

**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
ÖLÇME VE DEĞERLENDİRME PROGRAMI**

**ÇOKLU GRUP DEĞİŞEN MADDE FONKSİYONU
BELİRLEME TEKNİKLERİNİN FARKLI
SİMÜLASYON KOŞULLARINDA
KARŞILAŞTIRILMASI**

DOKTORA TEZİ

ÖZKAN SAATÇIOĞLU

**ANKARA
MAYIS, 2022**



**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
ÖLÇME VE DEĞERLENDİRME PROGRAMI**

**ÇOKLU GRUP DEĞİŞEN MADDE FONKSİYONU
BELİRLEME TEKNİKLERİNİN FARKLI
SİMÜLASYON KOŞULLARINDA
KARŞILAŞTIRILMASI**

DOKTORA TEZİ

ÖZKAN SAATÇIOĞLU

DANIŞMAN: DOÇ. DR. ERGÜL DEMİR

**ANKARA
MAYIS, 2022**

Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Özkan SAATÇIOĞLU adlı öğrencinin hazırladığı “Çoklu Grup Değişen Madde Fonksiyonu Belirleme Tekniklerinin Farklı Simülasyon Koşullarında Karşılaştırılması” başlıklı bu çalışma Ölçme ve Değerlendirme Anabilim Dalı / Ölçme ve Değerlendirme Programı’nda jüri üyelerince oy birliği ile **Doktora Tezi** olarak kabul edilmiştir.

	<u>Jüri Üyeleri</u>	<u>İmza</u>
Başkan	Prof. Dr. Nizamettin KOÇ	
Üye	Doç. Dr. Ergül DEMİR (Danışman)	
Üye	Prof. Dr. Duygu ANIL	
Üye	Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU	
Üye	Doç. Dr. H. Deniz GÜLLEROĞLU	

ONAY

Bu tez Ankara Üniversitesi Lisansüstü Eğitim Öğretim Yönetmeliği’nin ilgili maddeleri uyarınca jüri üyeleri tarafından 23/05/2022 tarihinde, Enstitü Yönetim Kurulunca .../.../20... tarihinde kabul edilmiştir.

.....

Prof. Dr. İlhan YALÇIN

Eğitim Bilimleri Enstitüsü Müdürü

ETİK İLKELERE UYGUNLUK BİLDİRİMİ

Tez içindeki bütün bilgileri akademik yazım kurallarına uygun biçimde raporlaştırdığımı ve bunları etik ilkelere (atıfta bulunulan tüm yapıtlara kaynaklarda yer verilmesi, tezde kullanılan bilgi ve belgelere resmi yollarla ulaşılması ve bunların aslı bozulmadan kullanılması vb.) uygun olarak elde ettiğimi ve sunduğumu bildiririm.

