg & Muniz, 2000; French & Maller, 2007; Kristjansson, 2001; Wang & Su, 2004; Zwick, Donoghue & Grima, 1993). Ayrıca ilgili çalışmalar DMF büyüklüğünün yetenek kestirimi üzerinde etkili olduğunu göstermiştir (Walker vd., 2012; Wells vd., 2002; Zumbo, 2003). Bu araştırmada ise DMF büyüklükleri 0,5 ve 1,0 olarak belirlenmiştir.

Sabit ve değişen koşullar ile ilgili yukarıda yapılan açıklamalar doğrultusunda oluşturulan simülasyon deseni Tablo 8’de sunulmuştur.

Tablo 8.

Simülasyon Deseni

Koşul	Örneklem büyüklüğü	Örneklem büyüklüğü oranı	DMF büyüklüğü	DMF'li madde oranı	DMF'li madde numarası
1	600	300:300 (1/1)	0,5	%5	20,40
2				%10	10, 20, 30, 40
3				%20	5, 10, 15, 20, 25, 30, 35, 40
4			1,0	%5	20,40
5				%10	10, 20, 30, 40
6				%20	5, 10, 15, 20, 25, 30, 35, 40
7	600	200:400 (1/2)	0,5	%5	20,40
8				%10	10, 20, 30, 40
9				%20	5, 10, 15, 20, 25, 30, 35, 40
10			1,0	%5	20,40
11				%10	10, 20, 30, 40
12				%20	5, 10, 15, 20, 25, 30, 35, 40
13	1200	600:600 (1/1)	0,5	%5	20,40
14				%10	10, 20, 30, 40
15				%20	5, 10, 15, 20, 25, 30, 35, 40
16			1,0	%5	20,40
17				%10	10, 20, 30, 40
18				%20	5, 10, 15, 20, 25, 30, 35, 40
19	1200	400:800 (1/2)	0,5	%5	20,40
20				%10	10, 20, 30, 40
21				%20	5, 10, 15, 20, 25, 30, 35, 40
22			1,0	%5	20,40
23				%10	10, 20, 30, 40
24				%20	5, 10, 15, 20, 25, 30, 35, 40
25	2400	1200:1200 (1/1)	0,5	%5	20,40
26				%10	10, 20, 30, 40
27				%20	5, 10, 15, 20, 25, 30, 35, 40
28			1,0	%5	20,40
29				%10	10, 20, 30, 40
30				%20	5, 10, 15, 20, 25, 30, 35, 40
31	2400	800:1600 (1/2)	0,5	%5	20,40
32				%10	10, 20, 30, 40
33				%20	5, 10, 15, 20, 25, 30, 35, 40
34			1,0	%5	20,40
35				%10	10, 20, 30, 40
36				%20	5, 10, 15, 20, 25, 30, 35, 40

Tablo 8’de yer alan simülasyon desenine göre araştırma kapsamında üç farklı örneklem büyüklüğü (600, 1200 ve 2400), iki farklı odak ve referans grubun örneklem büyüklüğü oranı (1/1 ve 1/2), üç farklı DMF’li madde oranı (%5, %10 ve %20) ve iki farklı DMF büyüklüğü (0,5 ve 1,0) olmak üzere toplam 36 simülasyon koşulu oluşturulmuştur.

Simülasyon çalışmalarının örnekleme teorisine dayanmasından dolayı belirli sayıda replikasyon yapılarak bu araştırmada istenen özellikte verinin üretilmesi sağlanmıştır. Bu

sayede yapılan kestirimlerin bir veri setinden diğetine değışmemesi sağlanabilir. Replikasyon sayısının çok olması daha güvenilir ve uygun kestirimler elde etme anlamına gelmektedir. Buna rağmen replikasyon sayısının artması analiz sürecinin uzamasına ve belirli bir tekrar sayısından sonra kestirimlerdeki iyileşme miktarı azalmaktadır (Kılıç, Uysal & Doğan, 2018). Alanyazında yer alan çalışmalarda (Bolt, 2002; Narayanan & Swaminathan, 1996; Rogers & Swaminathan, 1993; Wen, 2014) ve 2009 yılında Eğitimde Ölçme Ulusal Konseyi'nin (NCME) yayınlarında replikasyon sayısının 100 olarak ele alındığı görülmüştür. Yapılan çalışmalar dikkate alınarak bu araştırmada her bir koşul için 100 replikasyon yapılmıştır. Böylece 36 koşulun her biri için yapılan 100 replikasyon neticesinde analizler toplam 3600 veri seti üzerinden gerçekleştirilmiştir.

3.3. Veri Üretimi

Bu araştırmada simülatif veri üretilirken öncelikli olarak WinGen yazılımı aracılığıyla 2PL modele uygun madde parametre değerleri üretilmiştir (bkz. Tablo 9). WinGen yazılımı gerçek veriye uygun olarak ikili ve çoklu puanlanan madde yanıt verilerini üretmek için geliştirilen ve kullanıcı dostu arayüz sunan bir yazılımdır. Yazılımla çeşitli dağılımlara (normal, uniform, beta veya lognormal) uygun olarak madde ve yetenek parametreleri üretilir. Ayrıca yazılım birçok MTK modelini de (1PLM, 2PLM, 3PLM, KPM, GKPM vd.) desteklemektedir. Aynı zamanda PARSCALE, BILOG-MG ve MULTILOG gibi MTK programları için syntax dosyaları oluşturabilir ve 1.000.000 veri kümesine kadar tekrar sağlayarak veri üretebilir (Han & Hambleton, 2014).

WinGen yazılımı ile öncelikle yetenek parametresinin normal dağılıma uygun olduğu varsayımı altında (ortalama=0, SS=1) madde sayısı 40, yanıt kategorisi iki, 2PL modele ve normal dağılıma uygun madde parametreleri (a ve b) üretilmiştir. Ardından R programı aracılığıyla referans grubuna ait madde güçlük parametrelerine araştırma kapsamında DMF gösterecek olan maddeler için belirlenen DMF büyüklükleri tanımlanmış ve böylece odak grubun madde güçlük parametre değerleri elde edilmiştir. Simülasyon çalışması için üretilen madde parametreleri Tablo 9'da sunulmuştur.

Tablo 9.

Simülasyon Çalışması İçin Üretilen Madde Parametre Değerleri

Madde No	Yanıt Kategorisi	Model	<i>a</i>	<i>b</i>	Madde No	Yanıt Kategorisi	Model	<i>a</i>	<i>b</i>
1	2	2PLM	1,484	-0,484	21	2	2PLM	1,105	-1,282
2	2	2PLM	1,005	0,233	22	2	2PLM	0,937	-0,170
3	2	2PLM	1,437	-0,989	23	2	2PLM	0,712	-1,428
4	2	2PLM	0,875	0,402	24	2	2PLM	1,128	0,853
5	2	2PLM	1,018	-0,323	25	2	2PLM	1,124	-1,036
6	2	2PLM	0,856	-0,430	26	2	2PLM	0,793	1,092
7	2	2PLM	0,926	-0,473	27	2	2PLM	0,840	0,364
8	2	2PLM	0,996	-0,504	28	2	2PLM	1,368	-1,820
9	2	2PLM	0,660	-0,812	29	2	2PLM	0,788	-1,161
10	2	2PLM	0,894	-1,087	30	2	2PLM	0,836	1,172
11	2	2PLM	0,873	0,640	31	2	2PLM	0,832	-0,398
12	2	2PLM	0,828	1,785	32	2	2PLM	0,567	0,292
13	2	2PLM	1,035	0,760	33	2	2PLM	1,166	0,570
14	2	2PLM	0,996	-1,333	34	2	2PLM	0,963	1,441
15	2	2PLM	1,319	0,128	35	2	2PLM	1,015	-0,598
16	2	2PLM	0,925	0,431	36	2	2PLM	0,946	-0,111
17	2	2PLM	1,026	-1,251	37	2	2PLM	1,162	0,362
18	2	2PLM	1,116	0,863	38	2	2PLM	1,258	0,191
19	2	2PLM	1,044	-0,902	39	2	2PLM	1,082	0,281
20	2	2PLM	0,923	-0,421	40	2	2PLM	0,992	1,172

Tablo 9’da sunulan madde parametrelerinin üretilmesinin ardından iki kategorili (1-0) yanıt verileri R programında yer alan “eRm” paketi aracılığıyla üretilmiştir (bkz. <https://CRAN.R-project.org/package=eRm>). Yanıt verilerini üretmede kullanılan syntax örneği Ek 11’de yer almaktadır.

Araştırma kapsamında kullanılan R programının en büyük avantajlarından biri ise çeşitli olasılık dağılımlarını (standart normal, bernoulli veya düzgün dağılımı vd.) kullanarak değişkenleri ve veri kümelerini oluşturmaktır. Bu program araştırmacıların psikometrik modeller ve yöntemler kullanarak istatistiksel analiz yapmalarını sağlayan birçok pakete sahiptir (Bulut & Sünbül, 2017).

3.4. Veri Analizi

DMF Belirleme Yöntemlerinin Analizleri ve Değerlendirme Ölçütleri

Araştırmada DMF analizleri Mantel-Haenszel, Lord’un χ^2 , Lojistik Regresyon, SIBTEST ve Raju’nun alan ölçümü yöntemleri kullanılarak gerçekleştirilmiştir. DMF yöntemlerinin analizleri R programında yer alan “difR” ve “mirt” paketleri aracılığıyla yapılmıştır.

Mantel-Haenszel yöntemine ilişkin analizler R programında yer alan “difR” paketi aracılığıyla gerçekleştirilmiştir. Analiz sonucunda elde edilen Δ_{MH} katsayısının değerlendirmesinde ETS tarafından önerilen ölçütler dikkate alınmıştır (Wiberg, 2007);

“A düzeyi ($|\Delta_{MH}| < 1$), B düzeyi ($1 \leq |\Delta_{MH}| < 1.5$) ve C düzeyi ($|\Delta_{MH}| \geq 1.5$)’dir”.

SIBTEST yöntemine ilişkin analizler R programında yer alan “mirt” paketi aracılığıyla gerçekleştirilmiştir. Analiz sonucunda elde edilen β_U katsayısının değerlendirmesinde Roussos ve Stout (1996) tarafından önerilen ölçütler dikkate alınmıştır;

“A düzeyi ($\beta_U < 0,059$), B düzeyi ($0,059 \leq \beta_U < 0,088$) ve C düzeyi ($\beta_U \geq 0,088$)’dir”.

Lojistik Regresyon yöntemine ilişkin analizler R programında yer alan “difR” paketi aracılığıyla gerçekleştirilmiştir. Analiz sonucunda elde edilen ΔR^2 katsayısının değerlendirmesinde Zumbo ve Thomas (1996) tarafından önerilen ölçütler dikkate alınmıştır;

“A düzeyi ($\Delta R^2 < 0,13$), B düzeyi ($0,13 \leq \Delta R^2 < 0,26$) ve C düzeyi ($0,26 \leq \Delta R^2$)’dir”.

Lord’un χ^2 yöntemine ilişkin analizler R programında yer alan “difR” paketi aracılığıyla gerçekleştirilmiştir. Analiz sonucunda elde edilen Δ_{Lord} istatistiğinin değerlendirmesinde ETS tarafından önerilen ölçütler dikkate alınmıştır (Holland & Thayer, 1985);

“A düzeyi ($\Delta_{Lord} < 1$), B düzeyi ($1 \leq \Delta_{Lord} < 1,5$) ve C düzeyi ($\Delta_{Lord} \geq 1,5$)’dir”.

Raju’nun alan ölçümü yöntemine ilişkin analizler R programında yer alan “difR” paketi aracılığıyla gerçekleştirilmiştir. Analiz sonucunda elde edilen NCDIF istatistiğinin değerlendirmesinde ise Wright ve Oshima (2015) tarafından önerilen ölçütler dikkate alınmıştır;

“A düzeyi ($NCDIF < 0,003$), B düzeyi ($0,003 \leq NCDIF < 0,008$) ve C düzeyi ($NCDIF \geq 0,008$)’dir”.

I. Tip Hata ve Güç Oranlarının İncelenmesi

Araştırma kapsamında ele alınan DMF belirleme yöntemlerinin belirlenen simülasyon koşulları altındaki performanslarının karşılaştırılması amacıyla yöntemlerin I. Tip hata ve güç oranları hesaplanmıştır. DMF çalışmalarında I. Tip hata gerçekte DMF’li olarak tanımlanmayan maddelerin yanlışlıkla DMF’li olarak belirlenmesi anlamına gelmektedir.

DMF’li maddelerin yanlış belirlenmesi değerdendirmenin geerliđini zayıflatabilir (Kim, 2010). DMF alıřmalarında g ise gerekte DMF’li olarak tanımlanan maddelerin dođru bir řekilde DMF’li olarak belirlenmesi anlamına gelmektedir (Vaughn & Wang, 2010).

Ampirik (deneysel) veri analizlerinde bir maddenin DMF’li olarak dođru belirlenip belirlenmediđini veya ka maddenin hatalı bir řekilde DMF’li olarak belirlendiđini tespit etmenin bir yolu yoktur. Fakat Monte Carlo gibi simlasyon alıřmalarında maniple edilen kořulların (DMF byklđ, yetenek dađılımları, rneklem byklđ, DMF’li madde sayısı vb.) bilinmesi kullanılan yntemlerin hangi kořullarda daha iyi sonular verdiđinin kesin olarak tespitini sađlar (Fidalgo, Ferreres & Muniz, 2004). Ayrıca birden fazla DMF belirleme ynteminin birlikte kullanılması da testte yer alan maddelerin ka tanesinin DMF’li olarak belirlendiđi konusundaki kesinlik seviyesini de artırır (Kim & Cohen, 1998; Shealy & Stout, 1993).

Bu arařtırmada her replikasyon iin hatalı řekilde DMF’li olarak belirlenen madde sayılarının oranı hesaplanarak I. Tip hata oranları elde edilmiřtir. rneđin, 40 maddelik bir testte 4 adet maddenin DMF’li olarak tanımlandıđı SIBTEST yntemi iin Tablo 10’da 100 replikasyon sayına iliřkin bilgiler yer almaktadır (0: DMF’siz olarak belirlenen maddeler; 1: DMF’li olarak belirlenen maddeler; “1, 5, 10 ve 15.” maddeler ise gerekte DMF’li olarak tanımlanan maddeler).

Tablo 10.

I. Tip Hata ve Güç Oranlarının Hesaplanmasına İlişkin Örnek Tablo

Madde	Replikasyon sayısı								
	1	2	3	4		97	98	99	100
1*	1	1	1	0		1	1	0	1
2	0	0	0	0		1	0	0	1
3	0	0	0	0		0	0	0	0
4	0	0	0	0		0	0	0	0
5*	1	1	0	0		0	1	0	1
6	0	0	0	1		0	0	0	0
7	0	0	0	0		0	0	0	0
8	0	1	1	0		0	0	0	0
9	0	0	0	0		0	0	0	0
10*	1	1	1	1		0	1	1	0
11	0	0	0	0		0	0	0	0
12	1	0	0	0		0	0	0	0
13	0	0	0	0		0	0	0	1
14	0	0	0	1		1	0	0	0
15*	0	0	1	1		1	0	0	1
16	0	0	0	0		0	0	0	0
17	0	0	0	0		0	0	0	0
18	0	1	0	0		0	0	0	0
19	0	0	0	0		0	0	0	0
20	0	0	0	0		0	0	0	0
21	0	0	0	0		0	0	0	0
22	0	0	0	0		0	0	0	0
23	0	0	0	0		0	0	0	0
24	0	0	0	0		0	0	0	0
25	0	0	0	0		0	0	0	0
26	1	0	1	1		1	1	1	1
27	0	0	0	0		0	0	0	0
28	0	0	0	0		0	0	0	0
29	0	0	0	0		0	0	0	0
30	1	0	1	1		1	1	1	1
31	0	0	0	0		0	0	0	0
32	0	0	0	0		0	0	0	0
33	0	0	0	0		0	0	0	0
34	0	0	0	0		0	0	0	0
35	0	0	0	1		0	0	0	0
36	0	0	0	0		0	0	1	0
37	0	0	0	1		0	0	0	0
38	0	0	0	0		0	0	0	0
39	0	0	1	0		0	0	0	0
40	0	0	0	0		0	0	0	0
H.DMF	3	2	4	6		4	2	3	4
D.DMF	3	3	3	2		2	3	1	3

Not: 1*, 5*, 10* ve 15* gerçekte DMF'li olarak tanımlanan maddeler. H.DMF: Hatalı olarak DMF'li belirlenen madde sayısı. D. DMF: Doğru olarak DMF'li belirlenen madde sayısı.

Tablo 10'da yer alan bilgilere göre birinci tekrarda üç, ikinci tekrarda iki, yüzüncü tekrarda dört vs. madde hatalı şekilde DMF'li olarak belirlenmiştir. Bu bilgilere göre;

I. Tip hata oranı = $1/100 \cdot (3/36 + 2/36 + 4/36 + 6/36 + \dots + 4/36 + 2/36 + 3/36 + 4/36) = 0,097$ 'dir.

Dolayısıyla SIBTEST yöntemi 40 maddelik testte ortalama % 9,7 maddeyi hatalı bir şekilde DMF'li olarak belirlemiştir.

Güç oranları da I. Tip hata oranlarına benzer şekilde hesaplanmıştır. Tabloda yer alan bilgilere göre;

Güç oranı = $1/100 \cdot (3/4 + 3/4 + 2/4 + \dots + 3/4) = 0,75$ 'tir.

I. Tip hata ve güç oranlarının değerlendirilmesinde Bradley'in (1978) kriteri kullanılmıştır. Bu kritere göre I. Tip hata olasılığı;

α ,05 düzeyinde “ $,025 \leq \text{I. Tip hata oranı} \leq ,075$ ” ve

α ,01 düzeyinde “ $,0055 \leq \text{I. Tip hata oranı} \leq ,015$ ” dir.

Yapılan hesaplamalar neticesinde elde edilen değerler bu aralıkta yer alıyorsa “I. Tip hata oranının uygun hata sınırları içinde olduğu” şeklinde yorumlanmıştır. Güç oranlarının değerlendirilmesinde ise elde edilen değerlerin “,80” den büyük olması durumunda, “kullanılan yöntemin yeterli düzeyde DMF belirleme gücüne sahip olduğu” şeklinde yorum yapılmıştır. Yapılan hesaplamaların ardından DMF belirleme yöntemlerine göre I. Tip hata ve güç oranlarının farklılaşma durumunu ve yöntemlerle simülasyon koşullarının etkileşimini test etmek amacıyla IBM SPSS Statistic programı aracılığıyla Faktöriyel ANOVA analizleri yapılmıştır. Ayrıca anlamlı farklılığın olduğu durumlarda etki büyüklükleri (η^2) raporlanarak grafiklerle sunulmuştur.

DMF Belirleme Yöntemlerinin Madde ve Yetenek Parametre Kestirimlerine Olan Etkisinin İncelenmesi

DMF belirleme yöntemlerine göre madde ve yetenek parametre kestirimlerinin incelenmesinde öncelikle her bir simülasyon koşulu için DMF içermeyen 40 maddeye ilişkin madde ve yetenek parametre kestirimleri yapılmıştır. Ardından beş farklı DMF yönteminin DMF'li olarak belirlediği maddeler cevap matrisinden çıkarılarak R programında tekrar madde ve yetenek parametre kestirimleri yapılmıştır. Yöntemlerin ilgili koşullar altında DMF'li olarak belirlediği madde oranları Ek 11'de yer almaktadır. Ardından her bir DMF belirleme yönteminin sonuçlarını karşılaştırmak için kestirilen madde ve yetenek parametre değerlerinin RMSE, standart hata ve yanlışlık (BIAS) değerleri hesaplanmıştır. Yapılan hesaplamaların ardından madde ve yetenek parametre değerlerinin yöntemlere göre farklılaşma durumunu ve yöntemlerle simülasyon koşullarının birbiriyle olan etkileşimini test etmek amacıyla Faktöriyel ANOVA analizleri yapılmıştır. Ayrıca anlamlı farklılığın

olduğu durumlarda etki büyüklükleri (η^2) raporlanmıştır. Hesaplanan kriterlere (RMSE, BIAS, SH ve η^2) ilişkin açıklamalar aşağıda yer almaktadır.

Yanlılık ve standart hata RMSE'nin iki bileşenidir ($RMSE^2 = BIAS^2 + SH^2$). Yanlılık, kestirilen parametre ile gerçek parametre değeri arasındaki farkın toplamının replikasyon sayısına bölünmesiyle elde edilmektedir. Yanlılığın hesaplanması Eşitlik 18'de verilmiştir.

$$BIAS = \frac{\sum_{i=1}^N (\hat{b}_i - b_i)}{N_n} \quad (18)$$

Eşitlik 18'de yer alan b gerçek parametre değerini, $\hat{b}$ kestirilen parametre değerini; N simülasyon çalışmasındaki replikasyon sayısını; n ise örneklem büyüklüğünü temsil etmektedir. Yanlılık değeri 0'a eşit olduğunda yapılan kestirimin yansız olduğu kabul edilmektedir (Wen, 2014).

RMSE ise kestirilen parametre ile gerçek parametre değerleri arasındaki mutlak farkın kareleri toplamının replikasyon sayısına oranının kareköküdür. RMSE değeri toplam hata olarak adlandırılır. RMSE ve yanlılık değerlerinin küçük olması, yani kestirilen parametre değerinin gerçek parametre değerine yakın olması ise yapılan kestirimin etkili olduğunun bir göstergesidir. RMSE'nin hesaplanması Eşitlik 19'da verilmiştir (Wen, 2014).

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{b}_i - b_i)^2}{N_n}} \quad (19)$$

Etki büyüklüğü bağımsız değişkenin düzeyleriyle ilişkilendirilen bağımlı değişken varyansı orantısıdır. Dolayısıyla etki büyüklüğü bağımlı değişken içindeki toplam varyans miktarını değerlendirir. Grup farklılıklarının değerlendirilmesinde kullanılan manidarlık testleri, karşılaştırmalar ve parametre kestirimleri bağımsız değişken(ler) ve bağımlı değişken(ler)in aralarındaki ilişkileri değerlendiremezler. Ayrıca manidarlık testleri bağımlı ve bağımsız değişkenlerin arasındaki bağlantının güvenilirliğini değerlendirirken, etki büyüklüğü var olan bağlantının miktarını ölçer. Etki büyüklüğünün hesaplanması Eşitlik 20'de verilmiştir (Tabachnick & Fidell, 2015).

$$\eta^2 = \frac{SS_{etki}}{SS_{etki} + SS_{hata}} \quad (20)$$

Etki büyüklüğünün yorumlanmasında kullanılan ölçütler ise (Kirk, 1996);

“ $\eta^2=0,01$, düşük etki”, “ $\eta^2=0,06$, orta etki” ve “ $\eta^2=0,138$, yüksek etki”dir.

BÖLÜM IV

BULGULAR VE YORUM

Bu bölümde araştırma sorularına ilişkin bulgu ve yorumlara yer verilmiştir.

4.1. Farklı Koşullar Altında DMF Belirleme Yöntemlerinin I. Tip Hata Oranları

“DMF belirleme yöntemlerinin I. Tip Hata oranları örneklem büyüklükleri (600, 1200 ve 2400), odak ve referans grup örneklem büyüklüğü oranları (1/1 ve 1/2), DMF büyüklükleri (0,5 ve 1,0) ve testte bulunan DMF’li maddelerin oranları (%5, %10 ve %20) koşullarında nasıl değişmektedir?” araştırma sorusuna yönelik elde edilen bulgular aşağıda verilmiştir.

DMF belirleme yöntemlerinin ilgili koşullar altında I. Tip Hata oranları Tablo 11’de sunulmuştur.

Tablo 11.

İlgili Koşullar Altında DMF Belirleme Yöntemlerinin I. Tip Hata Oranları

			DMF belirleme yöntemleri								Raju'nun alan ölçümü	
Koşullar			M-H		Lord'un χ^2		LR		SIBTEST			
Örneklem büyüklüğü	Örneklem oranı	DMF'li madde oranı	DMF büyüklüğü		DMF büyüklüğü		DMF büyüklüğü		DMF büyüklüğü		DMF büyüklüğü	
			0,5	1,0	0,5	1,0	0,5	1,0	0,5	1,0	0,5	1,0
600	300/300	%5	0,042	0,040	0,035	0,040	0,057	0,054	0,035	0,032	0,047	0,086
		%10	0,042	0,049	0,047	0,077	0,062	0,061	0,028	0,028	0,053	0,100
		%20	0,061	0,097	0,049	0,102	0,075	0,099	0,043	0,054	0,075	0,115
	200/400	%5	0,040	0,042	0,041	0,046	0,057	0,054	0,041	0,037	0,045	0,055
		%10	0,040	0,044	0,040	0,049	0,061	0,061	0,035	0,037	0,051	0,065
		%20	0,054	0,092	0,044	0,083	0,065	0,095	0,047	0,051	0,053	0,092
1200	600/600	%5	0,043	0,049	0,035	0,050	0,051	0,056	0,041	0,036	0,057	0,083
		%10	0,047	0,064	0,041	0,086	0,062	0,069	0,045	0,053	0,053	0,126
		%20	0,075	0,177	0,050	0,136	0,074	0,154	0,062	0,130	0,052	0,132
	400/800	%5	0,038	0,045	0,033	0,042	0,053	0,053	0,044	0,043	0,039	0,048
		%10	0,048	0,064	0,030	0,063	0,051	0,066	0,043	0,053	0,033	0,084
		%20	0,058	0,165	0,041	0,123	0,073	0,150	0,057	0,126	0,038	0,119
2400	1200/1200	%5	0,053	0,056	0,039	0,070	0,057	0,058	0,050	0,053	0,040	0,156
		%10	0,056	0,094	0,035	0,110	0,062	0,087	0,051	0,076	0,045	0,151
		%20	0,117	0,333	0,077	0,280	0,100	0,274	0,108	0,286	0,078	0,251
	800/1600	%5	0,040	0,055	0,032	0,056	0,054	0,059	0,044	0,052	0,028	0,068
		%10	0,056	0,088	0,038	0,092	0,053	0,081	0,058	0,078	0,031	0,074
		%20	0,110	0,299	0,068	0,232	0,099	0,247	0,096	0,248	0,046	0,162

Tablo 11'de genel olarak beş yönteme ilişkin bulgular incelendiğinde; DMF büyüklüğü 0,5'den 1,0'e yükseldiğinde genel olarak tüm yöntemlerin I. Tip hata oranlarının yükselme eğiliminde olduğu görülmektedir. Odak ve referans grupların örneklem büyüklüğü oranları açısından incelendiğinde; tüm yöntemlerde örneklem büyüklüğü oranı eşit olduğu durumdaki I. Tip hata oranlarının, örneklem büyüklüğü 1/2 oranda olduğu duruma göre nispeten daha yüksek olduğu söylenebilir. Testte bulunan DMF'li madde oranı %5'ten %20'ye yükseldikçe beş yönteminde I. Tip hata oranlarının artma eğiliminde olduğu gözlenmektedir.

DMF belirleme yöntemlerine göre ayrıntılı olarak bulgular incelendiğinde ise; örneklem büyüklüğü 600-1200 ve DMF büyüklüğü 0,5 iken Mantel-Haenszel yönteminin Tip hata oranlarının 0,040-0,075 aralığında değerler aldığı görülmektedir. Bu değerler Bradley'in (1978) referansına göre incelendiğinde ise uygun hata sınırları içinde olduğu söylenebilir ($0,025 \leq I. \text{ Tip hata oranı} \leq ,075$). Fakat DMF büyüklüğü 0,5 iken örneklem büyüklüğü 2400'e yükseldiğinde ve testte bulunan DMF'li madde oranı %20'ye çıktığında Mantel-Haenszel yönteminin hata oranı yükselmektedir (M-H $I. \text{ Tip hata} >,075$). DMF büyüklüğü 0,5'ten 1,0'e yükseldiğinde 600 ve 1200 örneklem büyüklüklerinde Mantel-Haenszel yöntemi sadece %5 ve %10 DMF'li madde oranlarında uygun hata sınırlarındadır. Örneklem büyüklüğü 2400'e yükseldiğinde ise Mantel-Haenszel yönteminin I. Tip hata oranları sadece %5 DMF'li madde oranında uygun aralıkta iken %10 ve %20 DMF'li madde oranlarında hata oranı yükselmektedir (M-H $I. \text{ Tip hata} >,075$).