Özkan Saatçioğlu

ÖZET

ÇOKLU GRUP DEĞİŞEN MADDE FONKSİYONU BELİRLEME TEKNİKLERİNİN FARKLI SİMÜLASYON KOŞULLARINDA KARŞILAŞTIRILMASI

SAATÇİOĞLU, Özkan

Doktora Tezi, Ölçme ve Değerlendirme Anabilim Dalı

Tez Danışmanı: Doç. Dr. Ergül DEMİR

Mayıs, 2022, xvi + 367 Sayfa

Bu çalışmada ikili puanlanan maddelerde çoklu grup Değişen Madde Fonksiyonu (DMF) tekniklerinin çeşitli koşullardaki performanslarının incelenmesi amaçlanmıştır. 2PLM'ye dayalı 20 maddelik veri setinin üretildiği bu simülasyon çalışmasında yetenek parametresi için normal dağılım, güçlük $[b \sim U(-1.06, 1.39)]$ ile ayırt edicilik $[a \sim U(0.84, 1.53)]$ parametreleri için ise tekdüze dağılım dikkate alınmış ve dört madde DMF'li olarak belirlenmiştir. Bununla beraber grup sayısı (3, 4), referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3), DMF miktarı $[a(0.0, 0.4, 0.7), b(0.0, 0.4, 0.7)]$ ve DMF türü [DMF Yok, TB, TBO(a), TBO(ab)] faktörlerine ilişkin koşullar değiştirilerek, çoklu grup tekniklerinin (GLK, GLR-LRT, GLR-Wald, GMH ve Lordif) performansları Tip I hata, istatistiksel güç ve Tip III hata oranları açısından karşılaştırılmıştır. Verilerin üretilmesi ve analiz edilmesi sürecinde R yazılımındaki çok sayıda paketten yararlanılmıştır. Simülasyonun geçerliğini belirlemek için mutlak yanlılık ve RMSE hesaplanmış ve madde güçlük ile ayırt edicilik parametrelerine ilişkin kestirimlerin doğrulandığı görülmüştür. Araştırmada dikkate alınan faktörlerin ana ve etkileşim etkilerini incelemek için Faktöriyel ANOVA kullanılmış ve eta kare ile kısmi eta kare etki büyüklüğü katsayıları yorumlanmıştır. Araştırma sonucunda çoklu grup teknikleri için hesaplanan temel Tip I hata oranlarının referans grubun örneklem büyüklüğü ve grup sayısından etkilenmeyerek, liberal kritere göre kabul edilebilir olduğu gözlemlenmiştir. Faktörler arasındaki etkileşim etkileri incelendiğinde; DMF türü, DMF miktarı ve referans grubun örneklem büyüklüğünün çoklu grup teknikleri için hesaplanan Tip I hata, istatistiksel güç ve Tip III hata oranları üzerinde önemli birer faktör oldukları tespit edilmiştir. Grup sayısının ise bu tekniklerin

performansları üzerinde ciddi bir farklılaşmaya yol açmadığı belirlenmiştir. Tip I hata, istatistiksel güç ve Tip III hata açısından sonuçlar birlikte değerlendirildiğinde, farklı koşullarda farklı tekniklerin daha iyi performans gösterdikleri ortaya konulmuştur. Bu sonuçlardan hareketle gerçek veri setlerinde çoklu grup tekniklerinden en az ikisinin birlikte kullanılması önerilebilir.

Anahtar Sözcükler: Değişen madde fonksiyonu, çoklu grup, ikili puanlanan madde, kültürlerarası değerlendirme.



ABSTRACT

COMPARING THE TECHNIQUES FOR DETERMINING MULTIPLE-GROUP DIFFERENTIAL ITEM FUNCTIONING UNDER DIFFERENT SIMULATION CONDITIONS

SAATÇIOĞLU, Özkan

Dissertation, Department of Measurement and Evaluation

Supervisor: Assoc. Prof. Dr. Ergül DEMİR

May, 2022, xvi + 367 Pages

The aim of this study was to examine the performance of multiple-group Differential Item Functioning (DIF) in dichotomous items under different conditions. In this simulation where a dataset of 20 items was generated using 2PLM, normal distribution was considered for the ability parameter and uniform distribution for the difficulty [$b \sim U(-1.06, 1.39)$] and discrimination [$a \sim U(0.84, 1.53)$] parameters, and thus four items were found to have DIF. Multiple-group techniques (GLK, GLR- LRT, GLR-Wald, GMH, and Lordif) were compared in terms of Type I error, statistical power, and Type III error rates by varying the conditions regarding the number of groups (3, 4), the sample size of the reference group (600, 1800), the ratio of the sample sizes of the focus groups to the reference group (1:1/1, 1:1/2 and 1:2/3), the amount of DIF [$a(0.0, 0.4, 0.7)$, $b(0.0, 0.4, 0.7)$] and the type of DIF [No DIF, uniform, non-uniform(a), non-uniform(ab)]. A number of R packages were employed in the process of generating and analyzing the data. Absolute bias and RMSE were calculated to determine the validity of the simulation, confirming the estimations regarding the item difficulty and discrimination parameters. Factorial ANOVA was used to examine the main and interaction effects of the factors addressed in the study, and eta-squared and partial eta-squared effect size coefficients were interpreted. The results showed that the basic Type I error rates calculated for the multiple-group techniques were not affected by the sample size of the reference group and the number of groups, and were acceptable based on the liberal criterion. An analysis of the interaction effects between the factors revealed that the type of DIF, the amount of DIF and the sample size of the reference group were important factors on the rates of Type I error, statistical power and Type III error

calculated for the multiple-group techniques. However, the number of groups did not lead to a considerable difference in the performance of these techniques. Evaluating the results for Type I error, statistical power and Type III error together, it was found that different techniques performed better in different conditions. Based on these results, it can be recommended to use at least two of the multiple-group techniques together in real datasets.

Key Words: Differential item functioning, multiple groups, dichotomous items, cross-cultural assessment.



ÖNSÖZ

Hayatımın en zor süresince yardımını ve hoşgörüsünü hiçbir zaman esirgemeyen, verdiği tavsiyelerle yolumu aydınlatan, kendime olan inancımı kaybetmememi sağlayan, her konuda bilgi ve tecrübelerini aktaran, en ümitsiz olduğum anlarda ne olursa olsun pes etmemeyi öğreten, çalışmamın en başından bir ürün olarak ortaya çıkmasına kadar her aşamasında destekleriyle araştırmaya katkı sunan, akademik ve kişisel hayatımda ilham aldığım, kıymetli danışmanım Doç. Dr. Ergül DEMİR'e akademik gelişimime yapmış olduğu katkıları için teşekkürü bir borç bilirim.

Değerli görüş ve önerileriyle tez çalışmama yapmış oldukları katkılarından dolayı saygıdeğer hocalarım Prof. Dr. Nizamettin KOÇ, Prof. Dr. Duygu ANIL, Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU ve Doç. Dr. Deniz GÜLLEROĞLU'na teşekkür ederim.

Akademik hayatımda ders almaktan onur duyduğum Ankara Üniversitesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Anabilim Dalının değerli hocalarına gelişimime yapmış oldukları katkıları için ne kadar teşekkür etsem azdır.

Tez sürecinin başından beri destekleriyle hep yanımda olduklarını bildiğim Muş Alparslan Üniversitesi Eğitim Bilimleri bölüm başkanı Doç. Dr. Canan DEMİR YILDIZ ile Eğitim Fakültesi'nin değerli hocaları Doç. Dr. Zekeriya ÇAM, Dr. Fuat ELKONCA, Dr. Görkem CEYHAN, Dr. Yusuf ÖZTÜRK ve Arş. Gör. Sedat GELİBOLU'na teşekkürlerimi sunarım.

Tezimin şekilsel düzenleme sürecinde gece gündüz demeden yardımını esirgemeyen değerli arkadaşım Arş. Gör. Ömer KAMIŞ'ın özel bir teşekkürü hak ettiğini düşünüyorum.

Ankara Üniversitesi'nde çalıştığım sürece hayatıma renk katan, onlarla olmaktan keyif aldığım başta Arş. Gör. Muharrem ŞENGÜL ve Arş. Gör. Fatih İMROL olmak üzere ismini sayamadığım tüm arkadaşlarıma teşekkür ederim.

Bu araştırmada yer alan tüm/kısmi nümerik hesaplamalar TÜBİTAK ULAKBİM, Yüksek Başarım ve Grid Hesaplama Merkezi'nde (TRUBA kaynaklarında) gerçekleştirildiğinden, kurmuş oldukları alt yapıyla analizlerimin gerçekleştirilmesine olanak sağlayan Türkiye Bilimsel ve Teknolojik Araştırma Kurumu'na teşekkürlerimi sunarım.



Umut Selim'e...

İÇİNDEKİLER

	Sayfa
ETİK İLKELERE UYGUNLUK BİLDİRİMİ.....	iii
ÖZET	iv
ABSTRACT	vi
ÖNSÖZ.....	viii
İÇİNDEKİLER.....	x
TABLolar DİZİNİ.....	xiii
ŞEKİLLER DİZİNİ.....	xv
KISALTMALAR/SİMGELER	xvi
BÖLÜM 1	1
GİRİŞ.....	1
Problem.....	1
Amaç.....	18
Önem.....	19
Sınırlılıklar	21
BÖLÜM 2.....	23
KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR.....	23
Lord Ki-Kare ve Genelleştirilmiş Lord Ki-Kare	40
Mantel Haenszel ve Genelleştirilmiş Mantel Haenszel	44
Lojistik Regresyon ve Genelleştirilmiş Lojistik Regresyon	49
Madde Tepki Kuramı'na Dayalı Lojistik Regresyon (Lordif).....	56
BÖLÜM 3	81
YÖNTEM	81
Araştırmanın Modeli.....	81
Simülasyon Deseni	82
Değişimlenen Koşullar.....	83
DMF Teknikleri.....	83
Grup Sayısı.....	84
Referans Grubun Örneklem Büyüklüğü ve Odak Grupların Örneklem Büyüklüklerinin Referans Gruba Oranı.	84
DMF Miktarı.	87
DMF Türü (Kaynağı).	88
Sabit Tutulan Koşullar	92
Veri Üretmede Kullanılan Model.....	93
Test Uzunluğu (Madde Sayısı).....	94
DMF'li Madde Yüzdesi (DMF İçeren Madde Sayısı).	95
Yetenek ve Madde Parametrelerinin Dağılımı.	97
Verilerin Üretilmesi	99
Verilerin Analizi	102