Lord'un χ^2 yöntemine ilişkin bulgular incelendiğinde, DMF büyüklüğü 0,5 iken; örneklem büyüklüğü 2400 ve örneklem oranı 1/1 koşulu dışındaki tüm koşullarda Tip hata oranlarının 0,030-0,680 aralığında değerler aldığı görülmektedir. Bu değerler Bradley'in (1978) referansına göre incelendiğinde ise uygun hata sınırları içinde olduğu söylenebilir ($0,025 \leq I. \text{ Tip hata oranı} \leq ,75$). DMF büyüklüğü 1,0'e yükseldiğinde ise 600 ve 1200 örneklem büyüklüğü koşulu altında; odak ve referans grupların örneklem sayısı birbirine eşit olduğunda sadece DMF'li madde oranı %5 iken uygun hata sınırlarındadır. Fakat aynı koşullar altında odak ve referans grup örneklem sayısı 1/2 oranında olduğunda ise DMF'li madde oranı %5 ve %10 iken uygun hata sınırlarındadır. Örneklem sayısı 2400'e yükseldiğinde ise odak ve referans grup örneklem büyüklüğü oranları farketmeksizin (hem 1/1 hem de 1/2 iken) DMF'li madde oranı %10 ve %20'ye çıktığında Lord'un χ^2 yönteminin hata oranı yükselmektedir (Lord'un $\chi^2 I. \text{ Tip hata} >,075$).

Lojistik Regresyon yöntemine ilişkin bulgular incelendiğinde; Mantel-Haenszel yöntemine benzer şekilde, örneklem büyüklüğü 600-1200 ve DMF büyüklüğü 0,5 iken Lojistik Regresyon yönteminin Tip hata oranlarının 0,051-0,075 aralığında değerler aldığı görülmektedir. Bu değerler Bradley'in (1978) referansına göre incelendiğinde ise uygun hata sınırları içinde olduğu söylenebilir ($0,025 \leq I. \text{ Tip hata oranı} \leq ,075$). Fakat DMF büyüklüğü 0,5 iken örneklem büyüklüğü 2400'e yükseldiğinde ve testte bulunan DMF'li madde oranı %20'ye çıktığında yine Mantel-Haenszel yöntemine benzer şekilde Lojistik Regresyon yönteminin hata oranı yükselmektedir (LR $I. \text{ Tip hata} >,075$). Yine Mantel-Haenszel yöntemine

benzer şekilde, DMF büyüklüğü 0,5'ten 1,0'e yükseldiğinde 600 ve 1200 örneklem büyüklüklerinde Lojistik Regresyon yönteminin I. Tip hata oranları sadece %5 ve %10 DMF oranlarında uygun aralıktadır. Örneklem büyüklüğü 2400'e yükseldiğinde ise Lojistik Regresyon yönteminin I. Tip hata oranları sadece %5 DMF'li madde oranlarında uygun aralıkta iken %10 ve %20 DMF'li madde oranlarında hata oranı yükselmektedir (LR I. Tip hata $>,075$).

SIBTEST yöntemine ilişkin bulgular incelendiğinde, Mantel-Haenszel ve Lojistik Regresyon yöntemine benzer şekilde, örneklem büyüklüğü 600-1200 ve DMF büyüklüğü 0,5 iken Lojistik Regresyon yönteminin Tip hata oranlarının 0,028-0,062 aralığında değerler aldığı görülmektedir. Bu değerler Bradley'in (1978) referansına göre incelendiğinde ise uygun hata sınırları içinde olduğu söylenebilir ($0,025 \leq \text{I. Tip hata oranı} \leq ,075$). Fakat DMF büyüklüğü 0,5 iken örneklem büyüklüğü 2400'e yükseldiğinde ve testte bulunan DMF'li madde oranı %20'ye çıktığında yine Mantel-Haenszel ve Lojistik Regresyon yöntemine benzer şekilde SIBTEST yönteminin hata oranları yükselmektedir (SIBTEST I. Tip hata $>,075$). DMF büyüklüğü 0,5'ten 1,0'e yükseldiğinde 600 örneklem büyüklüğünde %5, %10 ve %20 DMF'li madde oranlarında; 1200 örneklem büyüklüğünde %5 ve %10 DMF'li madde oranlarında; 2400 örneklem büyüklüğünde ise %5 DMF'li madde oranında hata oranları uygun aralıktadır. Dolayısıyla örneklem büyüklüğü 600'den 2400'e yükseldikçe ve DMF'li madde oranında %5'ten %20'ye yükseldikçe SIBTEST yönteminin I. Tip hata oranları artma eğilimindedir.

Raju'nun alan ölçümü yöntemine ilişkin bulgular incelendiğinde, DMF büyüklüğü 0,5 iken; örneklem büyüklüğü 2400 ve örneklem oranı 1/1 koşulu dışındaki tüm koşullarda Tip hata oranlarının 0,028-0,075 aralığında değerler aldığı görülmektedir. Bu değerler Bradley'in (1978) referansına göre incelendiğinde ise uygun hata sınırları içinde olduğu söylenebilir ($0,025 \leq \text{I. Tip hata oranı} \leq ,075$). DMF büyüklüğü 1,0'a yükseldiğinde ise örneklem oranı 1/1 koşulu altında üç farklı örneklem büyüklüğünde de hata oranları uygun aralıkta değildir (0,083-0,251). Fakat örneklem büyüklüğü oranı 1/2 koşulu altında üç farklı örneklem büyüklüğünde de DMF'li madde sayısı %5 iken hata oranları uygun aralıktadır.

4.1.1. Farklı Örneklem Büyüklüklerine Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının Kıyaslanması

Farklı örneklem büyüklüklerine (600, 1200 ve 2400) göre Mantel-Haenszel, Lord'un χ^2 , Lojistik Regresyon, SIBTEST ve Raju'nun alan ölçümü yöntemlerinin I. Tip hata oranları 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 12'de sunulmuştur.

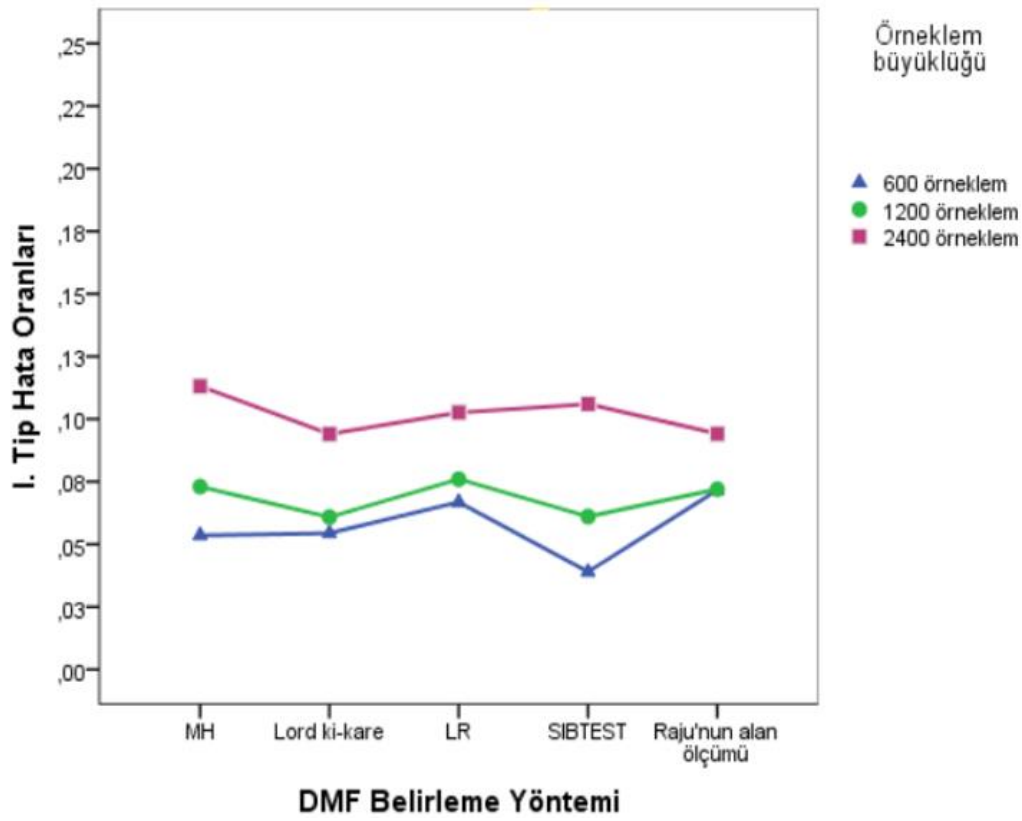
Tablo 12.

Farklı Örneklem Büyüklüklerine Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının İncelenmesi

	Örneklem büyüklüğü						Yöntem	Örneklem büyüklüğü	Etkileşim
	600		1200		2400				
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,054	0,020	0,073	0,047	0,113	0,098	F _{4;165} =0,480; p=,750	F _{2;165} =11,487; p= ,000 η^2 =0,122 <u>Anlamlı fark</u>	F _{8;165} =0,351; p= ,944
Lord χ^2	0,054	0,021	0,061	0,036	0,094	0,079			
LR	0,067	0,015	0,076	0,036	0,103	0,076			
SIBTEST	0,039	0,008	0,061	0,032	0,106	0,084			
Raju	0,072	0,023	0,072	0,037	0,094	0,070			
Anlamlı fark: 600-2400 (η^2 =0,131) ve 1200-2400 (η^2 =0,035)									

Tablo 12'de yer alan bulgular incelendiğinde, DMF belirleme yöntemlerine göre I. Tip hata oranları anlamlı farklılık göstermezken (F_{4;165}=0,480; p=,750); farklı örneklem büyüklüklerine göre anlamlı farklılık göstermektedir (F_{2;165}=11,487; p=,000). I. Tip hata oranlarındaki değişkenliğin yaklaşık %12'si örneklem büyüklüğü ile açıklanmakla birlikte, örneklem büyüklüklerinin I. Tip hata oranları üzerinde orta düzeyde etkiye sahip olduğu söylenebilir (η^2 =0,122). Yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın ise 600-2400 (η^2 =0,131) ve 1200-2400 (η^2 =0,035) örneklem büyüklükleri arasında olduğu saptanmıştır. Ayrıca DMF belirleme yöntemleri ile örneklem büyüklüklerine ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği tespit edilmiştir (F_{8;165}=0,351; p=,944).

Farklı örneklem büyüklüklerinde DMF belirleme yöntemlerinin I. Tip hata oranları Şekil 2'de sunulmuştur.



Şekil 2. Farklı örneklem büyüklüklerinde DMF belirleme yöntemlerinin I. Tip hata oranları

Şekil 2’de yer alan grafik incelendiğinde beş DMF belirleme yöntemine ilişkin I. Tip hata oranlarının örneklem büyüklüğü 600’den 2400’e yükseldikçe artma eğiliminde olduğu görülmektedir.

4.1.2. Farklı Örneklem Oranlarına Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının Kıyaslanması

Farklı örneklem oranlarına (1/1 ve 1/2) göre Mantel-Haenszel, Lord’un χ^2 , Lojistik Regresyon, SIBTEST ve Raju’nun alan ölçümü yöntemlerinin I. Tip hata oranları 5x2 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 13’te sunulmuştur.

Tablo 13.

Farklı Örneklem Oranlarına Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının İncelenmesi

	Örneklem oranı				Yöntem	Örneklem oranı	Etkileşim
	1/1		1/2				
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,083	0,071	0,077	0,064			
Lord χ^2	0,076	0,059	0,064	0,048	F _{4;170} =0,436;	F _{1;170} =1,944;	F _{4;170} =0,336;
LR	0,084	0,054	0,079	0,049	p= ,782	p= ,165	p= ,853
SIBTEST	0,071	0,066	0,066	0,051			
Raju	0,094	0,054	0,064	0,034			

Tablo 13'e göre I. Tip hata oranlarının hem DMF belirleme yöntemlerine hem de farklı örneklem oranlarına göre anlamlı farklılık göstermediği tespit edilmiştir (F_{4;170}=0,436; p=,782 ve F_{1;170}=1,944; p=,165). Ayrıca DMF belirleme yöntemleri ile örneklem oranlarına ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği saptanmıştır (F_{4;170}=0,336; p= ,853).

4.1.3. Farklı DMF Büyüklüklerine Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının Kıyaslanması

Farklı DMF büyüklüklerine (0,5 ve 1,0) göre Mantel-Haenszel, Lord'un χ^2 , Lojistik Regresyon, SIBTEST ve Raju'nun alan ölçümü yöntemlerinin I. Tip hata oranları 5x2 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 14'te sunulmuştur.

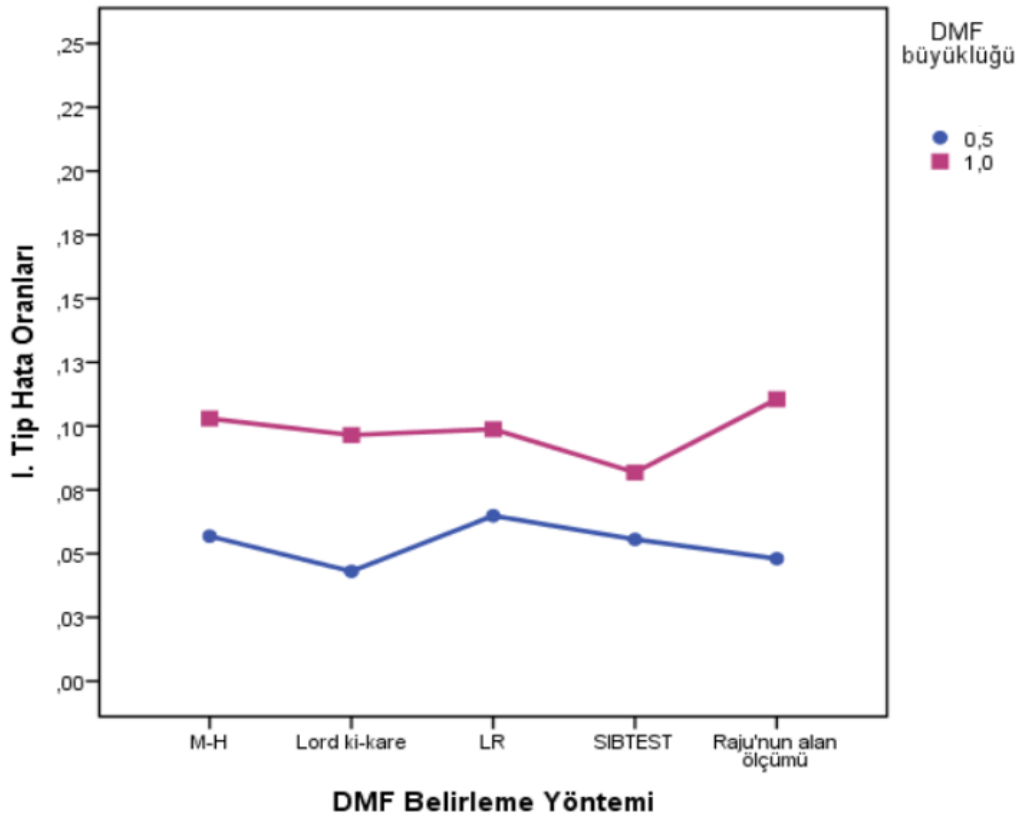
Tablo 14.

Farklı DMF Büyüklüklerine Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının İncelenmesi

	DMF büyüklüğü				Yöntem	DMF büyüklüğü	Etkileşim
	0,5		1,0				
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,057	0,023	0,103	0,087	F _{4;170} =0,521; p= ,720	F _{1;170} =33,983; p= ,000 η^2 =0,167	F _{4;170} =0,732; p= ,571
Lord χ^2	0,043	0,012	0,097	0,065			
LR	0,065	0,015	0,099	0,066			
SIBTEST	0,056	0,034	0,082	0,073			
Raju	0,048	0,013	0,111	0,048			

Tablo 14’te yer alan bulgular incelendiğinde, DMF belirleme yöntemlerine göre I. Tip hata oranları anlamlı farklılık göstermezken ($F_{4;170}=0,521$; $p=,720$); DMF büyüklüklerine göre anlamlı farklılık göstermektedir ($F_{1;170}=33,983$; $p=,000$). I. Tip hata oranlarındaki değişkenliğin yaklaşık %16,7’si DMF büyüklüğü ile açıklanmakla birlikte, DMF büyüklüklerinin I. Tip hata oranları üzerinde yüksek etkiye sahip olduğu söylenebilir ($\eta^2=0,167$). Ayrıca DMF belirleme yöntemleri ile DMF büyüklüklerine ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği tespit edilmiştir ($F_{4;170}=0,732$; $p=,571$).

Farklı DMF büyüklüklerinde DMF belirleme yöntemlerinin I. Tip hata oranları Şekil 3’te sunulmuştur.



Şekil 3. Farklı DMF büyüklüklerinde DMF belirleme yöntemlerinin I. Tip hata oranları

Şekil 3’te yer alan grafik incelendiğinde DMF büyüklüğü 0,5’ten 1,0’e yükseldiğinde beş DMF belirleme yöntemine ilişkin I. Tip hata oranlarının artma eğiliminde olduğu görülmektedir.

4.1.4. Farklı DMF’li Madde Oranlarına Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının Kıyaslanması

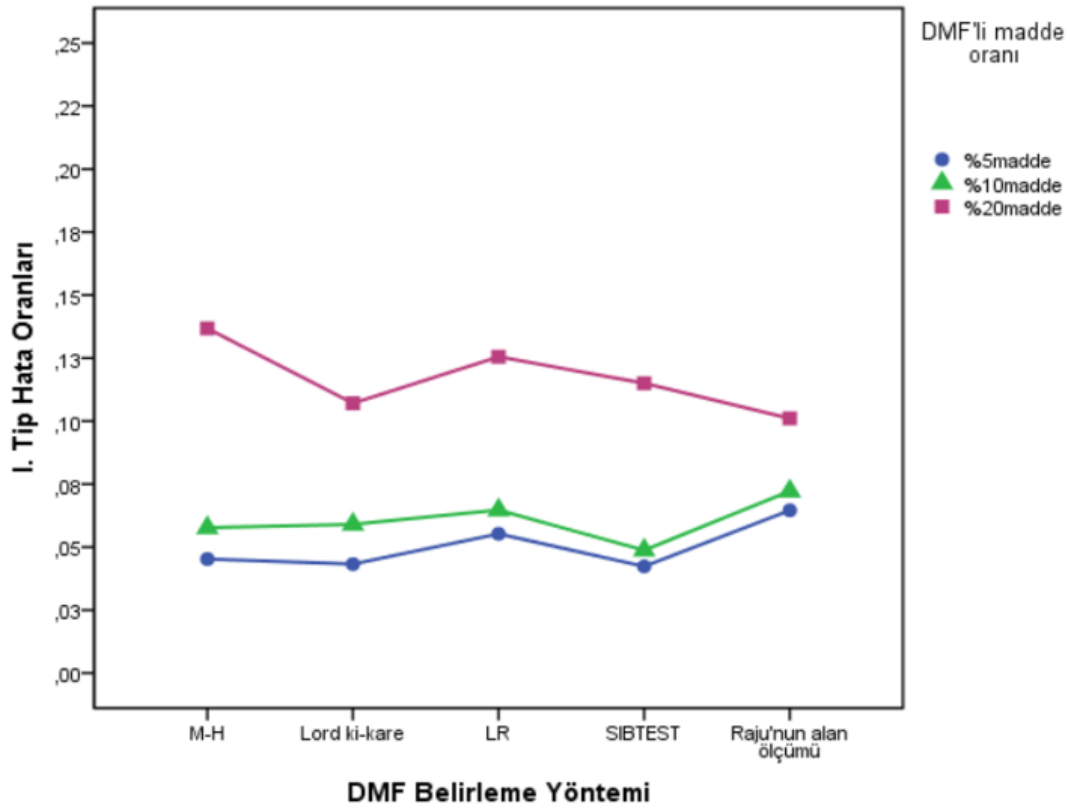
Farklı DMF’li madde oranlarına (%5, %10 ve %20) göre Mantel-Haenszel, Lord’un χ^2 , Lojistik Regresyon, SIBTEST ve Raju’nun alan ölçümü yöntemlerinin I. Tip hata oranları 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 15’te sunulmuştur. Tablo 15.

Farklı DMF’li Madde Oranlarına Göre DMF Belirleme Yöntemlerinin I. Tip Hata Oranlarının İncelenmesi

	DMF’li madde oranları						Yöntem	DMF’li madde oranları	Etkileşim
	%5		%10		%20				
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,045	0,006	0,058	0,017	0,137	0,093		$F_{2;165}=34,445;$	
Lord χ^2	0,043	0,011	0,059	0,026	0,107	0,077	$F_{4;165}=0,603;$	$p=,000$	$F_{8;165}=0,684;$
LR	0,055	0,002	0,065	0,010	0,126	0,069	$p=,661$	$\eta^2=0,295$	$p=,705$
SIBTEST	0,042	0,007	0,049	0,016	0,115	0,083		<u>Anlamlı fark</u>	
Raju	0,065	0,034	0,072	0,037	0,101	0,061			
Anlamlı fark: %5-%20 ($\eta^2=0,273$) ve %10-%20 ($\eta^2=0,206$)									

Tablo 15’te yer alan bulgular incelendiğinde, DMF belirleme yöntemlerine göre I. Tip hata oranları anlamlı farklılık göstermezken ($F_{4;165}=0,603;$ $p=,661$); DMF’li madde oranlarına göre anlamlı farklılık göstermektedir ($F_{2;165}=34,445;$ $p=,000$). I. Tip hata oranlarındaki değişkenliğin yaklaşık %29,5’i testte bulunan DMF’li madde oranları ile açıklanmakla birlikte, DMF’li madde oranlarının I. Tip hata oranları üzerinde yüksek etkiye sahip olduğu söylenebilir ($\eta^2=0,295$). Yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın ise %5-%20 ($\eta^2=0,273$) ve %10-%20 ($\eta^2=0,206$) DMF’li madde oranları arasında olduğu saptanmıştır. Ayrıca DMF belirleme yöntemleri ile DMF’li madde oranlarına ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği tespit edilmiştir ($F_{8;165}=0,684;$ $p=,705$).

Farklı DMF’li madde oranlarında DMF belirleme yöntemlerinin I. Tip hata oranları Şekil 4’te sunulmuştur.



Şekil 4. Farklı DMF’li madde oranlarında DMF belirleme yöntemlerinin I. Tip hata oranları

Şekil 4’te yer alan grafik incelendiğinde testte bulunan DMF’li madde oranları %5’ten %20’ye yükseldiğinde beş DMF belirleme yöntemine ilişkin I. Tip hata oranlarının artma eğiliminde olduğu görülmektedir.

4.2. Farklı Koşullar Altında DMF Belirleme Yöntemlerinin Güç Oranları

“DMF belirleme yöntemlerinin güç oranları örneklem büyüklükleri (600, 1200 ve 2400), odak ve referans grup örneklem büyüklüğü oranları (1/1 ve 1/2), DMF büyüklükleri (0,5 ve 1,0) ve testte bulunan DMF’li maddelerin oranları (%5, %10 ve %20) koşullarında nasıl değişmektedir?” araştırma sorusuna yönelik elde edilen bulgular aşağıda verilmiştir.

DMF belirleme yöntemlerinin ilgili koşullar altında güç oranları Tablo 16’da sunulmuştur.

Tablo 16.

İlgili Koşullar Altında DMF Belirleme Yöntemlerinin Güç Oranları

Koşullar			DMF belirleme yöntemleri									
			M-H		Lord'un χ^2		LR		SIBTEST		Raju'nun alan ölçümü	
			DMF büyüklüğü		DMF büyüklüğü		DMF büyüklüğü		DMF büyüklüğü		DMF büyüklüğü	
Örneklem büyüklüğü	Örneklem oranı	DMF'li madde oranı	0,5	1,0	0,5	1,0	0,5	1,0	0,5	1,0	0,5	1,0
600	300/300	%5	0,32	0,69	0,13	0,50	0,31	0,70	0,13	0,49	0,19	0,47
		%10	0,35	0,78	0,17	0,57	0,35	0,81	0,17	0,49	0,20	0,52
		%20	0,37	0,83	0,16	0,64	0,35	0,84	0,21	0,59	0,23	0,64
	200/400	%5	0,39	0,69	0,17	0,48	0,35	0,70	0,11	0,39	0,24	0,48
		%10	0,31	0,77	0,13	0,56	0,32	0,78	0,12	0,42	0,16	0,53
		%20	0,30	0,80	0,15	0,56	0,31	0,80	0,13	0,40	0,19	0,60
1200	600/600	%5	0,62	0,84	0,34	0,66	0,62	0,84	0,50	0,61	0,37	0,48
		%10	0,65	0,92	0,40	0,79	0,61	0,92	0,55	0,79	0,36	0,65
		%20	0,63	0,93	0,39	0,84	0,58	0,93	0,53	0,88	0,46	0,75
	400/800	%5	0,55	0,87	0,30	0,63	0,51	0,88	0,45	0,65	0,31	0,51
		%10	0,60	0,92	0,33	0,77	0,57	0,92	0,48	0,82	0,34	0,64
		%20	0,55	0,93	0,31	0,83	0,51	0,94	0,46	0,89	0,34	0,75
2400	1200/1200	%5	0,82	0,99	0,60	0,85	0,79	0,99	0,74	0,92	0,50	0,44
		%10	0,87	0,99	0,67	0,92	0,85	0,99	0,81	0,96	0,57	0,69
		%20	0,86	0,99	0,65	0,93	0,83	0,99	0,83	0,97	0,66	0,77
	800/1600	%5	0,75	0,99	0,62	0,89	0,75	0,97	0,71	0,95	0,51	0,54
		%10	0,83	0,98	0,68	0,92	0,82	0,98	0,83	0,97	0,58	0,74
		%20	0,83	0,98	0,61	0,93	0,80	0,98	0,82	0,97	0,65	0,80

Tablo 16’da genel olarak beş yönteme ilişkin bulgular incelendiğinde; DMF büyüklüğü 0,5, örneklem büyüklüğü 600 ve 1200 iken beş yöntemde güç oranlarının 0,11-0,65 aralığında değerler aldığı, dolayısıyla yeterli güce sahip olmadıkları görülmektedir. Örneklem büyüklüğü 2400, DMF büyüklüğü 1,0 olduğunda Raju’nun alan ölçümü yöntemi dışındaki dört yöntemin güç oranlarının 0,85-0,99 aralığında değer aldığı ve yeterli güce ulaştıkları söylenebilir. Dört yöntemden farklı olarak Raju’nun alan ölçümü yönteminin sadece 1,0 DMF büyüklüğü, 2400 örneklem büyüklüğü, 1/2 örneklem büyüklüğü oranı ve %20 DMF’li madde oranı koşulunda yeterli güce ulaştığı görülmektedir.

DMF belirleme yöntemlerine göre ayrıntılı olarak bulgular incelendiğinde ise; 0,5 DMF büyüklüğünde Mantel-Haenszel yönteminin sadece 2400 örneklem büyüklüğünde yeterli güce sahip olduğu görülmektedir. DMF büyüklüğü 1,0'e yükseldiğinde ise Mantel-Haenszel yöntemi 600 örneklem büyüklüğünde sadece %20 DMF'li madde oranında; 1200 ve 2400 örneklem büyüklüklerinde ise tüm koşullarda yeterli güce sahip olduğu belirlenmiştir.

Lord'un χ^2 yöntemine ilişkin bulgular incelendiğinde, DMF büyüklüğünde 0,5 iken tüm koşullarda yöntemin güç oranlarının 0,13-0,68 aralığında değerler aldığı; örneklem büyüklüğü 600 iken ise tüm koşullarda 0,13-0,64 aralığında değerler aldığı ve dolayısıyla yeterli güce sahip olmadığı görülmektedir. Örneklem büyüklüğü 1200'e yükseldiğinde ise sadece DMF büyüklüğü 1,0 ve DMF'li madde oranı %20 olduğunda yeterli güce sahip olduğu söylenebilir. Örneklem büyüklüğü 2400'e yükseldiğinde örneklem oranı ve DMF'li madde oranı farketmeksizin Lord'un χ^2 yönteminin sadece DMF büyüklüğü 1.0 olduğunda yeterli güce sahip olduğu söylenebilir.

Raju'nun alan ölçümü yönteminde Lord'un χ^2 yöntemine benzer şekilde DMF büyüklüğünde 0,5 iken tüm koşullarda yöntemin güç oranlarının 0,16-0,66 aralığında değerler aldığı dolayısıyla yeterli güce sahip olmadığı söylenebilir. Raju'nun alan ölçümü yönteminin güç oranının sadece örneklem büyüklüğü 2400, örneklem büyüklüğü oranı 1/2 ve DMF'li madde oranı ise %20 olduğunda 0,80 olduğu belirlenmiştir.

Lojistik Regresyon yöntemine ilişkin bulgular incelendiğinde, DMF büyüklüğü 0,5, örneklem büyüklüğü ise 600 ve 1200 iken güç oranlarının 0,31-0,62 aralığında değerler aldığı dolayısıyla yeterli güce sahip olmadığı görülmektedir. DMF büyüklüğü 0,5 iken örneklem büyüklüğü 2400'e yükseldiğinde ve örneklem büyüklüğü oranı farketmeksizin DMF'li madde oranı %10 ve %20 iken yeterli güce ulaştığı söylenebilir. DMF büyüklüğü 1,0'a yükseldiğinde ise sadece 1200 ve 2400 örneklem büyüklüklerinde yeterli güce sahip olduğu görülmektedir.

SIBTEST yöntemine ilişkin bulgular incelendiğinde, Lojistik Regresyon yöntemine benzer şekilde DMF büyüklüğü 0,5, örneklem büyüklüğü ise 600 ve 1200 iken güç oranlarının 0,11-0,55 aralığında değerler aldığı dolayısıyla yeterli güce sahip olmadığı söylenebilir. Yine Lojistik Regresyon yöntemine benzer şekilde DMF büyüklüğü 0,5 iken örneklem büyüklüğü 2400'e yükseldiğinde ise, örneklem büyüklüğü oranı farketmeksizin, DMF'li madde oranı %10 ve %20 iken yeterli güce ulaştığı söylenebilir. DMF büyüklüğü 1,0'e yükseldiğinde ise örneklem büyüklüğü 600 iken yeterli güce sahip olmadığı belirlenmiştir. DMF büyüklüğü

1,0 koşulu altında; örneklem büyüklüğü 1200 iken %10 ve %20 DMF oranlarında; örneklem büyüklüğü 2400 iken ise tüm koşullarda yeterli güce ulaştığı görülmektedir.

Ayrıca bütün bu açıklamalara ek olarak, DMF büyüklüğü 0,5'ten 1,0'a yükseldiğinde beş yöntemde güç oranlarının genel olarak yükselme eğiliminde olduğu görülmektedir. Benzer şekilde örneklem büyüklüğü 600'den 2400'e arttığında yöntemlerin güç oranlarının yükseldiği söylenebilir. Odak ve referans grupların örneklem büyüklüğü oranı açısından incelendiğinde; bütün yöntemlerde grupların örneklem sayısı eşit olduğu (1/1) durumdaki güç oranlarının, örneklem büyüklüğü 1/2 oranda olduğu duruma göre nispeten daha yüksek olduğu görülmektedir.

4.2.1. Farklı Örneklem Büyüklüklerine Göre DMF Belirleme Yöntemlerinin Güç Oranlarının Kıyaslanması

Farklı örneklem büyüklüklerine (600, 1200 ve 2400) göre Mantel-Haenszel, Lord'un χ^2 , Lojistik Regresyon, SIBTEST ve Raju'nun alan ölçümü yöntemlerinin güç oranları 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 17'de sunulmuştur.

Tablo 17.

Farklı Örneklem Büyüklüklerine Göre DMF Belirleme Yöntemlerinin Güç Oranlarının İncelenmesi

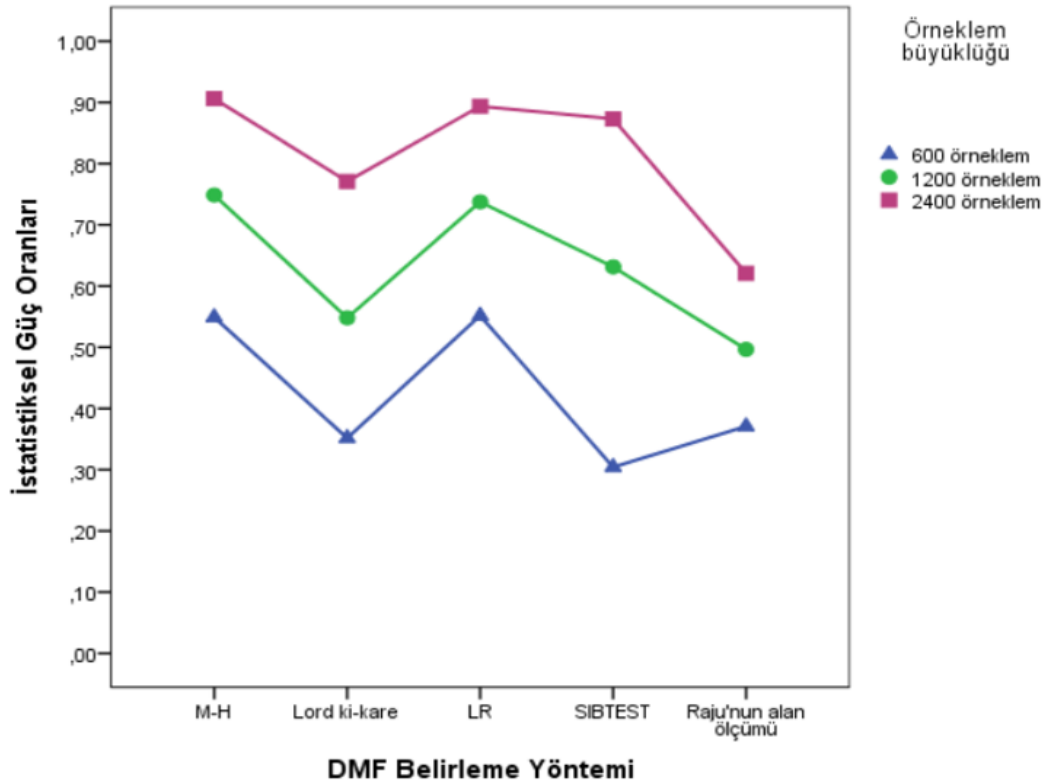
	Örneklem						Yöntem	Örneklem büyüklüğü	Etkileşim
	600		1200		2400				
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,549	0,225	0,749	0,163	0,906	0,089	$F_{4;165}=13,607$;	$F_{2;165}=76,946$;	
Lord	0,352	0,213	0,548	0,224	0,771	0,143	$p=,000$	$p=,000$	$F_{8;165}=1,488$;
LR	0,552	0,234	0,738	0,181	0,894	0,096	$\eta^2=0,248$	$\eta^2=0,483$	$p=,165$
SIBTEST	0,304	0,176	0,631	0,169	0,873	0,095	<u>*Anlamlı fark</u>	<u>**Anlamlı fark</u>	
Raju	0,371	0,184	0,497	0,163	0,621	0,115			
<u>*Anlamlı fark</u> : M-H ve Lord'un χ^2 ($\eta^2=0,124$); M-H ve SIBTEST ($\eta^2=0,066$); M-H ve Raju'nun alan ölçümü ($\eta^2=0,262$); Lord'un χ^2 ve LR ($\eta^2=0,114$); LR ve SIBTEST ($\eta^2=0,059$); LR ve Raju'nun alan ölçümü ($\eta^2=0,246$)									
<u>**Anlamlı fark</u> : 600-1200 ($\eta^2=0,191$); 600-2400 ($\eta^2=0,507$) ve 1200-2400 ($\eta^2=0,206$)									

Tablo 17'de yer alan bulgular incelendiğinde, hem DMF belirleme yöntemlerine hem de örneklem büyüklüklerine göre güç oranları anlamlı farklılık göstermektedir ($F_{4;165}=13,607$; $p=,000$ ve $F_{2;165}=76,946$; $p=,000$). Güç oranlarındaki değişkenliğin yaklaşık %24,8'i DMF belirleme yöntemleri ile açıklanmakla birlikte yöntemlerin güç oranları üzerinde yüksek

etkiye sahip olduğu söylenebilir ($\eta^2=0,248$). Benzer şekilde güç oranlarındaki değişkenliğin yaklaşık %48,3'ü ise örneklem büyüklükleri ile açıklanmakla birlikte örneklem büyüklüklerinin güç oranları üzerinde yüksek etkiye sahip olduğu söylenebilir ($\eta^2=0,483$). Ayrıca DMF belirleme yöntemleri ile örneklem büyüklüklerine ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği tespit edilmiştir ($F_{8;165}=1,488$; $p=,165$).

DMF belirleme yöntemleri üzerinden yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın ise Mantel-Haenszel ve Lord'un χ^2 ($\eta^2=0,124$); Mantel-Haenszel ve SIBTEST ($\eta^2=0,066$); Mantel-Haenszel ve Raju'nun alan ölçümü ($\eta^2=0,262$); Lord'un χ^2 ve Lojistik Regresyon ($\eta^2=0,114$); Lojistik Regresyon ve SIBTEST ($\eta^2=0,059$); Lojistik Regresyon ve Raju'nun alan ölçümü ($\eta^2=0,246$) yöntemleri arasında olduğu saptanmıştır. Örneklem büyüklükleri üzerinden yapılan ikili karşılaştırmalar sonucunda ise 600-1200 ($\eta^2=0,191$); 600-2400 ($\eta^2=0,507$) ve 1200-2400 ($\eta^2=0,206$) arasında olduğu tespit edilmiştir.

Farklı örneklem büyüklüklerinde DMF belirleme yöntemlerinin güç oranları Şekil 5'te sunulmuştur.



Şekil 5. Farklı örneklem büyüklüklerinde DMF belirleme yöntemlerinin güç oranları

Şekil 5’te yer alan grafik incelendiğinde beş DMF belirleme yöntemine ilişkin güç oranlarının örneklem büyüklüğü 600’den 2400’e yükseldikçe artma eğiliminde olduğu görülmektedir. Üç farklı örneklem büyüklüğünde de en yüksek güce sahip olan yöntemlerin Mantel-Haenszel ve Lojistik Regresyon; Raju’nun alan ölçümü yönteminin ise genel olarak düşük güce sahip olduğu söylenebilir.

4.2.2. Farklı Örneklem Büyüklüğü Oranlarına Göre DMF Belirleme Yöntemlerinin Güç Oranlarının Kıyaslanması

Farklı örneklem büyüklüğü oranlarına (1/1 ve 1/2) göre Mantel-Haenszel, Lord’un χ^2 , Lojistik Regresyon, SIBTEST ve Raju’nun alan ölçümü yöntemlerinin güç oranları 5x2 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 18’de sunulmuştur.

Tablo 18.

Farklı Örneklem Büyüklüğü Oranlarına Göre DMF Belirleme Yöntemlerinin Güç Oranlarının İncelenmesi

	Örneklem oranı				Yöntem	Örneklem oranı	Etkileşim
	1/1		1/2				
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,747	0,219	0,723	0,228			
Lord	0,566	0,256	0,548	0,267	$F_{4;170}=7,010$;	$F_{1;170}=0,335$;	$F_{4;170}=0,022$;
LR	0,740	0,223	0,715	0,232	p= ,335	p= ,564	p= ,999
SIBTEST	0,619	0,265	0,586	0,298			
Raju	0,497	0,179	0,495	0,194			

Tablo 18’de yer alan bulgular incelendiğinde, güç oranlarının hem DMF belirleme yöntemlerine hem de farklı örneklem büyüklüğü oranlarına göre anlamlı farklılık göstermediği tespit edilmiştir ($F_{4;170}=7,010$; $p=,335$ ve $F_{1;170}=0,335$; $p=,564$). Ayrıca DMF belirleme yöntemleri ile örneklem oranlarına ilişkin ortak etkinin de istatistiksel olarak anlamlı farklılık göstermediği saptanmıştır ($F_{4;170}=0,022$; $p=,999$).

4.2.3. Farklı DMF Büyüklüklerine Göre DMF Belirleme Yöntemlerinin Güç Oranlarının Kıyaslanması

Farklı DMF büyüklüklerine (0,5 ve 1,0) göre Mantel-Haenszel, Lord'un χ^2 , Lojistik Regresyon, SIBTEST ve Raju'nun alan ölçümü yöntemlerinin güç oranları 5x2 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 19'da sunulmuştur.

Tablo 19.

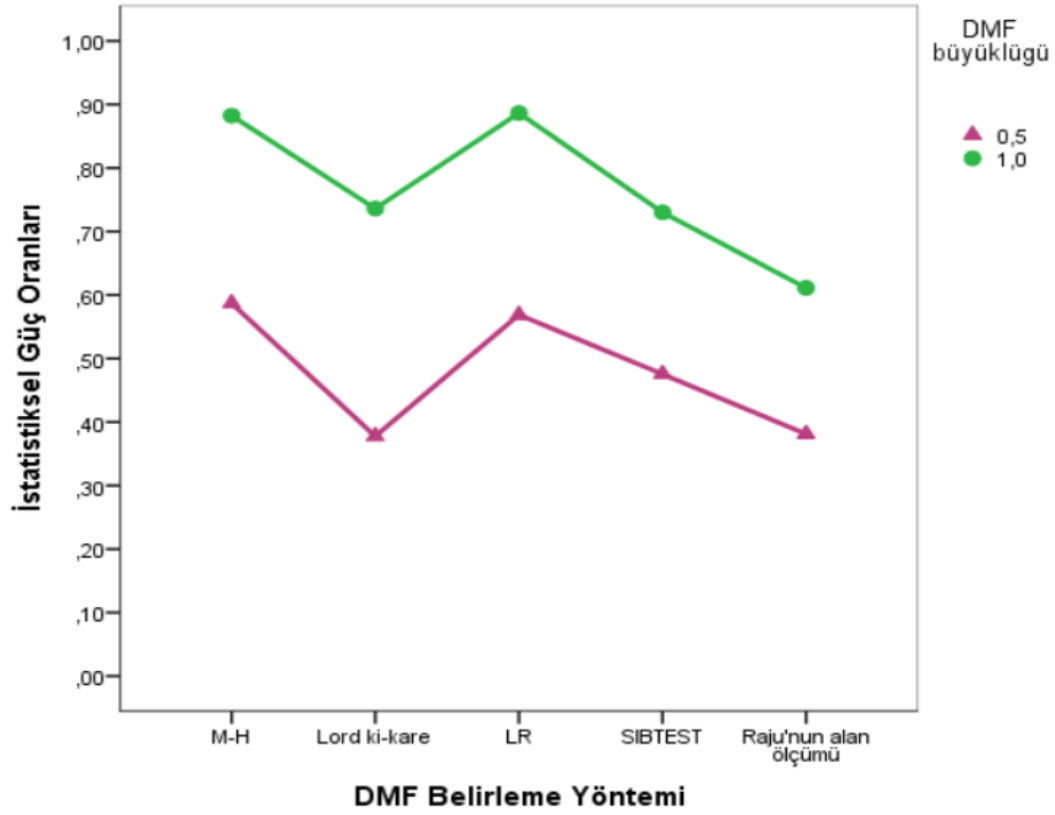
Farklı DMF Büyüklüklerine Göre DMF Belirleme Yöntemlerinin Güç Oranlarının İncelenmesi

Yöntem	DMF büyüklüğü				Yöntem	DMF büyüklüğü	Etkileşim
	0,5		1,0				
	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,587	0,208	0,882	0,104	F _{4;170} =11,721; p= ,000 η^2 =0,216 <u>*Anlamlı fark</u>	F _{1;170} =112,232; p= ,000 η^2 =0,398	F _{4;170} =0,683; p= ,604
Lord	0,378	0,208	0,736	0,160			
LR	0,569	0,202	0,887	0,097			
SIBTEST	0,476	0,275	0,730	0,223			
Raju	0,381	0,166	0,611	0,119			
<u>*Anlamlı fark:</u> M-H ve Lord'un χ^2 (η^2 =0,124); M-H ve SIBTEST (η^2 =0,066); M-H ve Raju'nun alan ölçümü (η^2 =0,262 Lord'un χ^2 ve LR (η^2 =0,114); LR ve SIBTEST (η^2 =0,059); LR ve Raju'nun alan ölçümü (η^2 =0,246); SIBTEST ve Raju'nun alan ölçümü (η^2 =0,050)							

Tablo 19'da yer alan bulgular incelendiğinde, hem DMF belirleme yöntemlerine hem de DMF büyüklüklerine göre güç oranları anlamlı farklılık göstermektedir (F_{4;170}=11,721; p=,000 ve F_{1;170}=112,232; p=,000). Güç oranlarındaki değişkenliğin yaklaşık %21,6'sı DMF belirleme yöntemleri ile açıklanmakla birlikte yöntemlerin güç oranları üzerinde yüksek etkiye sahip olduğu söylenebilir (η^2 =0,216). Benzer şekilde güç oranlarındaki değişkenliğin yaklaşık %39,8'i ise DMF büyüklükleri ile açıklanmakla birlikte DMF büyüklüklerinin de güç oranları üzerinde yüksek etkiye sahip olduğu söylenebilir (η^2 =0,398). Ayrıca DMF belirleme yöntemleri ile DMF büyüklüklerine ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği tespit edilmiştir (F_{4;170}=0,683; p=,604).

Yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın ise Mantel-Haenszel ve Lord'un χ^2 (η^2 =0,124); Mantel-Haenszel ve SIBTEST (η^2 =0,066); Mantel-Haenszel ve Raju'nun alan ölçümü (η^2 =0,262); Lord'un χ^2 ve Lojistik Regresyon (η^2 =0,114); Lojistik Regresyon ve SIBTEST (η^2 =0,059); Lojistik Regresyon ve Raju'nun alan ölçümü (η^2 =0,246); SIBTEST ve Raju'nun alan ölçümü (η^2 =0,050) yöntemleri arasında olduğu saptanmıştır.

Farklı DMF büyüklüklerinde DMF belirleme yöntemlerinin güç oranları Şekil 6’da sunulmuştur.



Şekil 6. Farklı DMF büyüklüklerinde DMF belirleme yöntemlerinin güç oranları

Şekil 6’da yer alan grafik incelendiğinde ise DMF büyüklüğü 0,5’ten 1,0’e yükseldiğinde beş DMF belirleme yöntemine ilişkin güç oranlarının artma eğiliminde olduğu görülmektedir. İki farklı DMF büyüklüğünde de gücü en düşük olan Lord’un χ^2 ve Raju’nun alan ölçümü yöntemleri; en yüksek olan ise Mantel-Haenszel ve Lojistik Regresyon yöntemleridir.

4.2.4. Farklı DMF’li Madde Oranlarına Göre DMF Belirleme Yöntemlerinin Güç Oranlarının Kıyaslanması

Farklı DMF’li madde oranlarına (%5, %10 ve %20) göre Mantel-Haenszel, Lord’un χ^2 , Lojistik Regresyon, SIBTEST ve Raju’nun alan ölçümü yöntemlerinin güç oranları 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 20’de sunulmuştur.

Tablo 20.

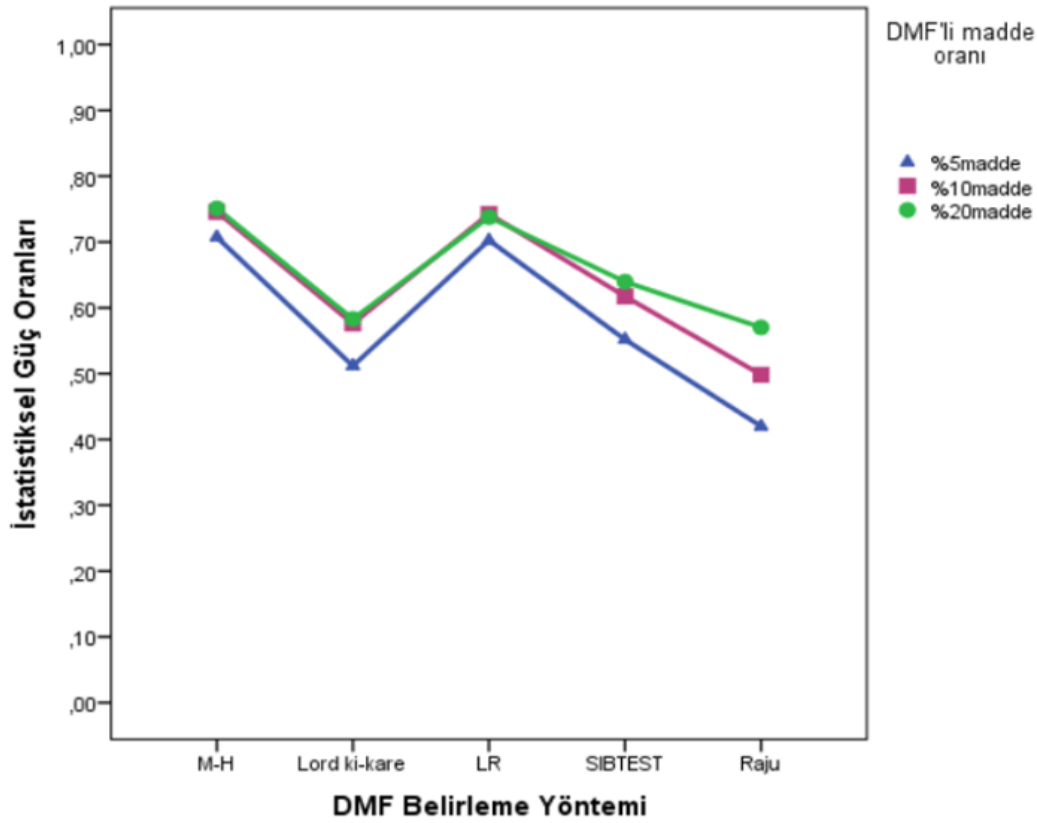
Farklı DMF'li Madde Oranlarına Göre DMF Belirleme Yöntemlerinin Güç Oranlarının İncelenmesi

	DMF'li madde oranları							DMF'li madde oranları	Etkileşim
	%5		%10		%20		Yöntem		
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,708	0,213	0,746	0,230	0,751	0,236	$F_{4;165}=6,965;$		
Lord	0,512	0,242	0,576	0,268	0,583	0,280	$p=,000$	$F_{2;165}=1,696;$	$F_{8;165}=0,116;$
LR	0,703	0,221	0,743	0,230	0,738	0,240	$\eta^2=0,144$	$p=,187$	$p=,999$
SIBTEST	0,552	0,266	0,617	0,290	0,639	0,295	<u>*Anlamlı fark</u>		
Raju	0,420	0,116	0,498	0,191	0,570	0,214			

*Anlamli fark: M-H ve Lord'un χ^2 ($\eta^2=0,124$); M-H ve SIBTEST ($\eta^2=0,066$); M-H ve Raju'nun alan ölçümü ($\eta^2=0,262$); Lord'un χ^2 ve LR ($\eta^2=0,114$); LR ve SIBTEST ($\eta^2=0,059$); LR ve Raju'nun alan ölçümü ($\eta^2=0,246$)

Tablo 20'de yer alan bulgular incelendiğinde, testte bulunan DMF'li madde oranlarına göre güç oranları anlamlı farklılık göstermezken ($F_{2;165}=1,696$; $p=,187$); DMF belirleme yöntemlerine ($F_{4;165}=6,965$; $p=,000$) göre anlamlı farklılık göstermektedir. Güç oranlarındaki değişkenliğin yaklaşık %14,4'ü DMF belirleme yöntemleri ile açıklanmakla birlikte, yöntemlerin güç oranları üzerinde yüksek etkiye sahip olduğu söylenebilir ($\eta^2=0,144$). Yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın ise Mantel-Haenszel ve Lord'un χ^2 ($\eta^2=0,124$); Mantel-Haenszel ve SIBTEST ($\eta^2=0,066$); Mantel-Haenszel ve Raju'nun alan ölçümü ($\eta^2=0,262$); Lord'un χ^2 ve Lojistik Regresyon ($\eta^2=0,114$); Lojistik Regresyon ve SIBTEST ($\eta^2=0,059$); Lojistik Regresyon ve Raju'nun alan ölçümü ($\eta^2=0,246$) yöntemleri arasında olduğu saptanmıştır. Ayrıca DMF belirleme yöntemleri ile DMF'li madde oranlarına ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği tespit edilmiştir ($F_{8;165}=0,116$; $p=,999$).

Farklı DMF'li madde oranlarında DMF belirleme yöntemlerinin güç oranları Şekil 7'de sunulmuştur.



Şekil 7. Farklı DMF’li madde oranlarında DMF belirleme yöntemlerinin güç oranları

Şekil 7’de yer alan grafik incelendiğinde ise testte bulunan DMF’li madde oranı %5’ten %10’a yükseldiğinde beş DMF belirleme yöntemine ilişkin güç oranlarının artma eğiliminde olduğu görülmektedir. Fakat DMF’li madde oranlarının %10’dan %20’ye yükselmesi Raju’nun alan ölçümü dışındaki yöntemlerin güç oranlarında nispeten farklılığa sebep olmadığı gözlenmektedir. Ayrıca DMF’li madde oranları %5 ve %10 iken güç oranı en düşük olan Raju’nun alan ölçümü yöntemi; en yüksek olanlar ise Mantel-Haenszel ve Lojistik Regresyon yöntemleridir.

4.3. Farklı Koşullar Altında DMF’li Maddelerin Testten Çıkarılmasında Madde Parametre Kestirimleri

“DMF belirleme yöntemlerine göre madde parametre kestirimlerinin RMSE değerleri; örneklem büyüklükleri (600, 1200 ve 2400), odak ve referans grup örneklem büyüklüğü oranları (1/1 ve 1/2), DMF büyüklükleri (0,5 ve 1,0) ve testte bulunan DMF’li maddelerin oranları (%5, %10 ve %20) koşullarında nasıl değişmektedir?” araştırma sorusuna yönelik elde edilen bulgular aşağıda verilmiştir.

4.3.1. Ayırt Edicilik Parametresine İlişkin RMSE Değerlerinin İncelenmesi

DMF belirleme yöntemlerine göre elde edilen ayırt edicilik parametre kestirimine ait RMSE değerleri; örneklem büyüklükleri, odak ve referans grup örneklem büyüklüğü oranları, DMF büyüklükleri ve testte bulunan DMF’li maddelerin oranları koşulları altında incelenmiştir. İlgili koşullar altında ayırt edicilik parametresine ilişkin RMSE değerleri Tablo 21’de sunulmuştur.

Tablo 21.

İlgili Koşullar Altında Ayırt Edicilik Parametresine İlişkin RMSE Değerleri

Koşullar				DMF belirleme yöntemleri				
Örnekleme	Örnekleme oranı	DMF büyüklüğü	DMF'li madde oranı	M-H	Lord'un χ^2	LR	SIBTEST	Raju'nun alan ölçümü
				RMSE	RMSE	RMSE	RMSE	RMSE
600	300/300	0,5	% 5	0,184098	0,181013	0,180686	0,185323	0,184342
			% 10	0,198379	0,193345	0,193474	0,194207	0,192843
			% 20	0,191176	0,189269	0,190769	0,188800	0,188445
		1,0	% 5	0,201705	0,198141	0,200853	0,197586	0,199402
			% 10	0,191411	0,189033	0,187577	0,190537	0,188361
			% 20	0,197419	0,195057	0,200774	0,201071	0,198719
	200/400	0,5	% 5	0,184365	0,184265	0,184178	0,185144	0,185459
			% 10	0,192320	0,195809	0,195101	0,193601	0,195809
			% 20	0,190798	0,192361	0,192361	0,193100	0,193199
		1,0	% 5	0,180323	0,182455	0,184660	0,180574	0,191378
			% 10	0,196737	0,193976	0,196737	0,194307	0,196537
			% 20	0,199018	0,197085	0,202802	0,192335	0,200009
1200	600/600	0,5	% 5	0,131871	0,131704	0,134623	0,132945	0,132945
			% 10	0,131926	0,130309	0,131552	0,129647	0,131670
			% 20	0,140438	0,137106	0,133504	0,137106	0,137106
		1,0	% 5	0,129773	0,134807	0,130075	0,136179	0,136179
			% 10	0,128296	0,132399	0,128618	0,128500	0,154995
			% 20	0,136509	0,141395	0,133878	0,136323	0,140165
	400/800	0,5	% 5	0,134064	0,134521	0,136029	0,135198	0,134064
			% 10	0,139229	0,140089	0,139338	0,137459	0,138117
			% 20	0,130349	0,133617	0,135035	0,133341	0,131663
		1,0	% 5	0,127527	0,127526	0,127678	0,129455	0,139594
			% 10	0,138923	0,135901	0,137881	0,129716	0,136944
			% 20	0,142688	0,141648	0,142059	0,141479	0,139063
2400	1200/1200	0,5	% 5	0,092319	0,089678	0,093076	0,090846	0,091687
			% 10	0,093466	0,094577	0,093057	0,093466	0,093660
			% 20	0,098596	0,097029	0,097787	0,097612	0,095889
		1,0	% 5	0,094254	0,095135	0,096121	0,095371	0,095135
			% 10	0,094346	0,093411	0,092275	0,094346	0,097041
			% 20	0,101145	0,108496	0,102185	0,101145	0,100348
	800/1600	0,5	% 5	0,090851	0,090612	0,090851	0,093752	0,093752
			% 10	0,096167	0,094805	0,094110	0,094317	0,094110
			% 20	0,095053	0,098538	0,095381	0,096071	0,098831
		1,0	% 5	0,093301	0,097417	0,093806	0,092856	0,097593
			% 10	0,093327	0,093465	0,093327	0,093749	0,099302
			% 20	0,096326	0,096017	0,097611	0,097422	0,096856

Tablo 21’de yer alan bulgular incelendiğinde, ayırt edicilik parametresine ilişkin RMSE değerleri örneklem büyüklüğü 600’den 2400’e yükseldikçe azalmış, aynı zamanda yanlışlık ve standart hata değerlerinde de bu eğilim görülmüştür (bkz. Ek 1, Ek 2 ve Ek 3). Örneklem büyüklüğü 600, 1200 ve 2400 iken DMF belirleme yöntemlerine göre RMSE değerlerinin genel olarak birbirlerine yakın değerler aldıkları da söylenebilir. Odak ve referans grubun örneklem büyüklüğü oranları açısından incelendiğinde ise hem 1/1 hem de 1/2 oranlarında RMSE değerlerinin birbirlerine yakın değerler aldıkları görülmüştür. DMF büyüklüğü 0,5 iken elde edilen RMSE değerlerinin 1,0’a göre nispeten daha düşük olduğu söylenebilir. Daha açık ifadeyle DMF büyüklüğü arttıkça RMSE değerleri artmış, yanlışlık ve standart hata değerlerinde de bu eğilim tespit edilmiştir (bkz. Ek 1, Ek 2 ve Ek 3). Testte bulunan DMF’li madde oranları %5’ten %20’ye yükseldikçe RMSE değerleri göreceli olarak artmış, aynı şekilde yanlışlık ve standart hata değerleride yükselmiştir (bkz. Ek 1, Ek 2 ve Ek 3). Ayrıca DMF’li madde oranı %5 iken Raju’nun alan ölçümü yöntemine göre elde edilen RMSE değerlerinin ortalamasının ise diğer yöntemlere göre nispeten yüksek olduğu söylenebilir.