Değerlendirme Ölçütleri.....	104
Tip I Hata, İstatistiksel Güç ve Tip III Hata.....	106
Yanlılık ve RMSE.	111
Ana Etkiler ile Etkileşimler için Faktöriyel ANOVA.	114
Simülasyonun Geçerliği.....	117
Görünüş Geçerliği	119
İstatistiksel Geçerlik.....	121
BÖLÜM 4.....	131
BULGULAR VE YORUMLAR	131
BÖLÜM 5.....	253
SONUÇLAR VE ÖNERİLER	253
Sonuçlar	253
İstatistiksel Geçerliğe İlişkin Sonuçlar.....	253
Tip I Hata Oranlarına İlişkin Sonuçlar.....	254
İstatistiksel Güç Oranlarına İlişkin Sonuçlar	255
Tip III Hata Oranlarına İlişkin Sonuçlar	257
Çoklu Grup Tekniklerinin Performanslarına İlişkin Genel Sonuçlar	258
Öneriler	260
Araştırmacılara Öneriler.....	260
Uygulayıcılara Öneriler.....	262
KAYNAKLAR.....	265
EKLER	299
EK 1. TB DMF Türü İçin Örnek MKE'ler.....	300
EK 2. TBO(a) DMF Türü İçin Örnek MKE'ler	301
EK 3. TBO(ab) DMF Türü İçin Örnek MKE'ler	302
EK 4. Referans ve Odak Grupların Örneklem Büyüklükleri.....	304
EK 5. Odak Gruplara Eklenen DMF Miktarları	305
EK 6. Üç Grup İçin Oluşturulan Simülasyon Deseni	307
EK 7. Referans ve Odak Grupların Madde Parametreleri	309
EK 8. Referans ve Odak Gruplar için Madde Karakteristik Eğrileri.....	310
EK 9. DMF'li Dört Maddenin Referans ve Odak Gruplardaki MKE'leri.....	311
EK 10. Madde Güçlük Parametresi İçin Ortalama Mutlak Yanlılık.....	312
EK 11. Madde Güçlük Parametresine İlişkin Ortalama Mutlak Yanlılık ve Ortalama RMSE Sonuçları	313
EK 12. Madde Güçlük Parametresi İçin Ortalama RMSE	315
EK 13. Madde Ayırt Edicilik Parametresi İçin Ortalama Mutlak Yanlılık	316
EK 14. Madde Ayırt Edicilik Parametresine İlişkin Ortalama Mutlak Yanlılık ve Ortalama RMSE Sonuçları	317
EK 15. Madde Ayırt Edicilik Parametresi İçin Ortalama RMSE	319
EK 16. Üç ve Dört Grupta TB DMF Türü İçin Tip I Hata Oranları.....	320
EK 17. Üç ve Dört Grupta TB DMF Türü İçin İstatistiksel Güç Oranları	322
EK 18. Üç ve Dört Grupta TB DMF Türü İçin Tip III Hata Oranları	324
EK 19. Üç ve Dört Grupta TBO(a) DMF Türü İçin Tip I Hata Oranları	326
EK 20. Üç ve Dört Grupta TBO(a) DMF Türü İçin İstatistiksel Güç Oranları.....	328