4.3.1.1. Farklı Örneklem Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre ayırt edicilik parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 22’de sunulmuştur.

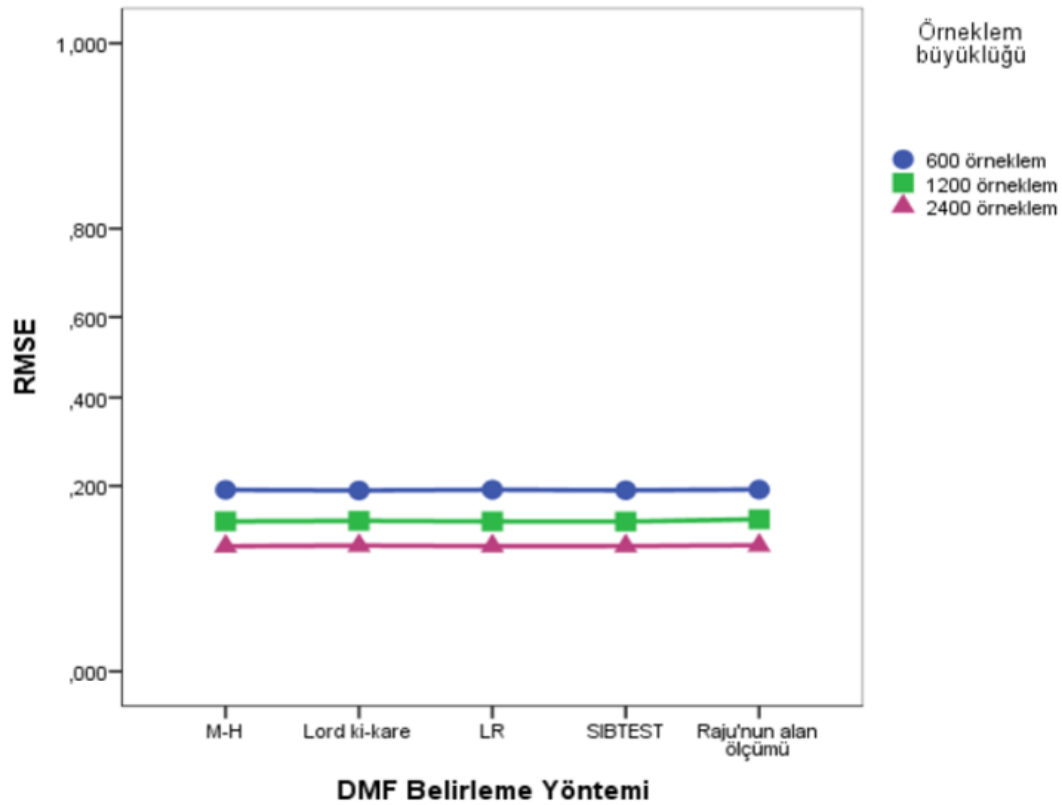
Tablo 22.

Farklı Örneklem Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin İncelenmesi

	Örneklem büyüklüğü						Yöntem	Örneklem büyüklüğü	Etkileşim
	600		1200		2400				
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,1923	0,0067	0,1343	0,0051	0,0949	0,0028		$F_{2;165}=5817,9;$	
Lord	0,1909	0,0058	0,1351	0,0044	0,0958	0,0048	$F_{4;165}=1,013;$	$p=,000$	$F_{8;165}=0,318;$
LR	0,1925	0,0072	0,1342	0,0043	0,0949	0,0031	$p=,402$	$\eta^2=0,986$	$p=,958$
SIBTEST	0,1914	0,0057	0,1339	0,0040	0,0951	0,0027		<u>*Anlamlı fark</u>	
Raju	0,1929	0,0054	0,1377	0,0062	0,0961	0,0026			
*Anlamlı fark: 600-1200 ($\eta^2=0,964$); 600-2400 ($\eta^2=0,990$) ve 1200-2400 ($\eta^2=0,948$)									

Tablo 22’de yer alan bulgular incelendiğinde DMF belirleme yöntemlerine göre RMSE değerleri anlamlı farklılık göstermezken ($F_{4;165}=1,013$; $p=,402$), örneklem büyüklüklerine göre anlamlı farklılık göstermektedir ($F_{2;165}=5817,9$; $p=,000$). RMSE değerlerindeki değişkenliğin yaklaşık %98,6’sı örneklem büyüklükleri ile açıklanmakla birlikte, örneklem büyüklüklerinin RMSE değerleri üzerinde yüksek etkiye sahip olduğu söylenebilir ($\eta^2=0,986$). Ayrıca DMF belirleme yöntemleri ile örneklem büyüklüklerine ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği tespit edilmiştir ($F_{8;165}=0,318$; $p=,958$). Yapılan ikili karşılaştırmalar neticesinde ise tespit edilen anlamlı farklılığın 600-1200 ($\eta^2=0,964$); 600-2400 ($\eta^2=0,990$) ve 1200-2400 ($\eta^2=0,948$) örneklem büyüklükleri arasında olduğu saptanmıştır.

Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre ayırt edicilik parametresinin RMSE değerleri Şekil 8’de sunulmuştur.



Şekil 8. Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre ayırt edicilik parametresinin RMSE değerleri

Şekil 8’de yer alan grafik incelendiğinde ayırt edicilik parametresine ilişkin RMSE değerleri örneklem büyüklüğü 600’den 2400’e yükseldikçe azalmaktadır. Buna rağmen farklı DMF belirleme yöntemlerine göre RMSE değerlerinin değişmediği görülmektedir.

4.3.1.2. Farklı Örneklem Oranları ve DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı örneklem büyüklüğü oranları ve farklı DMF belirleme yöntemlerine göre ayırt edicilik parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 23’te sunulmuştur.

Tablo 23.

Farklı Örneklem Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin İncelenmesi

Yöntem	Örneklem oranı				Yöntem	Örneklem oranı	Etkileşim
	1/1		1/2				
	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,1410	0,0419	0,1400	0,0409	F _{4;170} =0,015; p= ,999	F _{1;170} =0,001; p= ,981	F _{4;170} =0,003; p= ,999
Lord	0,1407	0,0403	0,1406	0,0407			
LR	0,1400	0,0413	0,1411	0,0418			
SIBTEST	0,1406	0,0415	0,1396	0,0404			
Raju	0,1422	0,0409	0,1424	0,0411			

Tablo 23’de yer alan bulgular incelendiğinde hem DMF belirleme yöntemlerine hem de örneklem büyüklüğü oranlarına göre RMSE değerlerinin anlamlı farklılık göstermediği tespit edilmiştir ($F_{4;170}=0,015$; $p=,999$ ve $F_{1;170}=0,001$; $p=,981$). Ayrıca DMF belirleme yöntemleri ile örneklem büyüklüğü oranlarına ilişkin ortak etkinde istatistiksel olarak anlamlı farklılık göstermediği saptanmıştır ($F_{4;170}=0,003$; $p=,999$).

4.3.1.3. Farklı DMF Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı DMF büyüklükleri ve farklı DMF belirleme yöntemlerine göre ayırt edicilik parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 24’te sunulmuştur.

Tablo 24.

Farklı DMF Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin İncelenmesi

Yöntem	DMF Büyüklüğü				Yöntem	DMF Büyüklüğü	Etkileşim
	0,5		1,0				
	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,1397	0,0406	0,1413	0,0423	F _{4;170} =0,015; p= ,999	F _{1;170} =0,167; p= ,683	F _{4;170} =0,014; p= ,999
Lord	0,1394	0,0403	0,1419	0,0406			
LR	0,1395	0,0404	0,1416	0,0427			
SIBTEST	0,1396	0,0405	0,1407	0,0414			
Raju	0,1396	0,0404	0,1449	0,0415			

Tablo 24’te yer alan bulgular incelendiğinde hem DMF belirleme yöntemlerine hem de DMF büyüklüklerine göre RMSE değerlerinin anlamlı farklılık göstermediği tespit edilmiştir (F_{4;170}=0,015; p=,999 ve F_{1;170}=0,167; p=,683). Ayrıca DMF belirleme yöntemleri ile DMF büyüklüklerine ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği saptanmıştır (F_{4;170}=0,014; p=,999).

4.3.1.4. Farklı DMF’li Madde Oranları ve DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı DMF’li madde oranları ve farklı DMF belirleme yöntemlerine göre ayırt edicilik parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 25’te sunulmuştur.

Tablo 25.

Farklı DMF’li Madde Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Ayırt Edicilik Parametresinin RMSE Değerlerinin İncelenmesi

	DMF'li madde oranları							DMF'li madde oranları	Etkileşim
	%5		%10		%20		Yöntem		
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,1370	0,0411	0,1412	0,0432	0,1433	0,0417			
Lord	0,1373	0,0402	0,1406	0,0425	0,1440	0,0403	F _{4;165} =0,014;	F _{2;165} =0,256;	F _{8;165} =0,006;
LR	0,1377	0,0407	0,1403	0,0430	0,1437	0,0425	p= ,999	p= ,774	p= ,999
SIBTEST	0,1379	0,0404	0,1395	0,0428	0,1429	0,0412			
Raju	0,1401	0,0411	0,1433	0,0419	0,1434	0,0418			

Tablo 25'te yer alan bulgular incelendiğinde hem DMF belirleme yöntemlerine hem de DMF'li madde oranlarına göre RMSE değerlerinin anlamlı farklılık göstermediği tespit edilmiştir ($F_{4;165}=0,014$; $p=,999$ ve $F_{2;165}=0,256$; $p=,774$). Ayrıca DMF belirleme yöntemleri ile DMF'li madde oranlarına ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği saptanmıştır ($F_{8;165}=0,006$; $p=,999$).

4.3.2. Güçlük Parametresine İlişkin RMSE Değerleri

DMF belirleme yöntemlerine göre elde edilen güçlük parametre kestirimine ait RMSE değerleri; örneklem büyüklükleri, odak ve referans grup örneklem büyüklüğü oranları, DMF büyüklükleri ve testte bulunan DMF'li maddelerin oranları koşulları altında incelenmiştir. İlgili koşullar altında güçlük parametresine ilişkin RMSE değerleri Tablo 26'da sunulmuştur.

Tablo 26.

İlgili Koşullar Altında Güçlük Parametresine İlişkin RMSE Değerleri

Koşullar				DMF Belirleme Yöntemleri				
Örneklem	Örneklem oranı	DMF büyüklüğü	DMF'li madde oranı	M-H	Lord'un χ^2	LR	SIBTEST	Raju'nun alan ölçümü
				RMSE	RMSE	RMSE	RMSE	RMSE
600	300/300	0,5	% 5	0,147590	0,146314	0,147029	0,149912	0,148499
			% 10	0,108299	0,107577	0,106206	0,108027	0,109285
			% 20	0,154545	0,156644	0,155333	0,152539	0,155960
		1,0	% 5	0,113993	0,107942	0,111242	0,107962	0,109684
			% 10	0,157111	0,158663	0,110745	0,159303	0,158234
			% 20	0,101577	0,172121	0,101219	0,213422	0,182205
	200/400	0,5	% 5	0,103351	0,155689	0,105687	0,152704	0,157297
			% 10	0,148092	0,147892	0,148039	0,147020	0,147892
			% 20	0,121900	0,122962	0,122962	0,121326	0,122155
		1,0	% 5	0,101339	0,102393	0,103229	0,102576	0,113152
			% 10	0,127890	0,129614	0,127890	0,131988	0,133528
			% 20	0,132172	0,130405	0,137867	0,127684	0,131799
1200	600/600	0,5	% 5	0,072927	0,081439	0,082924	0,081948	0,081948
			% 10	0,084397	0,084650	0,084588	0,083427	0,084749
			% 20	0,084814	0,083622	0,071870	0,083622	0,083622
		1,0	% 5	0,072950	0,084725	0,073709	0,083579	0,083579
			% 10	0,077885	0,103439	0,078545	0,102862	0,116792
			% 20	0,076622	0,083724	0,074475	0,077152	0,084395
	400/800	0,5	% 5	0,092798	0,092745	0,092385	0,093958	0,092798
			% 10	0,085679	0,083221	0,085239	0,084882	0,086079
			% 20	0,072844	0,090297	0,088911	0,087403	0,086079
		1,0	% 5	0,077248	0,077248	0,075102	0,074447	0,080011
			% 10	0,088978	0,087432	0,086917	0,078053	0,089371
			% 20	0,102919	0,101152	0,100582	0,102048	0,096549
2400	1200/1200	0,5	% 5	0,054544	0,052047	0,054349	0,054037	0,061884
			% 10	0,050479	0,061198	0,050827	0,050479	0,060639
			% 20	0,058509	0,058655	0,059519	0,057795	0,058865
		1,0	% 5	0,073213	0,073156	0,073663	0,072887	0,073156
			% 10	0,053836	0,054028	0,053639	0,053836	0,068811
			% 20	0,058336	0,069429	0,059295	0,058337	0,063917
	800/1600	0,5	% 5	0,054119	0,053441	0,054119	0,060845	0,060845
			% 10	0,057989	0,058521	0,057966	0,058179	0,057966
			% 20	0,050555	0,058565	0,047852	0,052701	0,059002
		1,0	% 5	0,050011	0,062357	0,050164	0,050423	0,061497
			% 10	0,052362	0,052144	0,052362	0,052702	0,068195
			% 20	0,051186	0,052999	0,053383	0,053204	0,068854

Tablo 26’da yer alan bulgular incelendiğinde güçlük parametresine ilişkin RMSE değerleri örneklem büyüklüğü 600’den 2400’e yükseldikçe azalmış, aynı zamanda yanlılık ve standart hata değerlerinde de bu eğilim görülmüştür (bkz. Ek 4, Ek 5 ve Ek 6). Örneklem büyüklüğü 600, 1200 ve 2400 iken DMF belirleme yöntemlerine göre elde edilen RMSE değerlerinin göreceli olarak farklılaştığı söylenebilir. Örneklem büyüklüğü 600 iken Lojistik Regresyon ve Mantel-Haenszel yöntemlerine göre ortalama RMSE değerlerinin nispeten daha düşük olduğu ($RMSE_{LR}=0,1231$ ve $RMSE_{M-H}=0,1265$); SIBTEST ve Raju’nun alan ölçümü yöntemlerine göre ise ortalama RMSE değerlerinin nispeten daha yüksek olduğu ($RMSE_{SIBTEST}=0,1395$ ve $RMSE_{Raju}=0,1391$). Örneklem büyüklüğü 1200 ve 2400 koşulu altında Lojistik Regresyon ve Mantel-Haenszel yöntemine göre elde edilen ortalama RMSE değerlerinin göreceli olarak daha düşük; Raju’nun alan ölçümü yöntemine göre ise göreceli olarak daha yüksek olduğu söylenebilir.

DMF büyüklüğü 0.5 koşulu altında ve örneklem büyüklüğü 600 iken tüm yöntemlere göre elde edilen RMSE değerlerinin, DMF büyüklüğü 1,0’a göre göreceli olarak daha yüksek olduğu görülmüştür. Benzer şekilde yanlılık ve standart hata değerlerinde de bu eğilim görülmüştür (bkz. Ek 4, Ek 5 ve Ek 6). Fakat örneklem büyüklüğü arttığında, 1,0 DMF büyüklüğündeki RMSE değerlerinin 0,5 DMF büyüklüğüne göre nispeten daha yüksek olduğu görülmüştür. Raju’nun alan ölçümü yöntemine göre elde edilen RMSE değerlerinin DMF büyüklüğü 1,0, örneklem büyüklüğü 1200 ve 2400 koşullarında diğer yöntemlere göre göreceli olarak daha yüksek olduğu söylenebilir. Aynı koşullar altında Lojistik Regresyon ve Mantel-Haenszel ve SIBTEST yöntemlerine göre RMSE değerlerinin göreceli olarak daha düşük olduğu görülmüştür.

Testte bulunan DMF’li madde oranları %5’ten %20’ye yükseldikçe Mantel-Haenszel yöntemi dışındaki tüm yöntemlere göre genel olarak RMSE değerleri artmış, aynı zamanda yanlılık ve standart hata değerlerinde de bu eğilim görülmüştür (bkz. Ek 4, Ek 5 ve Ek 6). Mantel-Haenszel yönteminde ise DMF’li madde oranı %10 iken elde edilen RMSE değerlerinin daha yüksek, %5 iken ise daha düşük olduğu tespit edilmiştir. DMF’li madde oranı %5, %10 ve %20 iken Lojistik Regresyon ve Mantel-Haenszel yöntemlerine göre en düşük; Raju’nun alan ölçümü yöntemine göre ise en yüksek RMSE değerleri elde edilmiştir. SIBTEST ve Lord’un χ^2 , yöntemlerine göre elde edilen RMSE değerlerinin ise tüm DMF’li madde oranlarında birbirlerine yakın değerler aldıkları görülmüştür.

4.3.2.1. Farklı Örneklem Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre güçlük parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 27’de sunulmuştur.

Tablo 27.

Farklı Örneklem Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin İncelenmesi

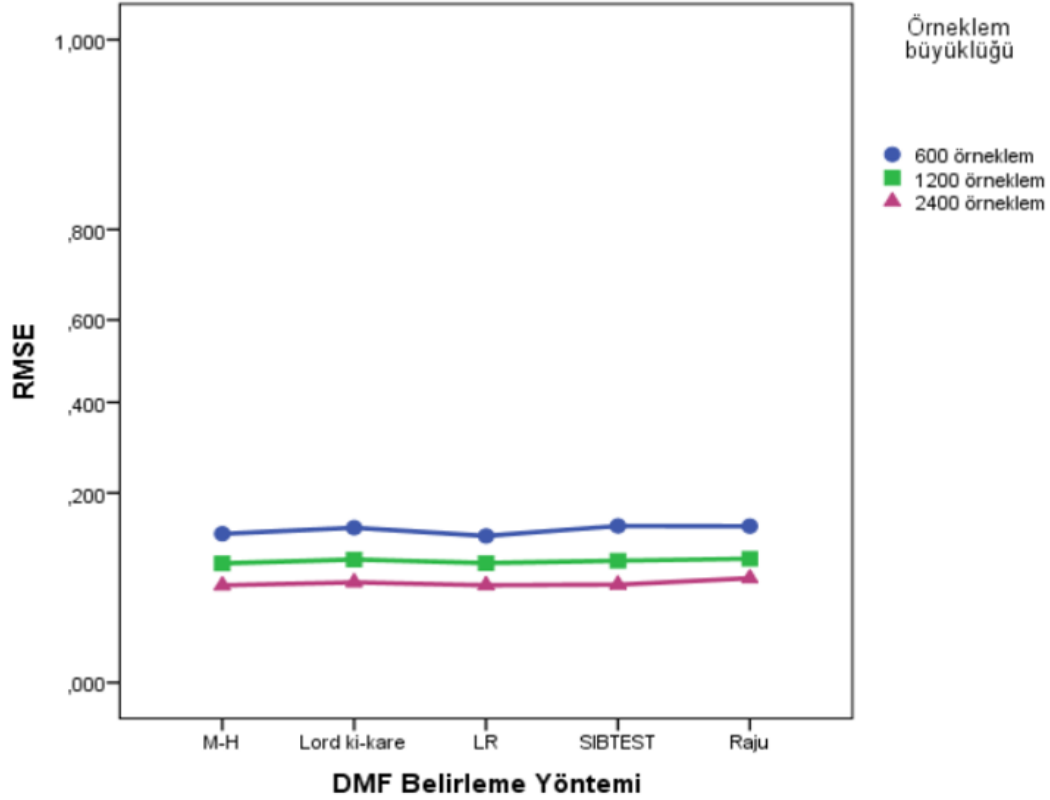
	Örneklem büyüklüğü						Yöntem	Örneklem büyüklüğü	Etkileşim
	600		1200		2400				
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,1265	0,0213	0,0825	0,0093	0,0554	0,0064	$F_{4;165}=2,930$;	$F_{2;165}=377,867$;	
Lord	0,1365	0,0231	0,0878	0,0079	0,0589	0,0068	$p=,023$	$p=,000$	$F_{8;165}=0,591$
LR	0,1231	0,0196	0,0829	0,0086	0,0556	0,0067	$\eta^2=0,066$	$\eta^2=0,821$	$p=,785$
SIBTEST	0,1395	0,0305	0,0861	0,0091	0,0562	0,0062	<u>*Anlamlı fark</u>	<u>**Anlamlı fark</u>	
Raju	0,1391	0,0229	0,0888	0,0100	0,0636	0,0049			
$\ast$ Anlamlı fark: M-H ve Raju’nun alan ölçümü ($\eta^2=0,018$); Lord’un χ^2 ve LR ($\eta^2=0,012$); LR ve Raju’nun alan ölçümü ($\eta^2=0,023$)									
$\ast\ast$ Anlamlı fark: 600-1200 ($\eta^2=0,634$); 600-2400 ($\eta^2=0,821$) ve 1200-2400 ($\eta^2=0,752$)									

Tablo 27’de yer alan bulgular incelendiğinde, hem DMF belirleme yöntemlerine hem de örneklem büyüklüklerine göre RMSE değerleri anlamlı farklılık göstermektedir ($F_{4;165}=2,930$; $p=,023$ ve $F_{2;165}=377,867$; $p=,000$). RMSE değerlerindeki değişkenliğin yaklaşık %6,6’sı DMF belirleme yöntemleri ile açıklanmakla birlikte RMSE değerleri üzerinde orta düzeyde etkiye sahip olduğu söylenebilir ($\eta^2=0,066$). Benzer şekilde RMSE değerlerindeki değişkenliğin yaklaşık %82,1’i ise örneklem büyüklükleri ile açıklanmakla birlikte örneklem büyüklüklerinin de RMSE değerleri üzerinde yüksek etkiye sahip olduğu söylenebilir ($\eta^2=0,821$). Ayrıca DMF belirleme yöntemleri ile örneklem büyüklüklerine ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği tespit edilmiştir ($F_{8;165}=0,591$; $p=,785$).

Yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın ise Mantel-Haenszel ve Raju’nun alan ölçümü ($\eta^2=0,018$); Lord’un χ^2 ve Lojistik Regresyon, ($\eta^2=0,012$); Lojistik Regresyon ve Raju’nun alan ölçümü ($\eta^2=0,023$) yöntemleri arasında olduğu saptanmıştır. Örneklem büyüklükleri üzerinden yapılan ikili karşılaştırmalar

sonucunda ise 600-1200 ($\eta^2=0,634$); 600-2400 ($\eta^2=0,821$) ve 1200-2400 ($\eta^2=0,752$) arasında olduğu tespit edilmiştir.

Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre güçlük parametresinin RMSE değerleri Şekil 9’da sunulmuştur.



Şekil 9. Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre güçlük parametresinin RMSE değerleri

Şekil 9’da yer alan grafik incelendiğinde güçlük parametresine ilişkin RMSE değerleri örneklem büyüklüğü 600’den 2400’e yükseldikçe azalmaktadır. Buna rağmen farklı DMF belirleme yöntemlerine göre RMSE değerlerinin değişmediği görülmektedir.

4.3.2.2. Farklı Örneklem Büyüklüğü Oranları ve DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı örneklem büyüklüğü oranları ve farklı DMF belirleme yöntemlerine göre güçlük parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 28’de sunulmuştur.

Tablo 28.

Farklı Örneklem Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin İncelenmesi

Yöntem	Örneklem oranı				Yöntem	Örneklem oranı	Etkileşim
	1/1		1/2				
	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,0889	0,3463	0,0873	0,0315	F _{4;170} = 0,540; p= ,706	F _{1;170} = 0,306; p= ,581	F _{4;170} = 0,084; p= ,987
Lord	0,0966	0,0380	0,0922	0,0337			
LR	0,0861	0,0305	0,0884	0,0319			
SIBTEST	0,0973	0,0445	0,0907	0,0339			
Raju	0,0992	0,0384	0,0952	0,0319			

Tablo 28’de yer alan bulgular incelendiğinde hem DMF belirleme yöntemlerine hem de örneklem büyüklüğü oranlarına göre RMSE değerlerinin anlamlı farklılık göstermediği tespit edilmiştir ($F_{4;170}= 0,540$; $p=,706$ ve $F_{1;170}= 0,306$; $p=,581$). Ayrıca DMF belirleme yöntemleri ile örneklem oranlarına ilişkin ortak etkinin de istatistiksel olarak anlamlı farklılık göstermediği saptanmıştır ($F_{4;170}= 0,084$; $p=,987$).

4.3.2.3. Farklı DMF Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı DMF büyüklükleri ve farklı DMF belirleme yöntemlerine göre güçlük parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 29’da sunulmuştur.

Tablo 29.

Farklı DMF Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin İncelenmesi

	DMF Büyüklüğü				Yöntem	DMF Büyüklüğü	Etkileşim
	0,5		1,0				
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,0891	0,0349	0,0872	0,0311			
Lord	0,0941	0,0369	0,0946	0,0350	$F_{4;170}=0,539$;	$F_{1;170}= 0,004$;	$F_{4;170}= 0,082$;
LR	0,0898	0,0349	0,0847	0,0269	$p= ,707$	$p= ,952$	$p= ,988$
SIBTEST	0,0934	0,0369	0,0946	0,0428			
Raju	0,0953	0,0361	0,0991	0,0346			

Tablo 29’da yer alan bulgular incelendiğinde hem DMF belirleme yöntemlerine hem de DMF büyüklüklerine göre RMSE değerlerinin anlamlı farklılık göstermediği tespit edilmiştir. ($F_{4;170}=0,539$; $p=,707$ ve $F_{1;170}= 0,004$; $p=,952$). Ayrıca DMF belirleme yöntemleri ile DMF büyüklüklerine ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği saptanmıştır ($F_{4;170}= 0,082$; $p=,988$).

4.3.2.4. Farklı DMF’li Madde Oranları ve DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı DMF’li madde oranları ve farklı DMF belirleme yöntemlerine göre güçlük parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 30’da sunulmuştur.

Tablo 30.

Farklı DMF’li Madde Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Güçlük Parametresinin RMSE Değerlerinin İncelenmesi

	DMF’li madde oranları							DMF’li madde oranları	
	%5		%10		%20		Yöntem		Etkileşim
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,0845	0,0286	0,0911	0,0371	0,0888	0,0341			
Lord	0,0908	0,0331	0,0940	0,0364	0,0984	0,0394	$F_{4;165}=0,526;$	$F_{2;165}=0,439;$	$F_{8;165}=0,290;$
LR	0,0853	0,0283	0,0869	0,0314	0,0894	0,0349	$p=,717$	$p=,645$	$p=,999$
SIBTEST	0,0904	0,0335	0,0926	0,0378	0,0989	0,0484			
Raju	0,0937	0,0326	0,0985	0,0344	0,0995	0,0402			

Tablo 30’da yer alan bulgular incelendiğinde hem DMF belirleme yöntemlerine hem de DMF’li madde oranlarına göre RMSE değerlerinin anlamlı farklılık göstermediği tespit edilmiştir ($F_{4;165}=0,526$; $p=,717$ ve $F_{2;165}=0,439$; $p=,645$). Ayrıca DMF belirleme yöntemleri ile DMF’li madde oranlarına ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği saptanmıştır ($F_{8;165}=0,290$; $p=,999$).

4.4. Farklı Koşullar Altında DMF’li Maddelerin Testten Çıkarılmasında Yetenek Parametre Kestirimleri

“DMF belirleme yöntemlerine göre örneklem büyüklükleri (600, 1200 ve 2400), odak ve referans grup örneklem büyüklüğü oranları (1/1 ve 1/2), DMF büyüklükleri (0,5 ve 1,0) ve

testte bulunan DMF'li maddelerin oranları (%5, %10 ve %20) koşullarında yetenek parametre kestirimlerinin RMSE, standart hata ve yanlılık değerleri nasıl değişmektedir?" araştırma sorusuna yönelik elde edilen bulgular aşağıda verilmiştir.

DMF belirleme yöntemlerine göre elde edilen yetenek parametre kestirimine ait RMSE değerleri; örneklem büyüklükleri, odak ve referans grup örneklem büyüklüğü oranları, DMF büyüklükleri ve testte bulunan DMF'li maddelerin oranları koşulları altında incelenmiştir. İlgili koşullar altında yetenek parametresine ilişkin RMSE değerleri Tablo 31'de sunulmuştur.

Tablo 31.

İlgili Koşullar Altında Yetenek Parametresine İlişkin RMSE Değerleri

Koşullar				DMF Belirleme Yöntemleri				
Örneklem	Örneklem oranı	DMF büyüklüğü	DMF'li madde oranı	M-H	Lord'un χ^2	LR	SIBTEST	Raju'nun alan ölçümü
				RMSE	RMSE	RMSE	RMSE	RMSE
600	300/300	0,5	% 5	0,261724	0,259438	0,266237	0,266912	0,263641
			% 10	0,265404	0,257498	0,267347	0,255903	0,255328
			% 20	0,273932	0,271433	0,272457	0,265588	0,269352
		1,0	% 5	0,271987	0,256146	0,267352	0,258652	0,260980
			% 10	0,270581	0,273303	0,280412	0,271606	0,271280
			% 20	0,278732	0,270239	0,280419	0,298431	0,271656
	200/400	0,5	% 5	0,267276	0,263068	0,266158	0,258741	0,266049
			% 10	0,263537	0,259429	0,257666	0,254224	0,259429
			% 20	0,264587	0,267394	0,267394	0,262056	0,262929
		1,0	% 5	0,273423	0,265438	0,275044	0,270114	0,266749
			% 10	0,269766	0,263461	0,269766	0,266115	0,265396
			% 20	0,279799	0,275120	0,283694	0,274890	0,277601
1200	600/600	0,5	% 5	0,267407	0,257426	0,265996	0,263729	0,263729
			% 10	0,271793	0,265434	0,270756	0,265159	0,268046
			% 20	0,277359	0,265890	0,271777	0,265890	0,265890
		1,0	% 5	0,273593	0,269749	0,276633	0,274508	0,274508
			% 10	0,288640	0,298116	0,290974	0,282356	0,416839
			% 20	0,300591	0,282346	0,289301	0,292340	0,286623
	400/800	0,5	% 5	0,269274	0,268343	0,271719	0,271299	0,269274
			% 10	0,278795	0,264451	0,275669	0,270495	0,263559
			% 20	0,268883	0,277817	0,280715	0,268583	0,270109
		1,0	% 5	0,268002	0,268002	0,269355	0,271476	0,267124
			% 10	0,275007	0,268943	0,271498	0,270263	0,272177
			% 20	0,297307	0,291433	0,295597	0,289764	0,291137
2400	1200/1200	0,5	% 5	0,277897	0,265083	0,285036	0,270821	0,266238
			% 10	0,274121	0,268697	0,274748	0,274121	0,265172
			% 20	0,300544	0,289429	0,294669	0,296392	0,283706
		1,0	% 5	0,278969	0,266498	0,272999	0,280394	0,266498
			% 10	0,281665	0,276046	0,273380	0,281665	0,283024
			% 20	0,339406	0,346220	0,348834	0,339406	0,313935
	800/1600	0,5	% 5	0,268897	0,265466	0,268897	0,268118	0,268118
			% 10	0,272964	0,273112	0,269471	0,272584	0,269471
			% 20	0,278322	0,269949	0,282863	0,287391	0,270312
		1,0	% 5	0,268471	0,269023	0,269832	0,260405	0,267428
			% 10	0,278479	0,280107	0,278479	0,281266	0,316215
			% 20	0,323135	0,303752	0,315082	0,318767	0,294520

Tablo 31’de yer alan bulgular incelendiğinde yetenek parametresine ilişkin RMSE değerleri örneklem büyüklüğü 600’den 2400’e yükseldikçe genel olarak artmış, aynı zamanda yanlılık ve standart hata değerlerinde de bu eğilim görülmüştür (bkz. Ek. 7, Ek. 8 ve Ek. 9). Odak ve referans grup sayısı eşit (1/1) iken elde edilen yetenek parametresine ilişkin RMSE değerlerinin örneklem oranının 1/2’ye göre göreceli olarak yüksek olduğu söylenebilir. DMF büyüklüğü 0,5 iken elde edilen RMSE değerlerinin 1,0’a göre nispeten düşük olduğu görülmüştür. Daha açık ifadeyle DMF büyüklüğü arttıkça yetenek parametresine ilişkin RMSE değerleri göreceli olarak artmış, yanlılık ve standart hata değerlerinde de bu artış gözlemlenmiştir (bkz. Ek. 7, Ek. 8 ve Ek. 9).

DMF büyüklüğü 0,5 iken Lord’un χ^2 ve Raju’nun alan ölçümü yöntemlerine göre ortalama RMSE değerlerinin nispeten düşük olduğu; Mantel-Haenszel ve Lojistik Regresyon yöntemlerine göre ise ortalama RMSE değerlerinin nispeten daha yüksek söylenebilir. DMF büyüklüğü 1,0 iken ise Lord’un χ^2 yöntemine göre ortalama RMSE değerlerinin göreceli olarak düşük; Mantel-Haenszel ve Raju’nun alan ölçümü yöntemlerine göre ise ortalama RMSE değerlerinin göreceli olarak yüksek olduğu söylenebilir.

Testte bulunan DMF’li madde oranları %5’ten %20’ye yükseldikçe genel olarak RMSE değerleri artmış, aynı zamanda yanlılık ve standart hata değerlerinde de bu eğilim görülmüştür (bkz. Ek. 7, Ek. 8 ve Ek. 9). DMF’li madde oranı %5 iken Mantel-Haenszel ve Lojistik Regresyon yöntemine göre ortalama RMSE değerleri göreceli olarak daha yüksek; Lord’un χ^2 yöntemine göre ise göreceli olarak daha düşüktür. DMF’li madde oranı %20 ye yükseldiğinde ise Mantel-Haenszel ve Lojistik Regresyon yöntemine göre ortalama RMSE değerleri göreceli olarak daha yüksek; Raju’nun alan ölçümü yöntemine göre ise göreceli olarak daha düşük olduğu söylenebilir.

4.4.1. Farklı Örneklem Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre yetenek parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 32’de sunulmuştur.

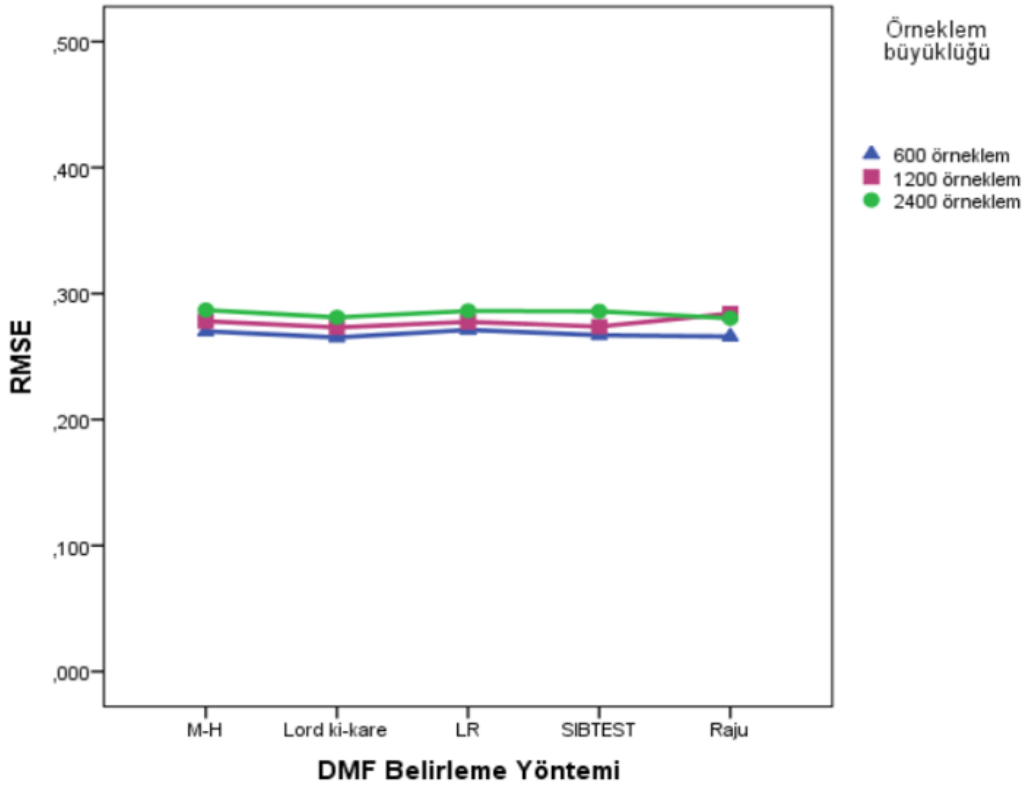
Tablo 32.

Farklı Örneklem Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin İncelenmesi

Yöntem	Örneklem						Yöntem	Örneklem büyüklüğü	Etkileşim
	600		1200		2400				
	$\bar{X}$	SS	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,2701	0,0058	0,2781	0,0114	0,2870	0,0226		$F_{2;165}=11,871;$	
Lord	0,2652	0,0064	0,2732	0,0119	0,2811	0,0235	$F_{4;165}=0,499;$	$p=,000$	$F_{8;165}=0,381;$
LR	0,2712	0,0075	0,2780	0,0096	0,2862	0,0238	$p=,736$	$\eta^2=0,126$	$p=,929$
SIBTEST	0,2669	0,0118	0,2738	0,0094	0,2860	0,0226		<u>*Anlamlı fark</u>	
Raju	0,2659	0,0061	0,2841	0,0427	0,2804	0,0186			
*Anlamlı fark: 600-1200 ($\eta^2=0,083$); 600-2400 ($\eta^2=0,202$); 1200-2400 ($\eta^2=0,025$)									

Tablo 32’de yer alan bulgular incelendiğinde, DMF belirleme yöntemlerine göre RMSE değerleri anlamlı farklılık göstermezken ($F_{4;165}=0,499$; $p=,736$); örneklem büyüklüklerine göre anlamlı farklılık göstermektedir ($F_{2;165}=11,871$; $p=,000$). RMSE değerlerindeki değişkenliğin yaklaşık %12,6’sı örneklem büyüklükleri ile açıklanmakla birlikte örneklem büyüklüklerinin RMSE değerleri üzerinde yüksek etkiye sahip olduğu söylenebilir ($\eta^2=0,126$). Ayrıca DMF belirleme yöntemleri ile örneklem büyüklüklerine ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği tespit edilmiştir ($F_{8;165}=0,381$; $p=,929$). Örneklem büyüklükleri üzerinden yapılan ikili karşılaştırmalar neticesinde 600-1200 ($\eta^2=0,083$); 600-2400 ($\eta^2=0,202$) ve 1200-2400 ($\eta^2=0,025$) arasında olduğu tespit edilmiştir.

Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre yetenek parametresinin RMSE değerleri Şekil 10’da sunulmuştur.



Şekil 10. Farklı örneklem büyüklükleri ve farklı DMF belirleme yöntemlerine göre yetenek parametresinin RMSE değerleri

Şekil 10’da yer alan grafik incelendiğinde ise yetenek parametresine ilişkin RMSE değerlerinin örneklem büyüklüğü 600’den 2400’e yükseldikçe artmaktadır. Buna rağmen farklı DMF belirleme yöntemlerine göre RMSE değerlerinin değişmediği görülmektedir.

4.4.2. Farklı Örneklem Büyüklüğü Oranları ve DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı örneklem büyüklüğü oranları ve farklı DMF belirleme yöntemlerine göre yetenek parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 33’te sunulmuştur.

Tablo 33.

Farklı Örneklem Büyüklüğü Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin İncelenmesi

Yöntem	Örneklem oranı				Yöntem	Örneklem oranı	Etkileşim
	1/1		1/2				
	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,2808	0,0179	0,2759	0,0141	F _{4;170} =0,450; p= ,772	F _{1;170} =2,741; p= ,999	F _{8;170} =0,066; p= ,992
Lord	0,2744	0,0210	0,2719	0,0109			
LR	0,2805	0,0191	0,2761	0,0122			
SIBTEST	0,2779	0,0196	0,2731	0,0146			
Raju	0,2804	0,0365	0,2732	0,0139			

Tablo 33'te yer alan bulgular incelendiğinde hem DMF belirleme yöntemlerine hem de örneklem büyüklüğü oranlarına göre RMSE değerlerinin anlamlı farklılık göstermediği tespit edilmiştir ($F_{4;170}=0,450;$ $p=,772$ ve $F_{1;170}=2,741;$ $p=,999$). Ayrıca DMF belirleme yöntemleri ile örneklem büyüklüğü oranlarına ilişkin ortak etkinde istatistiksel olarak anlamlı farklılık göstermediği saptanmıştır ($F_{8;170}=0,066;$ $p=,992$).

4.4.3. Farklı DMF Büyüklükleri ve DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı DMF büyüklükleri ve farklı DMF belirleme yöntemlerine göre yetenek parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 34'te sunulmuştur.

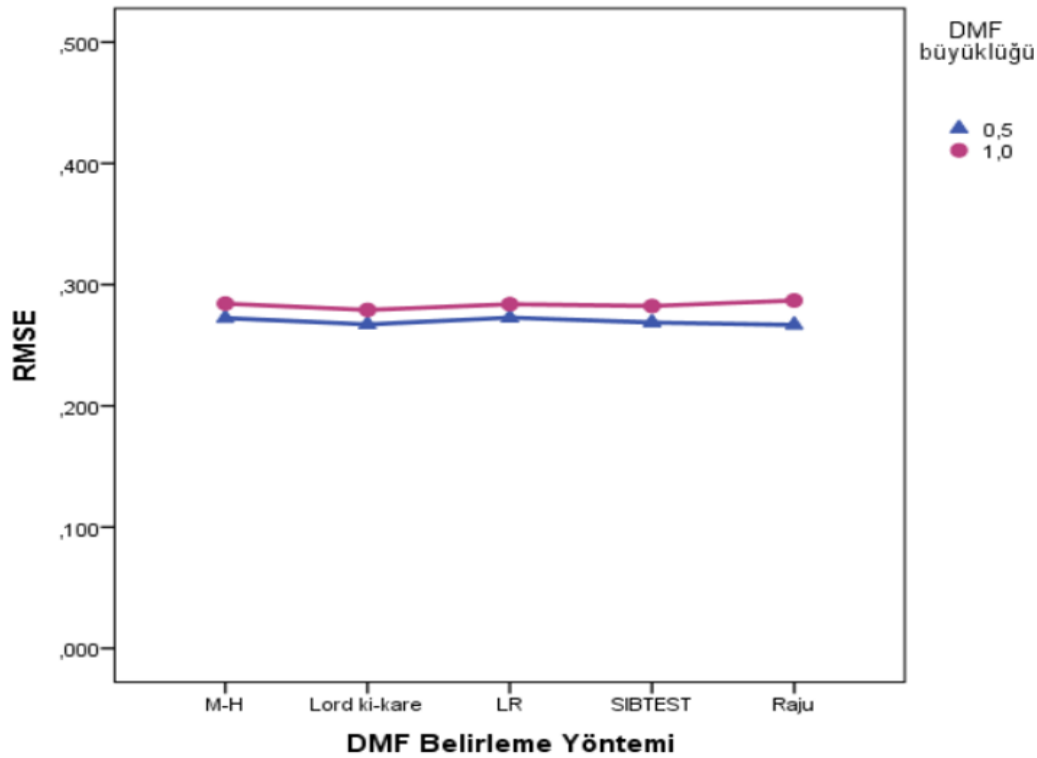
Tablo 34.

Farklı DMF Büyüklükleri ve Farklı DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin İncelenmesi

	DMF Büyüklüğü				Yöntem	DMF Büyüklüğü	Etkileşim
	0,5		1,0				
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,2724	0,0088	0,2844	0,0196	F _{4;170} =0,514; p= ,726	F _{1;170} =25,887; p= ,000 η^2 =0,132	F _{4;170} =0,379; p= ,823
Lord	0,2672	0,0077	0,2791	0,0207			
LR	0,2728	0,0086	0,2838	0,0201			
SIBTEST	0,2688	0,0102	0,2824	0,0202			
Raju	0,2667	0,0058	0,2869	0,0362			

Tablo 34’te yer alan bulgular incelendiğinde DMF belirleme yöntemlerine göre RMSE değerleri anlamlı farklılık göstermezken ($F_{4;170}=0,514$; $p=,726$); DMF büyüklüklerine göre anlamlı farklılık göstermektedir ($F_{1;170}=25,887$; $p=,000$). RMSE değerlerindeki değişkenliğin yaklaşık %13,2’si DMF büyüklükleri ile açıklanmakla birlikte, DMF büyüklüklerinin RMSE değerleri üzerinde yüksek etkiye sahip olduğu söylenebilir ($\eta^2=0,132$). Ayrıca DMF belirleme yöntemleri ile DMF büyüklüklerine ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği tespit edilmiştir ($F_{4;170}=0,379$; $p=,823$).

Farklı DMF büyüklükleri ve farklı DMF belirleme yöntemlerine göre yetenek parametresinin RMSE değerleri Şekil 11’de sunulmuştur.



Şekil 11. Farklı DMF büyüklükleri ve farklı DMF belirleme yöntemlerine göre yetenek parametresinin RMSE değerleri

Şekil 11’de yer alan grafik incelendiğinde ise DMF büyüklüğü 1,0 iken elde edilen yetenek parametresine ilişkin RMSE değerlerinin nispeten yüksek olduğu söylenebilir. Dolayısıyla beş DMF belirleme yönteminde de DMF büyüklüğü 0,5’ten 1,0’a yükseldiğinde RMSE değerlerinin artma eğiliminde olduğu görülmektedir. Buna rağmen farklı DMF belirleme yöntemlerine göre RMSE değerlerinin değişmediği görülmektedir.

4.4.4. Farklı DMF'li Madde Oranları ve DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin Kıyaslanması

Farklı DMF'li madde oranları ve farklı DMF belirleme yöntemlerine göre yetenek parametresine ilişkin RMSE değerleri 5x3 Faktöriyel ANOVA ile karşılaştırılmıştır. Elde edilen sonuçlar Tablo 35'te sunulmuştur.

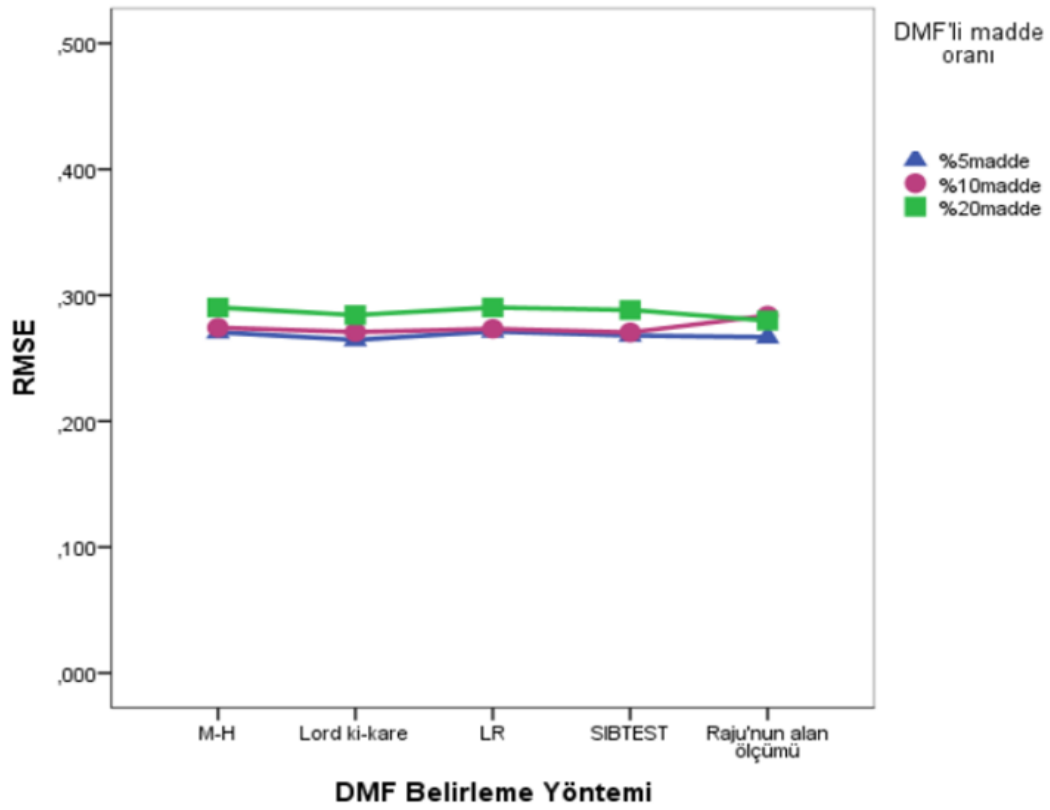
Tablo 35.

Farklı DMF'li Madde Oranları ve Farklı DMF Belirleme Yöntemlerine Göre Yetenek Parametresinin RMSE Değerlerinin İncelenmesi

	DMF'li madde oranları							DMF'li madde oranları	Etkileşim
	%5		%10		%20		Yöntem		
Yöntem	$\bar{X}$	SS	$\bar{X}$	SS	$\bar{X}$	SS			
M-H	0,2706	0,0048	0,2742	0,0069	0,2902	0,0228		$F_{2;165}=16,520$;	
Lord	0,2645	0,0046	0,2707	0,0109	0,2843	0,0226	$F_{4;165}=0,533$;	$p=,000$	$F_{8;165}=0,828$;
LR	0,2713	0,0056	0,2734	0,0081	0,2902	0,0225	$p=,712$	$\eta^2=0,167$	$p=,579$
SIBTEST	0,2679	0,0066	0,2705	0,0092	0,2883	0,0234		<u>*Anlamlı fark</u>	
Raju	0,2667	0,0033	0,2838	0,0447	0,2798	0,0148			
*Anlamlı fark: %5-%20 ($\eta^2=0,265$); %10-%20 ($\eta^2=0,076$)									

Tablo 35'te yer alan bulgular incelendiğinde DMF belirleme yöntemlerine göre RMSE değerleri anlamlı farklılık göstermezken ($F_{4;165}=0,533$; $p=,712$); DMF'li madde oranlarına göre anlamlı farklılık göstermektedir ($F_{2;165}=16,520$; $p=,000$). RMSE değerlerindeki değişkenliğin yaklaşık %16,7'si DMF'li madde oranları ile açıklanmakla birlikte, DMF'li madde oranlarının RMSE değerleri üzerinde yüksek etkiye sahip olduğu söylenebilir ($\eta^2=0,167$). Ayrıca DMF belirleme yöntemleri ile DMF'li madde oranlarına ilişkin ortak etkininde istatistiksel olarak anlamlı farklılık göstermediği tespit edilmiştir ($F_{8;165}=0,828$; $p=,579$). DMF'li madde oranları üzerinden yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın ise %5-%20 ($\eta^2=0,265$) ve %10-%20 ($\eta^2=0,076$) arasında olduğu saptanmıştır.

Farklı DMF'li madde oranları ve farklı DMF belirleme yöntemlerine göre yetenek parametresinin RMSE değerleri Şekil 12'de sunulmuştur.



Şekil 12. Farklı DMF’li madde oranları ve farklı DMF belirleme yöntemlerine göre yetenek parametresinin RMSE değerleri

Şekil 12’de yer alan grafik incelendiğinde ise testte bulunan DMF’li madde oranı %5’ten %20 ye yükseldikçe elde edilen yetenek parametresine ilişkin RMSE değerleri artmaktadır. Buna rağmen farklı DMF belirleme yöntemlerine göre RMSE değerlerinin değişmediği görülmektedir.

BÖLÜM V

SONUÇ VE TARTIŞMA

5.1. Sonuç ve Tartışma

Araştırmanın birinci alt probleminde beş DMF belirleme yönteminin I. Tip hata oranları farklı örneklem büyüklükleri, örneklem büyüklüğü oranları, DMF büyüklükleri ve testte bulunan DMF’li madde oranları kapsamında incelenmiştir. Bu doğrultuda örneklem büyüklüğü (600, 1200 ve 2400), örneklem büyüklüğü oranı (1/1 ve 1/2), DMF büyüklüğü (0,5 ve 1,0) ve testte bulunan DMF’li madde oranı (%5, %10 ve %20) olmak üzere toplamda 36 koşul ele alınmıştır. Her bir koşul için 100 tekrar yapılarak simülasyon çalışması gerçekleştirilmiştir.

Araştırmanın birinci alt probleminden elde edilen bulgular doğrultusunda I. Tip hata oranlarının örneklem büyüklüklerine göre anlamlı farklılık gösterdiği ortaya konulmuştur. Ayrıca I. Tip hata oranlarındaki değişkenliğin yaklaşık %12’si örneklem büyüklüğü ile açıklanmakla birlikte örneklem büyüklüklerinin I. Tip hata oranları üzerinde orta düzeyde etkiye sahip olduğu tespit edilmiştir ($\eta^2=0,122$). Yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın ise 600-2400 ($\eta^2=0,131$) ve 1200-2400 ($\eta^2=0,035$) örneklem büyüklükleri arasında olduğu saptanmıştır. Örneklem büyüklüğü 600’den 2400’e yükseldikçe beş DMF belirleme yöntemine ilişkin I. Tip hata oranlarında arttığı görülmüştür. Bu bulgu örneklem büyüklüğü arttıkça DMF belirlemede yöntemlerin I. Tip hata oranlarının arttığını belirten alanyazındaki çalışma sonuçlarıyla tutarlılık göstermektedir (Atar, 2007; Bolt, 2002; Duncan, 2006; Finch & French, 2007; Li vd., 2012; Tunç, 2016; Wood, 2011). Bu çalışmada örneklem büyüklüğü 600 iken SIBTEST yönteminin en düşük, Raju’nun alan ölçümü ile Lojistik Regresyon yöntemlerinin ise en

yüksek I. Tip hata oranına sahip olduğu saptanmıştır. Örneklem büyüklüğü 2400 olduğunda ise en düşük Lord'un χ^2 ve Raju'nun alan ölçümü yöntemlerinin; en yüksek ise Mantel-Haenszel yönteminin I. Tip hata oranlarına sahip olduğu belirlenmiştir. Bu araştırmanın bulgularıyla paralel olacak şekilde Li vd. (2012) örneklem büyüklüğü arttıkça Mantel-Haenszel ve Lojistik Regresyon yöntemlerinin I. Tip hata oranlarının arttığı sonucuna ulaşmışlardır. DeMars (2009) ise örneklem büyüklüğü arttığında Lojistik Regresyon ve Mantel-Haenszel yöntemlerinde şişirilmiş I. Tip hataya neden olduğu sonucuna ulaşmıştır.

Örneklem büyüklüğü oranlarına göre yapılan incelemeler doğrultusunda, beş DMF belirleme yöntemine ilişkin I. Tip hata oranlarının odak ve referans grubun örneklem büyüklüğü birbirine eşitken (1/1) daha yüksek olduğu görülmüştür. Bu bulgu eşit olan odak ve referans grubun örneklem büyüklükleri için elde edilen I. Tip hata oranlarının eşit olmayan gruplara göre daha yüksek olduğunu belirten çalışma sonuçlarıyla tutarlılık göstermektedir (Atalay Kabasakal vd., 2014; Kristjansson vd., 2005). Örneklem oranı 1/2 iken I. Tip hata oranları en düşük Lord'un χ^2 , SIBTEST ve Raju'nun alan ölçümü yöntemleri; en yüksek ise Mantel-Haenszel ve Lojistik Regresyon yöntemleri olduğu saptanmıştır. Bu bulgu ile paralel olacak şekilde Awuor (2008) odak ve referans grup oranlarının eşit olmadığı durumlarda, Mantel-Haenszel yönteminin I. Tip hata kontrolünün SIBTEST yöntemine göre daha güçlü olduğunu ifade etmiştir. Örneklem oranı 1/1 iken ise Raju'nun alan ölçümü yönteminin I. Tip hata oranının diğer dört yöntemle göre daha yüksek olduğu belirlenmiştir.

Araştırmada DMF büyüklüklerine (0,5 ve 1,0) göre I. Tip hata oranlarının anlamlı farklılık gösterdiği tespit edilmiştir. Ayrıca I. Tip hata oranlarındaki değişkenliğin yaklaşık %16,7'si DMF büyüklüğü ile açıklanmakla birlikte, DMF büyüklüklerinin I. Tip hata oranları üzerinde yüksek etkiye sahip olduğu saptanmıştır ($\eta^2=0,167$). DMF büyüklüğü 0,5'ten 1,0'e yükseldiğinde beş DMF belirleme yöntemine ilişkin I. Tip hata oranlarının da arttığı sonucuna ulaşılmıştır. Bu bulgu DMF büyüklüğü arttıkça DMF belirlemede yöntemlerin I. Tip hata oranlarının arttığını belirten alanyazındaki çalışma sonuçlarıyla paralellik göstermektedir (Awuor, 2008; Garrett, 2009; Su & Wang, 2005; Stark, Chernyshenko & Drasgow, 2006; Tunç, 2016; Wang & Su, 2004). Ek olarak Awuor (2008) DMF büyüklüğünün yöntemlerin davranışlarını etkilediği sonucuna ulaşmıştır. Ayrıca DMF büyüklüğü 0,5 iken I. Tip hata oranları en düşük olan Lord'un χ^2 ve Raju'nun alan ölçümü yöntemleri iken, en yüksek olan ise Lojistik Regresyon yöntemi olduğu belirlenmiştir. DMF

büyükülüğü 1,0'e yükseldiğinde ise I. Tip hata oranı en düşük SIBTEST yöntemi iken en yüksek Raju'nun alan ölçümü olduğu sonucuna ulaşılmıştır.

Testte bulunan DMF'li madde oranlarına (%5, %10 ve %20) göre I. Tip hata oranlarının anlamlı farklılık gösterdiği tespit edilmiştir. I. Tip hata oranlarındaki değişkenliğin yaklaşık %29,5'i testte bulunan DMF'li madde oranları ile açıklanmakla birlikte, DMF'li madde oranlarının I. Tip hata oranları üzerinde yüksek etkiye sahip olduğu ortaya konulmuştur ($\eta^2=0,295$). Yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın ise %5-%20 ($\eta^2=0,273$) ve %10-%20 ($\eta^2=0,206$) DMF'li madde oranları arasında olduğu saptanmıştır. DMF'li madde oranları %5'ten %20'ye yükseldiğinde beş DMF belirleme yöntemine ilişkin I. Tip hata oranlarının da arttığı sonucuna ulaşılmıştır. Wood (2011) testte bulunan DMF'li madde oranı %5'ten %20'ye yükseldiğinde I. Tip hata oranında bu doğrultuda arttığını belirtmiştir. Bu açıdan Wood (2011)'in yapmış olduğu çalışmanın sonuçları ile tutarlılık göstermektedir. DMF'li madde oranları %5 ve %10 koşulu altında I. Tip hata oranı en düşük Lord'un χ^2 , Mantel-Haenszel ve SIBTEST yöntemleri iken en yüksek Raju'nun alan ölçümü yöntemi olduğu belirlenmiştir. DMF'li madde oranı %20'ye yükseldiğinde ise I. Tip hata oranı en yüksek Mantel-Haenszel, en düşük Raju'nun alan ölçümü yöntemi olduğu sonucuna ulaşılmıştır.

Araştırmanın ikinci alt probleminde beş DMF belirleme yönteminin güç oranları farklı örneklem büyüklükleri, örneklem büyüklüğü oranları, DMF büyüklükleri ve testte bulunan DMF'li madde oranları kapsamında incelenmiştir. Bu doğrultuda örneklem büyüklüğü (600, 1200 ve 2400), örneklem büyüklüğü oranı (1/1 ve 1/2), DMF büyüklüğü (0,5 ve 1,0) ve testte bulunan DMF'li madde oranı (%5, %10 ve %20) olmak üzere toplamda 36 koşul ele alınmıştır. Her bir koşul için 100 tekrar yapılarak simülasyon çalışması gerçekleştirilmiştir.

Araştırmanın ikinci alt probleminden elde edilen bulgular doğrultusunda hem DMF belirleme yöntemlerine hem de örneklem büyüklüklerine göre güç oranlarının anlamlı farklılık gösterdiği tespit edilmiştir. Güç oranlarındaki değişkenliğin yaklaşık %24,8'i DMF belirleme yöntemleriyle, %48,3'ü ise örneklem büyüklükleri ile açıklandığı sonucuna ulaşılmıştır. Hem DMF belirleme yöntemlerinin hem de örneklem büyüklüklerinin güç oranları üzerinde yüksek etkiye sahip olduğu saptanmıştır. DMF belirleme yöntemleri üzerinden yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın Mantel-Haenszel ve Lord'un χ^2 ($\eta^2=0,124$); Mantel-Haenszel ve SIBTEST ($\eta^2=0,066$); Mantel-Haenszel ve Raju'nun alan ölçümü ($\eta^2=0,262$); Lord'un χ^2 ve Lojistik Regresyon ($\eta^2=0,114$); Lojistik

Regresyon ve SIBTEST ($\eta^2=0,059$); Lojistik Regresyon ve Raju'nun alan ölçümü ($\eta^2=0,246$) yöntemleri arasında olduğu saptanmıştır. Örneklem büyüklükleri üzerinden yapılan ikili karşılaştırmalar sonucunda ise 600-1200 ($\eta^2=0,191$); 600-2400 ($\eta^2=0,507$) ve 1200-2400 ($\eta^2=0,206$) arasında olduğu tespit edilmiştir. Beş DMF belirleme yöntemine ilişkin güç oranlarının örneklem büyüklüğü 600'den 2400'e yükseldikçe arttığı ortaya konulmuştur. Bu bulgu örneklem büyüklüğü arttıkça DMF belirlemede yöntemlerin güç oranlarının arttığını belirten alanyazındaki çalışma sonuçlarıyla tutarlılık göstermektedir (Carter, 2011; Tunç, 2016; Wang vd., 2013; Whitmore & Schumacker, 1999). Üç farklı örneklem büyüklüğünde de en yüksek güce sahip olan yöntemlerin Mantel-Haenszel ve Lojistik Regresyon olduğu; Raju'nun alan ölçümü yönteminin ise genel olarak daha düşük güce sahip olduğu belirlenmiştir.

Örneklem büyüklüğü oranlarına (1/1 ve 1/2) göre yapılan incelemeler doğrultusunda beş DMF belirleme yöntemine ilişkin güç oranlarının odak ve referans grubun örneklem büyüklükleri eşit iken (1/1) nispeten daha yüksek olduğu belirlenmiştir. Carter (2011), Herrera ve Gomez (2008), Atalay Kabasakal vd. (2014), Kristjansson vd. (2005), Tunç (2016), Wyse ve Mapuranga (2009) yaptıkları çalışmalarında odak ve referans grubun örneklem büyüklükleri birbirine eşitken elde edilen güç oranlarının daha yüksek olduğu sonucuna ulaşmışlardır. Bu bulgu alanyazındaki çalışma sonuçlarıyla paralellik göstermektedir. Ayrıca hem 1/1 hem de 1/2 örneklem oranlarında da gücü en düşük olan Lord'un χ^2 ve Raju'nun alan ölçümü yöntemleri; en yüksek ise Mantel-Haenszel ve Lojistik Regresyon yöntemleri olduğu sonucuna ulaşılmıştır. Araştırma sonuçlarına benzer şekilde Atalay Kabasakal vd. (2014) SIBTEST, Mantel-Haenszel ve MTK olabilirlik oran yöntemlerinin performanslarını (I. Tip hata ve güç oranları) karşılaştırmışlardır. Çalışma sonucuna eşit olmayan örneklem büyüklüklerinin eşit olan örneklem büyüklüklerine göre testin gücünün görece düşük olduğunu bulmuşlardır.

Araştırmada hem DMF büyüklüklerine (0,5 ve 1,0) göre hem de DMF belirleme yöntemlerine göre güç oranlarının anlamlı farklılık gösterdiği tespit edilmiştir. Güç oranlarındaki değişkenliğin yaklaşık %21,6'sının DMF belirleme yöntemleriyle, %39,8'inin ise DMF büyüklükleri ile açıklandığı sonucuna ulaşılmıştır. Hem DMF belirleme yöntemlerinin hem de DMF büyüklüklerinin güç oranları üzerinde yüksek etkiye sahip olduğu saptanmıştır. Yapılan ikili karşılaştırmalar neticesinde ise tespit edilen anlamlı farklılığın Mantel-Haenszel ve Lord'un χ^2 ($\eta^2=0,124$); Mantel-Haenszel ve SIBTEST ($\eta^2=0,066$);

Mantel-Haenszel ve Raju'nun alan ölçümü ($\eta^2=0,262$); Lord'un χ^2 ve Lojistik Regresyon ($\eta^2=0,114$); Lojistik Regresyon ve SIBTEST ($\eta^2=0,059$); Lojistik Regresyon ve Raju'nun alan ölçümü ($\eta^2=0,246$); SIBTEST ve Raju'nun alan ölçümü ($\eta^2=0,050$) yöntemleri arasında olduğu sonucuna ulaşılmıştır. Araştırmada DMF büyüklüğü 0,5'ten 1,0'e yükseldiğinde beş DMF belirleme yöntemine ilişkin güç oranlarının arttığı görülmüştür. Alanyazında DMF büyüklüğü arttıkça güç oranlarında bu doğrultuda yükseldiğini gösteren çalışmalar yer almaktadır (Garrett, 2009; Su & Wang, 2005; Tunç, 2016; Stark vd., 2006; Wang & Su, 2004). Stephens-Bonty (2008) Mantel-Haenszel ve Lojistik Regresyon yöntemlerini karşılaştırdığı çalışmada DMF büyüklüğü arttıkça güç oranlarında daha kabul edilebilir düzeylerde olduğu sonucuna ulaşmıştır. Tunç (2016) Mantel-Haenszel, LordDif ve ANOVA yöntemlerini karşılaştırmıştır. Çalışma sonucunda DMF büyüklüğü 0,25'ten 1,5'e yükseldikçe güç oranlarının arttığını ve DMF büyüklüğüne göre güç oranlarının anlamlı farklılık gösterdiğini bulmuştur. Ayrıca bu araştırmada hem 0,5 hem de 1,0 DMF büyüklüğünde gücü en düşük olan Lord'un χ^2 ve Raju'nun alan ölçümü yöntemleri; en yüksek olan ise Mantel-Haenszel ve Lojistik Regresyon yöntemleri olduğu belirlenmiştir.

Testte bulunan DMF'li madde oranı %5'ten %10'a yükseldiğinde beş DMF belirleme yöntemine ilişkin güç oranlarının arttığı sonucuna ulaşılmıştır. Fakat DMF'li madde oranlarının %10'dan %20'ye yükselmesi Raju'nun alan ölçümü dışındaki yöntemlerin güç oranlarında nispeten farklılığa sebep olmadığı ortaya konulmuştur. DMF'li madde oranları %5 ve %10 iken güç oranı en düşük olan Raju'nun alan ölçümü yöntemi; en yüksek olanlar ise Mantel-Haenszel ve Lojistik Regresyon yöntemleri olduğu sonucuna ulaşılmıştır. Bu bulgularla paralel olacak şekilde Wyse ve Mapuranga (2009) yaptıkları çalışmalarında DMF'li madde oranı %20'ye yükseldiği durumda elde edilen gücün %10'a göre daha düşük olduğu sonucuna ulaşmışlardır. Fidalgo, Mellenberg ve Muniz (2000), testte bulunan DMF'li madde oranı %10 iken Mantel-Haenszel yönteminin yeterli güce ulaştığı fakat %15 ve %30 oranlarında gücün yeterli olmadığı sonucuna ulaşmışlardır. Hidalgo ve Lopez Pina (2004), DMF belirlemede Lojistik Regresyon yönteminin güç oranlarının Mantel-Haenszel yöntemine göre daha yüksek olduğu sonucuna ulaşmışlardır.

Araştırmanın üçüncü alt probleminde DMF belirleme yöntemlerine göre elde edilen madde parametre kestirimlerine ait RMSE değerleri; örneklem büyüklükleri, odak ve referans grup

örneklem büyüklüğü oranları, DMF büyüklükleri ve testte bulunan DMF’li maddelerin oranları koşulları altında incelenmiştir.

Ayırt edicilik parametresine ait RMSE değerlerinin örneklem büyüklüğüne göre anlamlı farklılık gösterdiği sonucuna ulaşılmıştır. Ayrıca RMSE değerlerindeki değişkenliğin yaklaşık %98.6’sının örneklem büyüklükleri ile açıklandığı tespit edilmiştir. Yapılan ikili karşılaştırmalar neticesinde ise tespit edilen anlamlı farklılığın 600-1200 ($\eta^2=0,964$); 600-2400 ($\eta^2=0,990$) ve 1200-2400 ($\eta^2=0,948$) örneklem büyüklükleri arasında olduğu sonucuna ulaşılmıştır. Ayırt edicilik parametresine ilişkin RMSE değerlerinin örneklem büyüklüğü 600’den 2400’e yükseldikçe azaldığı saptanmıştır. Alanyazındaki çalışmalar incelendiğinde örneklem büyüklüğü arttıkça RMSE değerlerinde buna paralel olarak azaldığı bulunmuştur (Bolt & Lall, 2003; Çakıcı Eser, 2015; DeMars, 2003; Sünbül, 2011; Zhang, 2012). Buna rağmen örneklem büyüklüğü oranlarına, DMF büyüklüklerine ve testte bulunan DMF’li madde oranlarına göre RMSE değerlerinin anlamlı farklılık göstermediği sonucuna ulaşılmıştır. Ayrıca DMF belirleme yöntemlerine göre RMSE değerlerinin farklılaşmadığı belirlenmiştir.

Güçlük parametresine ait RMSE değerlerinin hem DMF belirleme yöntemlerine hem de örneklem büyüklüklerine göre anlamlı farklılık gösterdiği tespit edilmiştir. RMSE değerlerindeki değişkenliğin yaklaşık %82,1’inin örneklem büyüklüğü ile açıklandığı ve dolayısıyla örneklem büyüklüğünün RMSE değerleri üzerinde yüksek etkiye sahip olduğu belirlenmiştir. Ayrıca DMF yöntemlerinin RMSE değerleri üzerinde orta düzeyde bir etkiye sahip olduğu ve RMSE değerlerindeki değişkenliğin %6,6’sının yöntemlerle açıklandığı tespit edilmiştir. Yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın ise Mantel-Haenszel ve Raju’nun alan ölçümü ($\eta^2=0,018$); Lord’un χ^2 ve Lojistik Regresyon ($\eta^2=0,012$); Lojistik Regresyon ve Raju’nun alan ölçümü ($\eta^2=0,023$) yöntemleri arasında olduğu saptanmıştır. Örneklem büyüklükleri üzerinden yapılan ikili karşılaştırmalar sonucunda ise 600-1200 ($\eta^2=0,634$); 600-2400 ($\eta^2=0,821$) ve 1200-2400 ($\eta^2=0,752$) arasında olduğu tespit edilmiştir. Örneklem büyüklüğü 2400 iken ortalama RMSE değerleri en düşük Mantel-Haenszel ve Lojistik Regresyon yöntemlerinden; en yüksek ortalama RMSE değerleri ise SIBTEST, Raju’nun alan ölçümü ve Lord’un χ^2 yöntemlerinden elde edilmiştir. Ayrıca güçlük parametresine ait RMSE değerlerinin örneklem büyüklüğü 600’den 2400’e yükseldikçe azaldığı saptanmıştır. Alanyazındaki çalışmalar incelendiğinde örneklem büyüklüğü arttıkça RMSE değerlerinde buna paralel olarak azaldığı

bulunmuştur (Bolt & Lall, 2003; Çakıcı Eser, 2015; DeMars, 2003; Lee, 2012; Sünbül, 2011; Zhang, 2012). Buna rağmen örneklem büyüklüğü oranlarına, DMF büyüklüklerine ve testte bulunan DMF’li madde oranlarına göre RMSE değerlerinin anlamlı farklılık göstermediği sonucuna ulaşılmıştır.

Araştırmanın dördüncü alt problemde DMF belirleme yöntemlerine göre elde edilen yetenek parametre kestirimine ait RMSE değerleri; örneklem büyüklükleri, odak ve referans grup örneklem büyüklüğü oranları, DMF büyüklükleri ve testte bulunan DMF’li madde oranları koşulları altında incelenmiştir. Yetenek parametresine ait RMSE değerlerinin DMF belirleme yöntemlerine göre anlamlı farklılık göstermediği fakat örneklem büyüklüklerine göre anlamlı farklılık gösterdiği tespit edilmiştir. Örneklem büyüklüğünün RMSE değerleri üzerinde yüksek etkiye sahip olduğu ve RMSE değerlerindeki değişkenliğin yaklaşık %12,6’sını açıkladığı belirlenmiştir. Örneklem büyüklükleri üzerinden yapılan ikili karşılaştırmalar neticesinde 600-1200 ($\eta^2=0,083$); 600-2400 ($\eta^2=0,202$) ve 1200-2400 ($\eta^2=0,025$) arasında olduğu belirlenmiştir. Buna rağmen örneklem büyüklüğü oranlarına göre RMSE değerlerinin anlamlı farklılık göstermediği sonucuna ulaşılmıştır.

Yetenek parametresine ait RMSE değerlerinin DMF büyüklüklerine göre anlamlı farklılık gösterdiği tespit edilmiştir. DMF büyüklüklerinin RMSE değerleri üzerinde yüksek etkiye sahip olduğu ve RMSE değerlerindeki değişkenliğin yaklaşık %13,2’sini açıkladığı belirlenmiştir. Ayrıca DMF büyüklüğü 0,5’ten 1,0’a yükseldikçe RMSE değerlerinin artma eğiliminde olduğu sonucuna ulaşılmıştır. Alanyazındaki çalışmalar incelendiğinde DMF büyüklüğünün yetenek kestirimi üzerinde etkisinin olduğu bulunmuştur (Golia, 2014; Li & Zumbo, 2009; Lee & Zhang, 2017). Bu araştırma sonuçlarına benzer şekilde Li ve Zumbo (2009) DMF’li madde sayısına ve DMF büyüklüğüne göre DMF’li maddelerin teste dahil edilmesi ve dahil edilmemesi durumunun etkilerini araştırmışlardır. Yapılan çalışma sonucunda DMF büyüklüğü düşük düzeyde iken hatanın ve etki büyüklüğünün anlamlı şekilde değişmediği sonucuna ulaşmışlardır. Fakat DMF büyüklüğü arttığında hata miktarının arttığını tespit etmişlerdir. Golia (2014) ise 15 maddeden oluşan bir testte DMF’li maddelerin etkisini incelemiştir. DMF büyüklüğü arttığında yetenek kestiriminin etkilendiği sonucuna ulaşmıştır.

Bu araştırmada testte bulunan DMF’li madde oranlarına göre yetenek parametresine ait RMSE değerlerinin anlamlı farklılık gösterdiği belirlenmiştir. Testte bulunan DMF’li madde oranlarının RMSE değerleri üzerinde yüksek etkiye sahip olduğu ve RMSE değerlerindeki

değişkenliğin yaklaşık %16,7'sini açıkladığı belirlenmiştir. DMF'li madde oranları üzerinden yapılan ikili karşılaştırmalar neticesinde tespit edilen anlamlı farklılığın ise %5-%20 ($\eta^2=0,265$) ve %10-%20 ($\eta^2=0,076$) arasında olduğu saptanmıştır. Ayrıca DMF'li madde oranı %5'ten %20 ye yükseldikçe elde edilen yetenek parametresine ilişkin RMSE değerlerinin arttığı görülmüştür. Alanyazında testte bulunan DMF'li madde oranlarının da yetenek kestirimini etkilediği sonucuna ulaşan çalışmalara rastlanmıştır (Golia, 2010, 2014; Li & Zumbo, 2009; Lee & Zhang, 2017). Bu araştırma sonuçlarına benzer şekilde Lee ve Zhang (2017) DMF'li madde oranları ile orta ve yüksek düzey DMF'nin etkilerini incelemiştir. Çalışma sonucuna göre testte bulunan DMF'li madde oranı arttıkça yetenek kestirimlerinin de farklılık gösterdiğini bulmuşlardır. Fakat Golia (2010) tek biçimli DMF içeren üç maddenin testte tutulması ve çıkarılması durumunun yetenek kestirimi üzerindeki etkilerini değerlendirmiştir. Çalışma sonucunda birden fazla DMF'li maddenin teste dahil edilmesinin yetenek kestirimini olumsuz yönde etkilemediği sonucuna ulaşmıştır. Bu maddelerin testten çıkarılması durumunda ise yetenek kestiriminin olumsuz yönde etkilendiğini bulmuştur.

5.2. Öneriler

Bu araştırmadan elde edilen sonuçlara dayalı olarak araştırmacılara şu önerilerde bulunulabilir:

Araştırma Sonuçlarına Dayalı Öneriler

Bu araştırma sonucuna göre örneklem büyüklüğü 600 ve 1200 iken Mantel-Haenszel ve Lojistik Regresyon yöntemlerinin; 2400 iken ise Mantel-Haenszel, Lojistik Regresyon ve SIBTEST yöntemlerinin kullanılması önerilebilir. Odak ve referans grup örneklem büyüklüğü oranı 1/1 ve 1/2 iken Mantel-Haenszel, Lojistik Regresyon ve SIBTEST yöntemleri tercih edilebilir.

DMF büyüklüğü 0,5 iken Mantel-Haenszel ve Lojistik Regresyon; DMF büyüklüğü 1,0 iken ise Mantel-Haenszel, Lojistik Regresyon ve SIBTEST yöntemlerinin kullanılması önerilebilir.

Testte bulunan DMF’li madde oranı %5 ve %10 iken Mantel-Haenszel ve Lojistik Regresyon; %20 iken ise Mantel-Haenszel, Lojistik Regresyon ve SIBTEST yöntemleri tercih edilebilir.

Araştırmacılar İçin Öneriler

Araştırmada DMF belirleme yöntemlerine göre farklı koşullar altında madde ve yetenek parametre kestirimleri incelenmiştir. Araştırma kapsamında alanyazında yer alan çalışmalar dikkate alınarak üç farklı örneklem büyüklüğü (600, 1200 ve 2400), iki farklı örneklem büyüklüğü oranı (1/1 ve 1/2), iki farklı DMF büyüklüğü (0,5 ve 1,0) ve üç farklı DMF’li madde oranı (%5, %10 ve %20) koşulları belirlenmiştir. Fakat gerçek test uygulamalarında bu koşullar farklılık gösterebilir. Bu nedenle farklı koşullar ele alınarak benzer çalışmalar yürütülebilir.

Bu araştırmada iki kategoride puanlanan maddelere göre veri üretilmiştir. Çok kategorili puanlanan veya iki kategorili ve çok kategorili puanlanan verilerin birlikte ele alındığı benzer çalışmalar yapılabilir.

Bu araştırmada veri analizleri için Mantel-Haenszel, Lord’un χ^2 , Lojistik Regresyon, SIBTEST ve Raju’nun alan ölçümü yöntemleri ile sadece tek biçimli DMF türüne göre incelemeler yapılmıştır. Tek biçimli olmayan DMF türüne göre incelemeler yapılabilir.

Araştırmanın verileri 2PL modele ve tek biçimli DMF türüne göre üretilmiştir. Tek biçimli ve tek biçimli olmayan DMF türüne göre veriler üretilerek madde ve yetenek parametre kestirimleri incelenebilir.

Bu araştırmada veriler üretilirken test uzunluğu (40 madde) ve model türü (2PL) sabit koşul olarak ele alınmıştır. Farklı test uzunluklarında DMF’li maddelerin testten çıkarılmasının madde ve yetenek parametre kestirimleri üzerindeki etkileri incelenebilir.

KAYNAKLAR

- Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, 7(4), 255-278.
- AERA, APA, & NCME (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Akalın, Ş. (2014). *Kamu personeli seçme sınavı genel yetenek testinin madde yanlılığı açısından incelenmesi*. (Doktora tezi). <https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Akın Arkan, Ç., Uğurlu, S., & Atar, B. (2015). MIMIC, SIBTEST, lojistik regresyon ve mantel-haenszel yöntemleriyle gerçekleştirilen DMF ve yanlılık çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 31(1), 34-52.
- Angoff, W. H. (1993). *Perspectives on differential item functioning methodology*. Hillsdale, NJ: Erlbaum.
- Ankenmann, R. D., Witt, E. A., & Dunbar, S. B. (1999). An investigation of the power of the likelihood ratio goodness of fit statistic in detecting differential item functioning. *Journal of Educational Measurement*, 36, 277-300. <http://dx.doi.org/10.1111/j.1745-3984.1999.tb00558.x>
- Atalay Kabasakal, K., Arsan, N., Gök, B., & Kelecioğlu, H. (2014). Comparing performances (Type I error and power) of IRT likelihood ratio SIBTEST and Mantel-Haenszel methods in the determination of differential item functioning. *Educational Sciences: Theory & Practice*, 14(6), 2786-2193. <https://files.eric.ed.gov/fulltext/EJ1050496.pdf> sayfasından erişilmiştir.
- 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*, 43(43), 270-281. <https://dergipark.org.tr/tr/pub/hunefd/issue/7795/102030> sayfasından erişilmiştir.
- Atar, B. (2007). *Differential item functioning analyses for mixed response data IRT likelihood-ratio test, logistic regression, and GLLAMM procedures*. Doctoral Dissertation, The Florida State University, College of Education, Florida.
- Atar, B., & Kamata, A., (2011). Comparison of IRT likelihood ratio test and logistic regression DIF detection procedures. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 41, 36-47.
- Ateşok Deveci, N. (2008). *Üniversitelerarası kurul yabancı dil sınavının madde yanlılığı bakımından incelenmesi*. (Doktora tezi). <https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Awuor, R. A. (2008). *Effect of unequal sample sizes on the power of DIF detection: an IRT based monte carlo study with SIBTEST and Mantel-Haenszel procedures*. State University, Virginia Polytechnic Institute, Virginia.
- Ayan, C. (2011). *PISA 2009 fen okuryazarlığı alt testinin değişen madde fonksiyonu açısından incelenmesi*. (Yüksek lisans tezi). <https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Baker, F. (2001). *The basics of item response theory*. America: Eric.
- 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. <http://www.psicothema.com/pdf/3835.pdf> sayfasından erişilmiştir.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulanması*. Ankara: ÖSYM.
- Bolt, D. M. (2002). A Monte Carlo comparison of parametric and nonparametric polytomous DIF detection methods. *Applied Measurement in Education*, 15(2), 113-141. <https://eric.ed.gov/?id=EJ654822> sayfasından erişilmiştir.
- Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov Chain Monte Carlo. *Applied Psychological Measurement*, 27(6), 395-414.

- Bradley, J. V. (1978). Robustness? *The British Journal of Mathematical & Statistical Psychology*, 31(2), 144-152. <http://dx.doi.org/10.1111/j.2044-8317.1978.tb00581.x>
- Broer, M., Lee, Y., Rizavi, S., & Powers, D. (2005). *Ensuring The Fairness of GRE Writing Prompts: Assessing Differential Difficulty* (Report No: 02-07R). Princeton, NJ: Educational Testing Service.
- Brown, W. K., Cozby, C. P., Kee, D. W., & Worden, P. E.(1999). *Research methods in human development*. USA: Mayfield.
- Bulut, O., & Sünbül, Ö. (2017). R programlama dili ile Madde Tepki Kuramında monte carlo simülasyon çalışmaları. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(3), 266-287. <http://dx.doi.org/10.21031/epod.305821>
- Camilli, G., & Congdon, P. (1999). Application of a method of estimating DIF for polytomous test items. *Journal of Educational and Behavioral Statistics*, 24(4), 323-341. <http://dx.doi.org/10.2307/1165366>
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. California: Sage.
- Cappaert, K. (2014). *Dissecting the impact of DIF/DBF on ability estimation and person fit*. Doctoral Dissertation, The University of Wisconsin-Milwaukee in Educational Psychology.
- Carter, N. T. (2011). *Applications of differential functioning methods to the generalized graded unfolding model*. Doctoral Dissertation, Bowling Green State University, Psychology, ABD.
- Chaimongkol, S., Huffer, F. W., & Kamata, A. (2007). An explanatory differential item functioning (DIF) model by the WinBUG 1.4. *Songklanakarin Journal of Science and Technology*, 29, 449-458. https://rdo.psu.ac.th/sjstweb/journal/29-2/19item-response-theory_449-458.pdf sayfasından erişilmiştir.
- Chan, D. (2000). Detection of differential item functioning on the kirton adaptation-inventory using multiple-group mean and covariance structure analyses. *Multivariate Behavioral Research*, 35, 169-199. http://dx.doi.org/10.1207/S15327906MBR3502_2

- Chang, H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement*, 33(3), 333-353. <https://www.jstor.org/stable/1435376?seq=1> sayfasından erişilmiştir.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 31-44. <http://dx.doi.org/10.1111/j.1745-3992.1998.tb00619.x>
- Clauser, B., Mazor, K., & Hambleton, R. K. (1993). The effects of purification of the matching criterion on the identification of DIF using the Mantel-Haenszel procedure. *Applied Measurement in Education*, 6(4), 269-279. <https://eric.ed.gov/?id=EJ478666> sayfasından erişilmiştir.
- Cohen, A. S., Kim, S., & Wollack, J. A. (1996). An investigation of the likelihood ratio test for detection of differential item functioning. *Applied Psychological Measurement*, 20(1), 15-26. <http://dx.doi.org/10.1177/014662169602000102>
- Cohen, J. (1988). *Statistical power analysis for behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Crocker, L., & Algina, J. (1986). *Introduction to modern test theory*. California: Thomson Learning.
- Çakıcı Eser, D. (2015). *Çok boyutlu madde tepki kuramının farklı modellerinden çeşitli koşullar altında kestirilen parametrelerin incelenmesi*. (Doktora tezi). <https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Çepni, Z. (2011). *Değişen madde fonksiyonlarının SIBTEST, Mantel-Haenzsel, lojistik regresyon ve madde tepki kuramı yöntemleriyle incelenmesi*. (Doktora tezi). <https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Dainis, A. M. (2008). *Methods for identifying differential item and test functioning: an investigation of Type I error rates and power*. Doctoral Dissertation, James Madison University Assessment and Measurement, Harrisonburg, Virginia.
- De Ayala, R. J. (2009). *Theory and practice of item response theory*. New York: Guilford.
- DeMars, C. (2003). Sample size and the recovery of nominal response model item parameters. *Applied Psychological Measurement*, 27(4), 275-288.

- DeMars, C. E. (2009). Modification of the Mantel-Haenszel and logistic regression DIF procedures to incorporate the SIBTEST regression correction. *Journal of Educational and Behavioral Statistics*, 34, 149-170. <http://dx.doi.org/10.3102/1076998607313923>
- DeMars, C. E. (2010). *Item response theory. Understanding statistics measurement*. New York: Oxford.
- DeMars, C. E. (2010). Type I error inflation for detecting DIF in the presence of impact. *Educational and Psychological Measurement*, 70(6) 961-972. <http://dx.doi.org/10.1177/0013164410366691>
- 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. <http://dx.doi.org/10.14687/ijhs.v11i1.2798>
- 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.
- Duncan, S. C. (2006). *Improving the prediction of differential item functioning: A comparison of the use of an effect size for logistic regression DIF and Mantel-Haenszel DIF methods*. Doctoral Dissertation, Texas A&M University College Station Educational Psychology, Texas.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum.
- Erdem Keklik, D. (2012). *İki kategorili maddelerde tek biçimli değişen madde fonksiyonu belirleme tekniklerinin karşılaştırılması: Bir simülasyon çalışması*. (Doktora tezi). <https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Erdem, B. (2015). *Ortaöğretime geçişte kullanılan ortak sınavların değişen madde fonksiyonu açısından kitapçık türlerine göre farklı yöntemlerle incelenmesi*. (Yüksek lisans tezi). <https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Fidalgo, A. M., Ferreres, A., & Muniz, J. (2004). Utility of the Mantel-Haenszel procedure for detecting differential item functioning in small samples. *Educational and*

Psychological Measurement, 64(6), 925-936.
<http://dx.doi.org/10.1177/0013164404267288>

Fidalgo, A. M., Mellenberg, G. J., & Muniz, J. (2000). Effects of amount of DIF, test length, and purification type on robustness and power of Mantel-Haenszel procedures. *Methods of Psychological Research Online*, 5(3), 43-53.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.586.7639&rep=rep1&type=pdf> sayfasından erişilmiştir.

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. (2007). Detection of crossing differential item functioning a comparison four methods. *Educational Psychological Measurement*, 67(4), 565-582.
<http://dx.doi.org/10.1177/0013164406296975>

Finch, W. H., & French, B. F. (2008). Anomalous type I error rates for identifying one type of differential item functioning in the presence of the other. *Educational and Psychological Measurement*, 68, 742-759.
<http://dx.doi.org/10.1177/0013164407313370>

French, B. F., & Finch, W. H. (2013). Extensions of Mantel-Haenszel for multilevel DIF detection. *Educational and Psychological Measurement*, 73, 648-671.
<https://doi.org/10.1177/0013164412472341>

French, B. F., & Finch, W. H. (2015). Transforming SIBTEST to account for multilevel data structures. *Journal of Educational Measurement*, 52(2), 159-180.
<https://doi.org/10.1111/jedm.12071>

French, B. F., & Maller, S. J. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement*, 67 (3), 373-393. <https://doi.org/10.1177/0013164406294781>

Garrett, P. (2009). *A monte carlo study investigating missing data differential item functioning and effect size*. (Doctoral dissertation). Georgia State University Educational Policy Studies in the College of Education, Atlanta.

- Gierl, M. J. (2005). Using dimensionality-based DIF analyses to identify and interpret constructs that elicit group differences. *Educational Measurement: Issues and Practice*, 24(1), 3-14. <https://doi.org/10.1111/j.1745-3992.2005.00002.x>
- Gierl, M. J., Gotzmann, A., & Boughton, K. A. (2004). Performance of SIBTEST when the percent of DIF items is large. *Applied Measurement in Education*, 17(3), 241-264. https://doi.org/10.1207/s15324818ame1703_2
- 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.
- Gierl, M., Khaliq, S. N., & Boughton, K. (1999, June). *Gender differential item functioning in mathematics and science: Prevalence and policy implications*. Paper presented at the Symposium entitled "Improving Large-Scale Assessment in Education" at the Annual Meeting of the Canadian Society for the Study of Education, Canada.
- Golia, S. (2010). The assessment of DIF on Rasch measures with an application to job satisfaction. *Electronic Journal of Applied Statistical Analysis: Decision Support Systems and Services Evaluation*, 1(1) 16-25.
- Golia, S. (2014) Assessing the impact of uniform and nonuniform differential item functioning items on Rasch measure: The polytomous case. *Computational Statistics*, 30(2), 441-461. <https://link.springer.com/article/10.1007%2Fs00180-014-0542-x> sayfasından erişilmiştir.
- Gonzalez-Roma, V., Hernandez, A., & Gomez-Benito, J. (2006). Power and type I error of the mean and covariance structure analysis model for detecting differential item functioning in graded response items. *Multivariate Behavioral Research*, 41(1), 29-53. https://doi.org/10.1207/s15327906mbr4101_3
- Gök, B., Kelecioğlu, H., & Doğan, N. (2010). Değişen madde fonksiyonunu belirlemede mantel-haenszel ve lojistik regreyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 35(156), 3-16. <http://egitimvebilim.ted.org.tr/index.php/EB> sayfasından erişilmiştir.
- Güler, N., & Penfield, R. D. (2009). A comparison of the logistic regression and contingency table methods for simultaneous detection of uniform and nonuniform DIF. *Journal*

of *Educational Measurement*, 46(3), 314-329. <https://doi.org/10.1111/j.1745-3984.2009.00083.x>

Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (pp. 147-200). New York: American Council on Education and Macmillan.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory principles and applications*. Boston-USA: Kluwer-Nijhoff.

Hambleton, R. K., & Cook, L. L. (1983). Robustness of item response models and effects of test length and sample size on the precision of ability estimates. In D. J. Weiss (Ed.), *New horizons in testing* (pp. 31-49). NY: Academic.

Hambleton, R. K., & Jones R. W. (1993). An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47. <https://doi.org/10.1111/j.1745-3992.1993.tb00543.x>

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

Hambleton, R., Merenda, P. F., & Spielberger, C. D. (2017). *Eğitimde ve psikolojide kullanılan testlerin kültürlerarası değerlendirme amacıyla uyarlanması*. (N. Koç & A. Yıldırım, Çev.). Ankara: Pegem.

Han, K. T., & Hambleton, R. K. (2014). *User's manual for WinGen3*. Amherst, MA: University of Massachusetts.

Herrera, A., & Gomez, J. (2008). Influence of equal or unequal comparison group sample sizes on the detection of differential item functioning using the Mantel-Haenszel and logistic regression techniques. *Quality and Quantity*, 42(6), 739-755. <https://link.springer.com/article/10.1007%2Fs11135-006-9065-z> sayfasından erişilmiştir.

Higaldo, 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, 903-915. <https://psycnet.apa.org/record/2004-21302-001> sayfasından erişilmiştir.

- Holland P. W., & Wainer, H. (1993). *Differential item functioning*. New Jersey: Lawrence Erlbaum.
- Holland, P. W., & Thayer, D. T. (1985). An Alternative definition of the ETS delta scale of item difficulty. Research Report RR-85-43. Princeton, N-J: Educational Testing Service.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the mantel-haenszel procedure. In H. Holland & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14, 329-349. https://doi.org/10.1207/S15324818AME1404_2
- Johnson Frotman, K., A. (2007). *The evaluation of new criteria for polytomous DIF in the DFIT framework*. Doctoral Dissertation, Concordia University Illinois Institute of Technology, Chigago.
- Karasar, N. (1998). *Araştırmalarda rapor hazırlama yöntemi*. Ankara: Pars.
- Kılıç, A. F., Uysal, İ., & Doğan, N. (2018, Nisan). *The effect of replication number on generated data sets in simulation studies*. 27. Uluslararası Eğitim Bilimleri Kongresi'nde sunulmuş bildiri, Antalya.
- Kim, J. (2010). *Controlling type I error rate in evaluating differential item functioning for four DIF methods: Use of three procedures for adjustment of multiple item testing*. Doctoral Dissertation, Georgia State University, Department of Educational Policy Studies, Atlanta.
- Kim, S. H. (2000, April). *An investigation of the likelihood ratio test, the mantel test, and the generalized mantel haenszel test of DIF*. Paper Presented at the Annual Meeting of the American Educational Research Association, New Orleans.
- Kim, S. H., & Cohen, A. S. (1998). Detection of differential item functioning under the graded response model with the likelihood ratio test. *Applied Psychological Measurement*, 22(4), 345-355. <https://doi.org/10.1177/014662169802200403>

- Kim, S. H., Cohen, A. S., & Kim, H. O. (1994). An investigation of Lord's procedure for the detection of differential item functioning. *Applied Psychological Measurement*, 18(3), 217-228. <https://doi.org/10.1177/014662169401800303>
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746-759. <https://doi.org/10.1177/0013164496056005002>
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling and linking: Methods and practices*. New York: Springer.
- Koyuncu, İ., Aksu, G., & Kelecioğlu, H. (2018). Mantel-Haenszel, lojistik regresyon ve olabilirlik oranı değişen madde fonksiyonu inceleme yöntemlerinin farklı yazılımlar kullanılarak karşılaştırılması. *İlköğretim Online*, 17(2), 909-925. <http://ilkogretim-online.org.tr/index.php/io/article/view/2639> sayfasından erişilmiştir.
- Kristjansson, E. (2001). *Detecting DIF in polytomous items: An empirical comparison of the ordinal logistic regression, logistic discriminant function analysis, mantel and generalized mantel-haenszel procedures*. Doctoral Dissertation, University of Ottawa Faculty of Education, Ottawa.
- Kristjansson, E., Aylesworth, R., Mcdowell, I., & Zumbo, B. D. (2005). A comparison of four methods for detecting differential item functioning in ordered response items. *Educational and Psychological Measurement*, 65(6), 935-953. <https://doi.org/10.1177/0013164405275668>
- 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. <https://psycnet.apa.org/record/2009-07222-006> sayfasından erişilmiştir.
- Lee, Y.-H., & Zhang, J. (2017). Effects of differential item functioning on examinees' test performance and reliability of test. *International Journal of Testing*, 17(1), 23-54.
- Li, H., & Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika*, 61(4), 647-677. <https://link.springer.com/article/10.1007/BF02294041> sayfasından erişilmiştir.
- Li, Y., Brooks, G. P., & Johanson, G. A. (2012). Item discrimination and Type I error in the detection of differential item functioning. *Educational and Psychological*

- Measurement*, 72(5), 847-861. <https://journals.sagepub.com/home/epm> sayfasından erişilmiştir.
- Li, Z. (2009). *Impact of differential item functioning on statistical conclusions*. Doctoral Dissertation, University of British Columbia, Canada.
- Li, Z., & Zumbo, B. D. (2009). Impact of differential item functioning on subsequent statistical conclusions based on observed test score data. *Psicologica*: 30, 343-370. <https://www.redalyc.org/pdf/169/16911991011.pdf> sayfasından erişilmiştir.
- Lord, F. M. (1977). A study of item bias using item characteristic curve theory. In Y. H. Poortinga (Ed.), *Basic problems in cross-cultural psychology* (pp. 19-29). Amsterdam: Swets & Zeitlinger.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley.
- 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.
- Mazor, K. M., Clauser, B. E., & Hambleton, R. K. (1992). The effect of sample size on the functioning of the Mantel-Haenszel statistic. *Educational and Psychological Measurement*, 52, 443-451. <https://doi.org/10.1177/0013164492052002020>
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
- Millsap, R. E., & Meredith, W. (1992). Inferential conditions in statistical detection of measurement bias. *Applied Psychological Measurement*, 16, 389-402. <https://doi.org/10.1177/014662169201600411>
- 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. <https://doi.org/10.1177/014662169401800403>

- Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*, 20, 257-274. <https://doi.org/10.1177/014662169602000306>
- Osterlind S. J., & Everson H. T. (2009). *Differential item functioning*. California: Sage.
- Özdemir, B. (2015). A comparison of IRT-based methods for examining differential item functioning in TIMSS 2011 mathematics subtest. *Procedia Social and Behavioral Sciences*, 174, 2075-2083. <https://doi.org/10.1016/j.sbspro.2015.02.004>
- Özmen, D. T. (2013). *PISA 2009 okuma becerileri testi maddelerinin yanlılık açısından Türkiye, Amerika Birleşik Devletleri ve Birleşik Krallık uygulamalarında karşılaştırılması*. (Doktora tezi). <https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Pae, T. I. (2004). DIF for examinees with different academic backgrounds. *Language Testing*, 21(1), 53-73. <https://doi.org/10.1191/0265532204lt274oa>
- Paek, I., & Guo, H. (2011). Accuracy of DIF estimates and power in unbalanced designs using the Mantel-Haenszel DIF detection procedure. *Applied Psychological Measurement*, 35(7), 518-535. <https://doi.org/10.1177/0146621611420559>
- Perrone, M. (2006). Differential item functioning and item bias: Critical considerations in test fairness. *Teachers College Columbia University Working Papers in TESOL and Applied Linguistics*, 6(2), 1-3.
- Potenza, M. T., & Dorans, N. J. (1995). DIF assessment for polytomously scored items: Framework for classification and evaluation. *Applied Psychological Measurement*, 19(1), 23-37. <https://doi.org/10.1177/014662169501900104>
- Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, 53(4), 495-502.
- Raju, N., Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement*, 19(4), 353- 368. <https://doi.org/10.1177/014662169501900405>
- Reckase, M. D. (2009). *Multidimensional item response theory (statistics for social and behavioral sciences)*. New York: Springer.
- 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.
<https://doi.org/10.1177/014662169301700201>
- Roussos, L., & Stout, W. (1996). A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement*, 20, 355-371.
<https://doi.org/10.1177/014662169602000404>
- Roznowski, M., & Reith, J. (1999). Examining the measurement quality of tests containing differentially functioning items: Do biased items result in poor measurement? *Educational and Psychological Measurement*, 59(2), 248-269.
- Rupp, A. A., & Zumbo, B. D. (2003). Which model is best? Robustness properties to justify model choice among unidimensional IRT models under item parameter drift. *Alberta Journal of Educational Research*, 49, 264-276.
- Ryan, C. H. (2008). *Using hierarchical generalized linear modeling for detection of differential item functioning in a polytomous item response theory framework: An evaluation and comparison with generalized Mantel-Haenszel*. Doctoral Dissertation, Georgia State University, College of Education, Atlanta.
- Selvi, H. (2013). *Klasik test ve madde tepki kuramlarına dayalı değişen madde fonksiyonu belirleme tekniklerinin farklı puanlama durumlarında incelenmesi*. (Doktora tezi).
<https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58(2), 159-194.
- Shepard, L. A., Camilli, G., & Williams, D. M. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*, 22, 77-105.
<https://doi.org/10.1111/j.1745-3984.1985.tb01050.x>
- Sireci, S. G., & Rios, J. A. (2013). Decisions that make a difference in detecting differential item functioning. *Educational Research and Evaluation*, 19(2-3), 170-187.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, 91(6), 1292-1306.

- Stephens-Bonty, T. A. (2008). *Using three different categorical data analysis techniques to detect differential item functioning*. Doctoral Dissertation, Georgia State University College of Education, Atlanta.
- Su, Y. H., & Wang, W. C. (2005). Efficiency of the mantel, generalized mantel-haenszel, and logistic discriminant function analysis methods in detecting differential item functioning for polytomous items. *Applied Measurement in Education*, 18(4), 313-350.
- Sünbül, Ö. (2011). *Çeşitli boyutluluk özelliklerine sahip yapılarda madde parametrelerinin değişmezliğinin klasik test teorisi, tek boyutlu madde tepki kuramı, çok boyutlu madde tepki kuramı çerçevesinde incelenmesi*. (Doktora tezi). <https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Sünbül, Ö., & Ömür Sünbül, S. (2016). Değişen madde fonksiyonunun belirlenmesinde kullanılan yöntemlerde I. tip hata ve güç çalışması. *İlköğretim Online*, 15(3), 882-897.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370.
- Swaminathan, J., & Gifford, J. A. (1983). Estimation of parameters in the three-parameter latent trait model. In D. J. Weiss (Ed.), *New horizons in testing* (pp. 13-30). New York: Academic.
- 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. Baloglu, Çev.). Ankara: Nobel.
- Taylor, C. S., & Lee, Y. (2012). Gender DIF in reading and mathematics tests with mixed item formats. *Applied Measurement in Education*, 25, 246-280.
- Tennant, A., & Pallant, J. F. (2007). DIF matters: A practical approach to test if Differential Item Functioning makes a difference. *Rasch Measurement Transactions*, 20(4), 1082-1084.

- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 147-169). Hillside, NJ: Erlbaum.
- 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). Hillside, NJ: Erlbaum.
- Tunç, E. B. (2016). *İki ve çok kategorili puanlanan maddelerde değişen madde fonksiyonlarının karşılaştırılması*. (Doktora tezi). <https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Uzun, N. B., & Gelbal, S. (2017). PISA fen başarı testinin madde yanlılığının kültür ve dil açısından incelenmesi. *Kastamonu Eğitim Dergisi*, 25(6), 2427-2446.
- Vaughn, B. K., & Wang, Q. (2010). DIF trees: Using classification trees to detect differential item functioning. *Educational and Psychological Measurement*, 70, 941-952. <https://doi.org/10.1177/0013164410379326>
- Walker, C. M., Zhang, B., Banks, K., & Cappaert, K. (2012). Establishing effect size guidelines for interpreting the results for differential bundle functioning analyses using SIBTEST. *Educational and Psychological Measurement*, 72(3), 415-434.
- Wang, W. C., & Su, Y. H. (2004). Factors influencing the mantel and generalized Mantel-Haenszel methods for the assessment of differential item functioning in polytomous items. *Applied Psychological Measurement*, 28, 450-481. <https://doi.org/10.1177/0146621604269792>
- Wang, W. C., & Yeh, L. Y. (2003). Effects of anchor item methods on differential item functioning detection with the likelihood ratio test. *Applied Psychological Measurement*, 27(6), 479-498.
- Wang, W., Tay, L., & Drasgow, F. (2013). Detecting differential item functioning of polytomous items for an ideal point response process. *Applied Psychological Measurement*, 37(4), 316-335.
- Wells, C. S., Subkoviak, M. J., & Serlin, R. C. (2002). The effect of item parameter drift on examinee ability estimates. *Applied Psychological Measurement*, 26(1), 77-87.

- Wen, Y. (2014). *DIF analyses in multilevel data: Identification and effects on ability estimates*. Doctoral Dissertation, The University of Wisconsin Milwaukee, Educational Psychology.
- Whitmore, M. L., & Schumacker, R. E. (1999). A comparison of logistic regression and analysis of variance differential item functioning detection methods. *Educational and Psychological Measurement*, 59(6), 910-927.
- Wiberg, M. (2007). *Measuring and detecting differential item functioning in criterion-referenced licensing test: A theoretic comparison of methods*. EM No. 60. Umea University Department of Educational Measurement, Umea.
- Wood, W. S. (2011). *Differential item functioning procedures for polytomous items when examinee sample sizes are small*. Doctoral Dissertation, Graduate College of The University of Iowa, Iowa.
- Wright, K. D., & Oshima, T. C. (2015). An effect size measure for Raju's differential functioning for items and tests. *Educational and Psychological Measurement*, 75(2), 338-358.
- Wyse, A. E., & Mapuranga, R. (2009). Differential item functioning analysis using Rasch item information functions. *International Journal of Testing*, 9(4), 333-357. <https://doi.org/10.1080/15305050903352040>
- Zhang, B. (2008). Application of unidimensional item response models to tests with items sensitive to secondary dimension. *The Journal of Experimental Education*, 77(2), 147-166.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & 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*. Ottawa: National Defense.
- 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. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223-233.
- 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). A measure of DIF effect size using logistic regression procedures. *National Board of Medical Examiners, Philadelphia, PA*.
- Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 233-251. <https://www.jstor.org/stable/1435045?seq=1> sayfasından erişilmiştir.
- Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Descriptive and inferential procedures for assessing differential item functioning in polytomous items. *Applied Measurement in Education*, 10(4), 321-344. https://doi.org/10.1207/s15324818ame1004_2

EKLER

EK 1. 600 Örneklem Büyüklüğü için İlgili Koşullar Altında Ayırt Edicilik Parametresine İlişkin Standart Hata, RMSE ve Yanlılık Değerleri

Koşullar			DMF Belirleme Yöntemleri					
Örnekleme oranı	DMF büyüklüğü	DMF'li madde oranı		M-H	Lord'un χ^2	LR	SIBTEST	Raju'nun alan ölçümü
300/300	0,5	%5	SH	0,182748	0,182427	0,179217	0,184027	0,183006
			RMSE	0,184098	0,181013	0,180689	0,185326	0,184342
			BIAS	0,022257	0,022666	0,023022	0,021911	0,022153
		%10	SH	0,194745	0,189901	0,189969	0,190642	0,189436
			RMSE	0,198379	0,193345	0,193474	0,194207	0,192843
			BIAS	0,037798	0,036327	0,036654	0,037042	0,036088
		%20	SH	0,187861	0,185929	0,187449	0,185489	0,185150
			RMSE	0,191176	0,189269	0,190769	0,188800	0,188445
			BIAS	0,035445	0,035395	0,035436	0,035205	0,035084
	1,0	%5	SH	0,194799	0,191376	0,194380	0,191439	0,192767
			RMSE	0,201705	0,198141	0,200853	0,197580	0,199402
			BIAS	0,052327	0,051331	0,050580	0,048900	0,051009
		%10	SH	0,188784	0,186405	0,185687	0,188039	0,185737
			RMSE	0,191411	0,189033	0,187577	0,190537	0,188361
			BIAS	0,031607	0,031409	0,026561	0,030754	0,031331
		%20	SH	0,194762	0,191202	0,198047	0,196987	0,195149
			RMSE	0,197419	0,195057	0,200774	0,201071	0,198719
			BIAS	0,032282	0,038588	0,032979	0,040321	0,037499
200/400	0,5	%5	SH	0,182775	0,182379	0,182558	0,183235	0,183623
			RMSE	0,184365	0,184265	0,184178	0,185144	0,185459
			BIAS	0,024165	0,026291	0,024378	0,026519	0,026033
		%10	SH	0,188938	0,192568	0,191913	0,190425	0,192568
			RMSE	0,192320	0,195809	0,195101	0,193601	0,195809
			BIAS	0,035912	0,035475	0,035125	0,034923	0,035475
		%20	SH	0,185957	0,187661	0,187661	0,188356	0,188604
			RMSE	0,190798	0,192361	0,192361	0,193100	0,193199
			BIAS	0,042706	0,042259	0,042259	0,042539	0,041887
	1,0	%5	SH	0,178435	0,180033	0,182821	0,178699	0,186387
			RMSE	0,180323	0,182455	0,184660	0,180574	0,191378
			BIAS	0,026025	0,029631	0,025999	0,025951	0,043425
		%10	SH	0,191444	0,188652	0,191444	0,189171	0,191355
			RMSE	0,196737	0,193976	0,196737	0,194307	0,196537
			BIAS	0,045330	0,045137	0,045330	0,044381	0,044834
		%20	SH	0,195239	0,193328	0,198943	0,188244	0,196455
			RMSE	0,199018	0,197085	0,202802	0,192335	0,200009
			BIAS	0,038599	0,038303	0,039374	0,039459	0,037536

EK 2. 1200 Örneklem Büyüklüğü için İlgili Koşullar Altında Ayırt Edicilik Parametresine İlişkin Standart Hata, RMSE ve Yanlılık Değerleri

Koşullar			DMF Belirleme Yöntemleri					
Örnekleme oranı	DMF büyüklüğü	DMF'li madde oranı		M-H	Lord'un χ^2	LR	SIBTEST	Raju'nun alan ölçümü
600/600	0,5	%5	SH	0,130796	0,130028	0,132936	0,131187	0,131187
			RMSE	0,131871	0,131704	0,134623	0,132945	0,132945
			BIAS	0,016797	0,020946	0,021242	0,021550	0,021550
		%10	SH	0,130346	0,12864	0,129869	0,127979	0,130044
			RMSE	0,131926	0,13031	0,131552	0,129647	0,131670
			BIAS	0,020351	0,02071	0,020975	0,020728	0,020628
		%20	SH	0,138321	0,134909	0,132093	0,134908	0,134909
			RMSE	0,140438	0,137106	0,133504	0,137106	0,137106
			BIAS	0,024289	0,024447	0,019358	0,024448	0,024448
	1,0	%5	SH	0,128899	0,131923	0,129197	0,133417	0,133417
			RMSE	0,129773	0,134807	0,130075	0,136179	0,136179
			BIAS	0,015032	0,027737	0,015083	0,027287	0,027287
		%10	SH	0,127406	0,130793	0,127745	0,126974	0,152381
			RMSE	0,128296	0,132399	0,128618	0,128500	0,154995
			BIAS	0,015088	0,020563	0,014960	0,019748	0,028342
		%20	SH	0,135559	0,138330	0,133005	0,135014	0,136987
			RMSE	0,136509	0,141395	0,133878	0,136323	0,140165
			BIAS	0,01608	0,029278	0,015261	0,018847	0,029682
400/800	0,5	%5	SH	0,131197	0,131636	0,133027	0,132401	0,131197
			RMSE	0,134064	0,134521	0,136029	0,135198	0,134064
			BIAS	0,027577	0,027711	0,028423	0,027358	0,027577
		%10	SH	0,135222	0,135625	0,134475	0,132415	0,133335
			RMSE	0,139229	0,140089	0,139338	0,137459	0,138117
			BIAS	0,033162	0,035086	0,036494	0,036899	0,036025
		%20	SH	0,128804	0,131382	0,132760	0,131043	0,129514
			RMSE	0,130349	0,133617	0,135035	0,133341	0,131663
			BIAS	0,020010	0,024337	0,024682	0,024651	0,023690
	1,0	%5	SH	0,126294	0,126294	0,126403	0,128206	0,132792
			RMSE	0,127527	0,127526	0,127678	0,129455	0,139594
			BIAS	0,017685	0,017685	0,018003	0,017941	0,043046
		%10	SH	0,135109	0,132317	0,134033	0,128367	0,133402
			RMSE	0,138923	0,135901	0,137881	0,129716	0,136944
			BIAS	0,032328	0,031003	0,032347	0,018657	0,030947
		%20	SH	0,140717	0,139634	0,140065	0,139545	0,137014
			RMSE	0,142688	0,141648	0,142059	0,141479	0,139063
			BIAS	0,023632	0,023802	0,023716	0,023311	0,023784

EK 3. 2400 Örneklem Büyüklüğü için İlgili Koşullar Altında Ayırt Edicilik Parametresine İlişkin Standart Hata, RMSE ve Yanlılık Değerleri

Koşullar			DMF Belirleme Yöntemleri					
Örnekleme oranı	DMF büyüklüğü	DMF'li madde oranı		M-H	Lord'un χ^2	LR	SIBTEST	Raju'nun alan ölçümü
1200/1200	0,5	%5	SH	0,091655	0,088960	0,092389	0,090167	0,090609
			RMSE	0,092319	0,089678	0,093076	0,090846	0,091687
			BIAS	0,011048	0,011321	0,011289	0,011085	0,014019
		%10	SH	0,092645	0,093449	0,092312	0,092645	0,092529
			RMSE	0,093466	0,094577	0,093057	0,093466	0,093660
			BIAS	0,012361	0,014561	0,011747	0,012361	0,014509
		%20	SH	0,097071	0,095546	0,096305	0,096121	0,094406
			RMSE	0,098596	0,097029	0,097787	0,097612	0,095889
			BIAS	0,017273	0,016902	0,016957	0,016994	0,016801
	1,0	%5	SH	0,092986	0,093898	0,094908	0,094076	0,093898
			RMSE	0,094254	0,095135	0,096121	0,095371	0,095135
			BIAS	0,015405	0,015295	0,015224	0,015663	0,015295
		%10	SH	0,093667	0,092749	0,091595	0,093667	0,095676
			RMSE	0,094346	0,093411	0,092275	0,094346	0,097041
			BIAS	0,011299	0,011091	0,011175	0,011299	0,016219
		%20	SH	0,100580	0,106448	0,101719	0,100580	0,098320
			RMSE	0,101145	0,108496	0,102185	0,101145	0,100348
			BIAS	0,010671	0,020979	0,009743	0,010671	0,020072
800/1600	0,5	%5	SH	0,090114	0,089903	0,090114	0,092076	0,092077
			RMSE	0,090851	0,090612	0,090851	0,093752	0,093752
			BIAS	0,011546	0,011311	0,011546	0,017643	0,017643
		%10	SH	0,094315	0,092878	0,092226	0,092369	0,092226
			RMSE	0,096167	0,094805	0,094110	0,094317	0,094110
			BIAS	0,018779	0,019018	0,018739	0,019069	0,018739
		%20	SH	0,094407	0,096417	0,094739	0,095561	0,096665
			RMSE	0,095053	0,098538	0,095381	0,096071	0,098831
			BIAS	0,011065	0,020334	0,011043	0,009891	0,020583
	1,0	%5	SH	0,092686	0,095276	0,093205	0,092252	0,095465
			RMSE	0,093301	0,097417	0,093806	0,092856	0,097593
			BIAS	0,010689	0,020314	0,010601	0,010579	0,020271
		%10	SH	0,092669	0,092855	0,092669	0,093117	0,097598
			RMSE	0,093327	0,093465	0,093327	0,093749	0,099302
			BIAS	0,011053	0,010667	0,011053	0,010875	0,018319
		%20	SH	0,095799	0,095399	0,096978	0,096747	0,095238
			RMSE	0,096326	0,096017	0,097611	0,097422	0,096856
			BIAS	0,010065	0,010877	0,011095	0,011441	0,017627

EK 4. 600 Örneklem Büyüklüğü için İlgili Koşullar Altında Güçlük Parametresine İlişkin Standart Hata, RMSE ve Yanlılık Değerleri

Koşullar			DMF Belirleme Yöntemleri					
Örneklem oranı	DMF büyüklüğü	DMF'li madde oranı		M-H	Lord'un χ^2	LR	SIBTEST	Raju'nun alan ölçümü
300/300	0,5	%5	SH	0,113248	0,112817	0,114632	0,114178	0,113019
			RMSE	0,147590	0,146314	0,147029	0,149912	0,148499
			BIAS	0,094645	0,093167	0,092071	0,097144	0,096325
		%10	SH	0,099469	0,099368	0,097495	0,099349	0,100809
			RMSE	0,108299	0,107577	0,106206	0,108027	0,109285
			BIAS	0,042832	0,041218	0,042125	0,042421	0,042198
		%20	SH	0,113610	0,114357	0,112705	0,113226	0,115194
			RMSE	0,154545	0,156644	0,155333	0,152539	0,155960
			BIAS	0,104770	0,107051	0,106893	0,102217	0,105138
	1,0	%5	SH	0,106068	0,100128	0,102890	0,099457	0,101676
			RMSE	0,113993	0,107942	0,111242	0,107962	0,109684
			BIAS	0,041761	0,040322	0,042289	0,042001	0,041142
		%10	SH	0,119886	0,124953	0,103829	0,121812	0,122121
			RMSE	0,157111	0,158663	0,110745	0,159303	0,158234
			BIAS	0,101544	0,097779	0,038522	0,102662	0,100619
		%20	SH	0,096777	0,118946	0,096194	0,134084	0,121861
			RMSE	0,101577	0,172121	0,101219	0,213422	0,182205
			BIAS	0,030857	0,124409	0,031499	0,166043	0,135457
200/400	0,5	%5	SH	0,097771	0,115199	0,100139	0,114370	0,116755
			RMSE	0,103351	0,155689	0,105687	0,152704	0,157297
			BIAS	0,033501	0,104729	0,033793	0,101183	0,105406
		%10	SH	0,113919	0,113327	0,114429	0,113731	0,113327
			RMSE	0,148092	0,147892	0,148039	0,147020	0,147892
			BIAS	0,094623	0,095021	0,093923	0,093167	0,095021
		%20	SH	0,110843	0,112079	0,112079	0,110559	0,111417
			RMSE	0,121900	0,122962	0,122962	0,121326	0,122155
			BIAS	0,050729	0,050576	0,050576	0,049969	0,050082
	1,0	%5	SH	0,098067	0,099116	0,099904	0,099292	0,103726
			RMSE	0,101339	0,102393	0,103229	0,102576	0,113152
			BIAS	0,025547	0,025700	0,025991	0,025748	0,045213
		%10	SH	0,107671	0,109791	0,107671	0,114014	0,112329
			RMSE	0,127890	0,129614	0,127890	0,131988	0,133528
			BIAS	0,069013	0,068889	0,069013	0,066496	0,072192
		%20	SH	0,113053	0,111218	0,113787	0,109031	0,112081
			RMSE	0,132172	0,130405	0,137867	0,127684	0,131799
			BIAS	0,068474	0,068088	0,077846	0,066449	0,069346

EK 5. 1200 Örneklem Büyüklüğü için İlgili Koşullar Altında Güçlük Parametresine İlişkin Standart Hata, RMSE ve Yanlılık Değerleri

Koşullar			DMF Belirleme Yöntemleri					
Örnekleme oranı	DMF büyüklüğü	DMF'li madde oranı		M-H	Lord'un χ^2	LR	SIBTEST	Raju'nun alan ölçümü
600/600	0,5	%5	SH	0,069231	0,073019	0,074165	0,073251	0,073251
			RMSE	0,072927	0,081439	0,082924	0,081948	0,081948
			BIAS	0,022922	0,022922	0,037092	0,036739	0,036739
		%10	SH	0,075159	0,075469	0,075571	0,074738	0,075325
			RMSE	0,084397	0,084650	0,084588	0,083427	0,084749
			BIAS	0,038394	0,038342	0,038002	0,037073	0,038842
		%20	SH	0,075617	0,074541	0,068841	0,074541	0,074541
			RMSE	0,084814	0,083622	0,071870	0,083622	0,083622
			BIAS	0,038412	0,037899	0,020648	0,037899	0,037899
	1,0	%5	SH	0,070455	0,077044	0,071214	0,076079	0,076079
			RMSE	0,072950	0,084725	0,073709	0,083579	0,083579
			BIAS	0,018915	0,035251	0,019015	0,034602	0,034602
		%10	SH	0,074909	0,084275	0,075272	0,082889	0,093212
			RMSE	0,077885	0,103439	0,078545	0,102862	0,116792
			BIAS	0,021324	0,059979	0,022437	0,060909	0,070369
		%20	SH	0,073241	0,077899	0,070994	0,073987	0,077789
			RMSE	0,076622	0,083724	0,074475	0,077152	0,084395
			BIAS	0,022512	0,030683	0,022504	0,021872	0,032732
400/800	0,5	%5	SH	0,079814	0,080393	0,080453	0,080344	0,079814
			RMSE	0,092798	0,092745	0,092385	0,093958	0,092798
			BIAS	0,047341	0,046244	0,045412	0,048714	0,047342
		%10	SH	0,077537	0,076852	0,078571	0,078719	0,077344
			RMSE	0,085679	0,083221	0,085239	0,084882	0,083366
			BIAS	0,036452	0,031929	0,033049	0,031752	0,031108
		%20	SH	0,069956	0,079254	0,077016	0,077440	0,076166
			RMSE	0,072844	0,090297	0,088911	0,087403	0,086079
			BIAS	0,020307	0,043269	0,044427	0,040525	0,040104
	1,0	%5	SH	0,074816	0,074816	0,072939	0,072146	0,075931
			RMSE	0,077248	0,077248	0,075102	0,074447	0,080011
			BIAS	0,019228	0,019228	0,017897	0,018366	0,025224
		%10	SH	0,081396	0,079712	0,079206	0,074965	0,081754
			RMSE	0,088978	0,087432	0,086917	0,078053	0,089371
			BIAS	0,035940	0,035920	0,035790	0,021736	0,036105
		%20	SH	0,083799	0,082449	0,082098	0,082036	0,078427
			RMSE	0,102919	0,101152	0,100582	0,102048	0,096549
			BIAS	0,059749	0,058599	0,058108	0,060695	0,056313

EK 6. 2400 Örneklem Büyüklüğü için İlgili Koşullar Altında Güçlük Parametresine İlişkin Standart Hata, RMSE ve Yanlılık Değerleri

Koşullar			DMF Belirleme Yöntemleri					
Örneklem oranı	DMF büyüklüğü	DMF'li madde oranı		M-H	Lord'un χ^2	LR	SIBTEST	Raju'nun alan ölçümü
1200/1200	0,5	%5	SH	0,052481	0,049935	0,052192	0,052031	0,055019
			RMSE	0,054544	0,052047	0,054349	0,054037	0,061884
			BIAS	0,014858	0,014676	0,015161	0,014588	0,028329
		%10	SH	0,048533	0,054127	0,048978	0,048533	0,053637
			RMSE	0,050479	0,061198	0,050827	0,050479	0,060639
			BIAS	0,013885	0,028557	0,013586	0,013885	0,028288
		%20	SH	0,053844	0,054411	0,055162	0,053807	0,055139
			RMSE	0,058509	0,058655	0,059519	0,057795	0,058865
			BIAS	0,022896	0,021905	0,022356	0,021097	0,020609
	1,0	%5	SH	0,057404	0,057732	0,058434	0,057599	0,057732
			RMSE	0,073213	0,073156	0,073663	0,072887	0,073156
			BIAS	0,045441	0,044931	0,044852	0,044662	0,044931
		%10	SH	0,050762	0,051212	0,050845	0,050762	0,056982
			RMSE	0,053836	0,054028	0,053639	0,053836	0,068811
			BIAS	0,017933	0,017217	0,017085	0,017933	0,038574
		%20	SH	0,055598	0,062285	0,056456	0,055598	0,057202
			RMSE	0,058336	0,069429	0,059295	0,058337	0,063917
			BIAS	0,017665	0,030674	0,018127	0,017665	0,028518
800/1600	0,5	%5	SH	0,051949	0,051271	0,051949	0,055303	0,055303
			RMSE	0,054119	0,053441	0,054119	0,060845	0,060845
			BIAS	0,015167	0,015072	0,015167	0,025371	0,025371
		%10	SH	0,052902	0,053714	0,053249	0,053406	0,053249
			RMSE	0,057989	0,058521	0,057966	0,058179	0,057966
			BIAS	0,023750	0,023228	0,022905	0,023081	0,022905
		%20	SH	0,048210	0,052083	0,045933	0,050379	0,052049
			RMSE	0,050555	0,058565	0,047852	0,052701	0,059002
			BIAS	0,015218	0,026782	0,013417	0,015468	0,027788
	1,0	%5	SH	0,047841	0,052604	0,047839	0,048123	0,052139
			RMSE	0,050011	0,062357	0,050164	0,050423	0,061497
			BIAS	0,014570	0,033484	0,015096	0,015055	0,032609
		%10	SH	0,049916	0,049752	0,049916	0,050107	0,057489
			RMSE	0,052362	0,052144	0,052362	0,052702	0,068195
			BIAS	0,015816	0,015612	0,015816	0,016334	0,036679
		%20	SH	0,049157	0,050891	0,051199	0,050738	0,056984
			RMSE	0,051186	0,052999	0,053383	0,053204	0,068854
			BIAS	0,014269	0,014798	0,015109	0,016009	0,038647

EK 7. 600 Örneklem Büyüklüğü için İlgili Koşullar Altında Yetenek Parametresine İlişkin Standart Hata, RMSE ve Yanlılık Değerleri

Koşullar			DMF Belirleme Yöntemleri					Raju'nun alan ölçümü
Örneklem oranı	DMF büyüklüğü	DMF'li madde oranı		M-H	Lord'un χ^2	LR	SIBTEST	
300/300	0,5	%5	SH	0,258338	0,255617	0,263163	0,263949	0,260377
			RMSE	0,261724	0,259438	0,266237	0,266912	0,263641
			BIAS	0,041959	0,044361	0,040339	0,039654	0,041361
		%10	SH	0,262308	0,253743	0,263103	0,252324	0,251747
			RMSE	0,265404	0,257498	0,267347	0,255903	0,255328
			BIAS	0,040422	0,043814	0,047449	0,042649	0,042617
		%20	SH	0,272039	0,269387	0,270362	0,263332	0,267259
			RMSE	0,273932	0,271433	0,272457	0,265588	0,269352
			BIAS	0,032139	0,033265	0,033722	0,034541	0,033509
	1,0	%5	SH	0,269724	0,252753	0,264821	0,255080	0,258136
			RMSE	0,271987	0,256146	0,267352	0,258652	0,260980
			BIAS	0,035011	0,041551	0,036696	0,042836	0,038423
		%10	SH	0,267887	0,271139	0,277961	0,268973	0,268689
			RMSE	0,270581	0,273303	0,280412	0,271606	0,271280
			BIAS	0,038088	0,034324	0,036996	0,037729	0,037399
		%20	SH	0,276510	0,268133	0,278201	0,296969	0,269651
			RMSE	0,278732	0,270239	0,280419	0,298431	0,271656
			BIAS	0,035119	0,033679	0,035201	0,029510	0,032944
200/400	0,5	%5	SH	0,264124	0,259653	0,262900	0,255343	0,262885
			RMSE	0,267276	0,263068	0,266158	0,258741	0,266049
			BIAS	0,040921	0,042251	0,041516	0,041797	0,040910
		%10	SH	0,260729	0,256217	0,254554	0,250602	0,256217
			RMSE	0,263537	0,259429	0,257666	0,254224	0,259429
			BIAS	0,038363	0,040697	0,039923	0,042763	0,040697
		%20	SH	0,261590	0,264677	0,264677	0,261590	0,259877
			RMSE	0,264587	0,267394	0,267394	0,262056	0,262929
			BIAS	0,039711	0,038022	0,038022	0,041599	0,039948
	1,0	%5	SH	0,270192	0,262775	0,272531	0,267079	0,264568
			RMSE	0,273423	0,265438	0,275044	0,270114	0,266749
			BIAS	0,041908	0,037503	0,037099	0,040377	0,034043
		%10	SH	0,267198	0,260789	0,267198	0,263301	0,262899
			RMSE	0,269766	0,263461	0,269766	0,266115	0,265396
			BIAS	0,037134	0,037428	0,037134	0,038599	0,036318
		%20	SH	0,277879	0,272837	0,281429	0,271907	0,275475
			RMSE	0,279799	0,275120	0,283694	0,274890	0,277601
			BIAS	0,032722	0,035371	0,035784	0,040391	0,034294

EK 8. 1200 Örneklem Büyüklüğü için İlgili Koşullar Altında Yetenek Parametresine İlişkin Standart Hata, RMSE ve Yanlılık Değerleri

Koşullar			DMF Belirleme Yöntemleri					
Örneklem oranı	DMF büyüklüğü	DMF’li madde oranı		M-H	Lord’un χ^2	LR	SIBTEST	Raju’nun alan ölçümü
600/600	0,5	%5	SH	0,264888	0,254279	0,263566	0,261073	0,261073
			RMSE	0,267407	0,257426	0,265996	0,263729	0,263729
			BIAS	0,036619	0,040127	0,035875	0,037335	0,037335
		%10	SH	0,269021	0,262321	0,267890	0,261925	0,265102
			RMSE	0,271793	0,265434	0,270756	0,265159	0,268046
			BIAS	0,038720	0,040531	0,039286	0,041288	0,039621
		%20	SH	0,275525	0,263098	0,268593	0,263098	0,263098
			RMSE	0,277359	0,265890	0,271777	0,265890	0,265890
			BIAS	0,031849	0,038435	0,041478	0,038435	0,038435
	1,0	%5	SH	0,271056	0,267406	0,274251	0,272270	0,272270
			RMSE	0,273593	0,269749	0,276633	0,274508	0,274508
			BIAS	0,037169	0,035474	0,036228	0,034978	0,034978
		%10	SH	0,286601	0,295739	0,288742	0,279771	0,405951
			RMSE	0,288640	0,298116	0,290974	0,282356	0,416839
			BIAS	0,034247	0,037569	0,035974	0,038117	0,094648
		%20	SH	0,298519	0,280612	0,286869	0,290495	0,284699
			RMSE	0,300591	0,282346	0,289301	0,292340	0,286623
			BIAS	0,035231	0,031239	0,037429	0,032794	0,033159
400/800	0,5	%5	SH	0,266935	0,265949	0,269592	0,268892	0,266935
			RMSE	0,269274	0,268343	0,271719	0,271299	0,269274
			BIAS	0,035418	0,035762	0,033934	0,036059	0,035418
		%10	SH	0,276235	0,261634	0,273491	0,267977	0,260635
			RMSE	0,278795	0,264451	0,275669	0,270495	0,263559
			BIAS	0,037695	0,038492	0,034588	0,036826	0,039149
		%20	SH	0,266364	0,275549	0,278499	0,266262	0,267309
			RMSE	0,268883	0,277817	0,280715	0,268583	0,270109
			BIAS	0,036725	0,035423	0,035196	0,035233	0,038783
	1,0	%5	SH	0,265607	0,265607	0,266951	0,269238	0,264879
			RMSE	0,268002	0,268002	0,269355	0,271476	0,267124
			BIAS	0,035748	0,035748	0,035906	0,034790	0,034557
		%10	SH	0,272770	0,266399	0,268979	0,267501	0,269912
			RMSE	0,275007	0,268943	0,271498	0,270263	0,272177
			BIAS	0,035006	0,036899	0,036891	0,038538	0,035039
		%20	SH	0,295879	0,289718	0,293598	0,287489	0,287590
			RMSE	0,297307	0,291433	0,295597	0,289764	0,291137
			BIAS	0,029101	0,031574	0,034317	0,036234	0,045302

EK 9. 2400 Örneklem Büyüklüğü İçin İlgili Koşullar Altında Yetenek Parametresine İlişkin Standart Hata, RMSE ve Yanlılık Değerleri

Koşullar			DMF Belirleme Yöntemleri					
Örneklem oranı	DMF büyüklüğü	DMF'li madde oranı		M-H	Lord'un χ^2	LR	SIBTEST	Raju'nun alan ölçümü
1200/1200	0,5	%5	SH	0,276012	0,262184	0,283206	0,268329	0,263618
			RMSE	0,277897	0,265083	0,285036	0,270821	0,266238
			BIAS	0,032314	0,039095	0,032252	0,036660	0,037263
		%10	SH	0,271546	0,266594	0,272106	0,271546	0,262730
			RMSE	0,274121	0,268697	0,274748	0,274121	0,265172
			BIAS	0,037485	0,033552	0,038009	0,037485	0,035902
		%20	SH	0,298256	0,287279	0,292628	0,294277	0,281611
			RMSE	0,300544	0,289429	0,294669	0,296392	0,283706
			BIAS	0,037015	0,035209	0,034625	0,035341	0,034407
	1,0	%5	SH	0,275701	0,263347	0,270035	0,276738	0,263347
			RMSE	0,278969	0,266498	0,272999	0,280394	0,266498
			BIAS	0,042579	0,040855	0,040113	0,045134	0,040855
		%10	SH	0,278518	0,273051	0,270455	0,278518	0,279673
			RMSE	0,281665	0,276046	0,273380	0,281665	0,283024
			BIAS	0,041983	0,040553	0,039888	0,041983	0,043424
		%20	SH	0,338543	0,345456	0,347700	0,338543	0,311564
			RMSE	0,339406	0,346220	0,348834	0,339406	0,313935
			BIAS	0,024193	0,022995	0,028103	0,024193	0,038504
800/1600	0,5	%5	SH	0,266485	0,262822	0,266485	0,265842	0,265842
			RMSE	0,268897	0,265466	0,268897	0,268118	0,268118
			BIAS	0,035934	0,037370	0,035934	0,034861	0,034861
		%10	SH	0,270679	0,270675	0,266906	0,270201	0,266906
			RMSE	0,272964	0,273112	0,269471	0,272584	0,269471
			BIAS	0,035244	0,036401	0,037093	0,035962	0,037093
		%20	SH	0,275778	0,267511	0,280020	0,285673	0,267606
			RMSE	0,278322	0,269949	0,282863	0,287391	0,270312
			BIAS	0,037549	0,036200	0,040003	0,031384	0,038148
	1,0	%5	SH	0,265300	0,265745	0,266692	0,256926	0,264386
			RMSE	0,268471	0,269023	0,269832	0,260405	0,267428
			BIAS	0,041139	0,041869	0,041041	0,042419	0,040219
		%10	SH	0,275374	0,276822	0,275374	0,278378	0,309832
			RMSE	0,278479	0,280107	0,278479	0,281266	0,316215
			BIAS	0,041478	0,042771	0,041478	0,040199	0,063215
		%20	SH	0,321091	0,301218	0,313033	0,317079	0,291715
			RMSE	0,323135	0,303752	0,315082	0,318767	0,294520
			BIAS	0,036283	0,039152	0,035870	0,032763	0,040557

EK 10. İlgili Koşullar Altında Yöntemlerin Belirlediği DMF’li Madde Oranları

Koşullar			DMF belirleme yöntemleri										Raju'nun alan ölçümü
			M-H		Lord'un χ^2		LR		SIBTEST				
Örneklem büyüklüğü	Örneklem oranı	DMF'li madde oranı	DMF büyüklüğü		DMF büyüklüğü		DMF büyüklüğü		DMF büyüklüğü		DMF büyüklüğü		
			0,5	1,0	0,5	1,0	0,5	1,0	0,5	1,0	0,5	1,0	
600	300/300	%5	2,1	2,9	1,6	2,5	2,6	3,4	1,7	2,2	1,8	4,2	
		%10	2,9	4,9	2,3	5,1	3,4	5,4	2,1	2,9	2,6	5,7	
		%20	4,9	9,7	2,8	8,4	5,2	9,9	3,1	6,5	4,3	8,8	
	200/400	%5	2,3	2,7	1,9	2,4	2,9	3,6	1,8	2,2	2,2	3,3	
		%10	2,7	4,7	2,0	3,9	3,5	5,3	1,8	3,0	2,5	4,5	
		%20	4,2	9,3	2,6	7,1	4,5	9,4	2,5	4,9	3,3	7,7	
1200	600/600	%5	2,9	3,5	2,0	3,2	3,2	3,8	2,5	2,6	2,9	4,1	
		%10	4,3	5,9	3,1	6,2	4,7	6,2	3,8	5,1	3,4	7,2	
		%20	7,5	13,1	4,7	11,0	7,0	12,4	6,2	11,2	5,4	10,2	
	400/800	%5	2,5	3,5	1,9	2,9	3,0	3,8	2,6	2,9	2,1	2,8	
		%10	4,1	5,9	2,4	5,3	4,1	6,1	3,4	5,2	2,5	5,6	
		%20	6,2	12,7	3,8	10,6	6,4	12,3	5,5	11,1	3,9	9,8	
2400	1200/1200	%5	3,1	4,1	2,6	4,4	3,4	4,2	3,2	3,9	2,6	6,8	
		%10	5,5	7,4	3,9	7,6	5,6	7,1	5,1	6,6	3,9	8,2	
		%20	10,7	18,6	7,7	16,4	9,8	16,7	10,1	16,9	7,7	14,2	
	800/1600	%5	3,0	4,1	2,4	3,9	3,5	4,2	3,1	3,9	2,1	3,7	
		%10	5,3	7,1	4,1	6,9	5,2	6,8	5,4	6,7	3,4	5,7	
		%20	10,2	17,4	7,1	14,9	9,6	15,7	9,6	15,7	6,6	11,6	

Not: DMF’li madde oranları: Her bir koşul için 100 replikasyona ait ortalama DMF’li madde sayılarıdır.

EK 11. Veri Üretmede Kullanılan Syntax (Birinci Koşul İçindir)

```
library(norm)
library(mice)
library(eRm)
library(mirt)
library(ltm)
library(irtos)
library(classInt)
library(difR)
library(DFIT)
library(lordif)
ni=40 #testte bulunan madde sayısı
t=1
ns1=300 #odak grup için örneklem büyüklüğü
ns2=300 #referans grup için örneklem büyüklüğü
ns=ns1+ns2 #örneklem büyüklüğü
b.min=-3 #güçlük parametresinin en düşük değeri
b.max=3 #güçlük parametresinin en yüksek değeri
theta.min=-3 #yetenek parametresinin en düşük değeri
theta.max=3 #yetenek parametresinin en yüksek değeri
theta1=seq(theta.min,theta.max,((theta.max-theta.min)/(ns1-1)))
theta2=seq(theta.min,theta.max,((theta.max-theta.min)/(ns2-1)))
beta1=seq(b.min,b.max,((b.max-b.min)/(ni-1)))
t1=c(-0,484) # 40 maddeye ilişkin güçlük parametre değerleri
t2=c(0,233)
t3=c(-0,989)
t4=c(0,402)
t5=c(-0,323)
t6=c(-0,430)
t7=c(-0,473)
t8=c(-0,504)
t9=c(-0,812)
t10=c(-1,087)
t11=c(0,640)
```



```

t12=c(1,785)
t13=c(0,760)
t14=c(-1,333)
t15=c(0,128)
t16=c(0,431)
t17=c(-1,251)
t18=c(0,863)
t19=c(-0,902)
t20=c(-0,421)
t21=c(-1,282)
t22=c(-0,170)
t23=c(-1,428)
t24=c(0,853)
t25=c(-1,036)
t26=c(1,092)
t27=c(0,364)
t28=c(-1,820)
t29=c(-1,161)
t30=c(1,172)
t31=c(-0,398)
t32=c(0,292)
t33=c(0,570)
t34=c(1,441)
t35=c(-0,598)
t36=c(-0,111)
t37=c(0,362)
t38=c(0,191)
t39=c(0,281)
t40=c(1,172)
th=rbind(t1,t2,t3,t4,t5,t6,t7,t8,t9,t10,t11,t12,t13,t14,t15,t16,t17,t18,t19,t20,t21,t22,t23,t24,t
25,t26,t27,t28,t29,t30,t31,t32,t33,t34,t35,t36,t37,t38,t39,t40)
beta2=beta1
beta2[c(20,40)]=beta2[c(20,40)]+0.50 #20 ve 40 DMF'li olarak belirlenen madde
numaraları
beta1=beta1+th

```

```

beta2=beta2+th
cur=0
repeat{
  datagen<-function(ns,ni,t,beta,theta){
    resp=matrix(0,ns,ni)
    prob=array(0,c(ns,(t+1),ni))
    for (i in 1:ni){
      ncat=rep(t+1,ni)
      beta_vec=unlist(beta[i,])
      beta_vec=c(0,beta_vec)
      num=matrix(0,ns,ncat)
      tsum=0
      den=rep(0,ns)
      cc=0
      for(k in 1:max(ncat)){
        tsum=tsum+beta_vec[k]
        num[,k]=exp(cc*theta-tsum)
        den=den+num[,k]
        cc=cc+1
      }
      for (k in 1:max(ncat)){
        prob[,k,i]=num[,k]/den
      }
    }
    random=matrix(runif(ns*ni),ns,ni)
    for (i in 1:ni){
      sump=numeric(ns)
      for (k in 1:t){
        sump=sump+prob[,k,i]
        resp[,i]=ifelse(random[,i]>sump, resp[,i]+1, resp[,i]+0)
      }
    }
    return(resp)
  }
}

```


EK 12. Özgeçmiş

Kişisel Bilgiler	
Adı Soyadı	Fatma Gökçen AYVA YÖRÜ
Doğum Yeri	Ankara/Çankaya
Eğitim Durumu	
Doktora	Gazi Üniversitesi
	Eğitim Bilimleri Enstitüsü
	Eğitimde Ölçme ve Değerlendirme Bilim Dalı
Yüksek Lisans	Abant İzzet Baysal Üniversitesi
	Sosyal Bilimleri Enstitüsü
	Eğitimde Ölçme ve Değerlendirme Bilim Dalı
Üniversite	Ondokuz Mayıs Üniversitesi
	Fen-Edebiyat Fakültesi
	İstatistik Bölümü
İş Deneyimi	
Araştırma Görevlisi 2013/ -	Gazi Üniversitesi
	Gazi Eğitim Fakültesi
	Eğitimde Ölçme ve Değerlendirme Bilim Dalı
Araştırma Görevlisi 2011/2013	Afyon Kocatepe Üniversitesi
	Eğitim Fakültesi
	Eğitim Bilimleri Bölümü
İletişim Bilgileri	
e-mail	korelasyon_@hotmail.com
	korelasyon03@gmail.com

EK 13. Turnitin Orjinallik Raporu

ORIJINALLIK RAPORU			
% 5	% 3	% 1	% 6
BENZERLIK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ
BİRİNCİL KAYNAKLAR			
1	pdfs.semanticscholar.org İnternet Kaynağı	% 2	
2	Submitted to Akdeniz University Öğrenci Ödevi	% 1	
3	Submitted to Ankara University Öğrenci Ödevi	% 1	
4	slidex.tips İnternet Kaynağı	% 1	
İnceleme Tarihi: 05.12.2019			



GAZİLİ OLMAK AYRICALIKTIR..