EK 21. Üç ve Dört Grupta TBO(<i>a</i>) DMF Türü İçin Tip III Hata Oranları	330
EK 22. Üç ve Dört Grupta TBO(<i>ab</i>) DMF Türü İçin Tip I Hata Oranları	332
EK 23. Üç ve Dört Grupta TBO(<i>ab</i>) DMF Türü İçin İstatistiksel Güç Oranları.....	338
EK 24. Üç ve Dört Grupta TBO(<i>ab</i>) DMF Türü İçin Tip III Hata Oranları	344
EK 25. Üç ve Dört Grup Birlikte Dikkate Alındığında Hesaplanan Tip I Hata Oranları	350
EK 26. Üç ve Dört Grup Birlikte Dikkate Alındığında Hesaplanan İstatistiksel Güç Oranları	355
EK 27. Üç ve Dört Grup Birlikte Dikkate Alındığında Hesaplanan Tip III Hata Oranları	359
EK 28. Etik Kurul Belgesi	363
BENZERLİK BİLDİRİMİ	365
ÖZGEÇMİŞ.....	367



TABLOLAR DİZİNİ

Tablo	Sayfa
Tablo 1. Simülasyon Deseni.....	90
Tablo 2. Madde Düzeyinde Hesaplanan Ortalama Mutlak Yanlılık ve Ortalama RMSE Değerleri.....	122
Tablo 3. Güçlük Parametresi İçin Faktöriyel ANOVA Sonucu (Ortalama Mutlak Yanlılık)	124
Tablo 4. Güçlük Parametresi İçin Faktöriyel ANOVA Sonucu (Ortalama RMSE) ..	125
Tablo 5. Ayırt Edicilik Parametresi İçin Faktöriyel ANOVA (Ortalama Mutlak Yanlılık)	126
Tablo 6. Ayırt Edicilik Parametresi İçin Faktöriyel ANOVA (Ortalama RMSE)	127
Tablo 7. Üç Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)	132
Tablo 8. Dört Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)	133
Tablo 9. Üç Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları)	138
Tablo 10. Dört Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları)	140
Tablo 11. Üç Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)	144
Tablo 12. Dört Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)	145
Tablo 13. Üç Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)	150
Tablo 14. Dört Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)	151
Tablo 15. Üç Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları).....	155
Tablo 16. Dört Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları).....	156
Tablo 17. Üç Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)	161
Tablo 18. Dört Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)	162
Tablo 19. Üç Grupta TBO(ab) DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)	165
Tablo 20. Dört Grupta TBO(ab) DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)	167
Tablo 21. Üç Grupta TBO(ab) Türü DMF için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları).....	172
Tablo 22. Dört Grupta TBO(ab) DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları).....	174
Tablo 23. Üç Grupta TBO(ab) DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)	178

Tablo 24. Dört Grupta TBO(<i>ab</i>) DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)	179
Tablo 25. Tip I Hata Oranları İçin Faktöriyel ANOVA Sonucu.....	185
Tablo 26. Çoklu Grup Teknikleri İçin Hesaplanan Tip I Hata Oranlarına İlişkin Ortalama ve Standart Hatalar	197
Tablo 27. İstatistiksel Güç Oranları için Faktöriyel ANOVA Sonucu.....	209
Tablo 28. Çoklu Grup Teknikleri İçin Hesaplanan İstatistiksel Güç Oranlarına İlişkin Ortalama ve Standart Hatalar	219
Tablo 29. Tip III Hata Oranları için Faktöriyel ANOVA Sonucu	228
Tablo 30. Çoklu Grup Teknikleri İçin Hesaplanan Tip III Hata Oranlarına İlişkin Ortalama ve Standart Hatalar	237
Tablo 31. Çoklu Grup Teknikleri İçin Hesaplanan Tip I Hata, İstatistiksel Güç ve Tip III Hata Oranlarına İlişkin Ortalamalar.....	244



ŞEKİLLER DİZİNİ

Şekil	Sayfa
Şekil 1. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü, DMF Türü, DMF Miktarı ve Grup Sayısı Etkileşimi İçin Tip I Hata Oranları	206
Şekil 2. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü, DMF Türü, DMF Miktarı ve Grup Sayısı Etkileşimi İçin İstatistiksel Güç Oranları.....	224
Şekil 3. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü, DMF Türü, DMF Miktarı ve Grup Sayısı Etkileşimi İçin Tip III Hata Oranları.....	241



KISALTMALAR/SİMGELER

KTK	Klasik Test Kuramı
MTK	Madde Tepki Kuramı
MKE	Madde Karakteristik Eğrisi
a	Madde Ayırt Edicilik Parametresi
b	Madde Güçlük Parametresi
c	Şans Parametresi
θ	Yetenek Düzeyi
$P_i(\theta)$	Maddeye Doğru Yanıt Verme Olasılığı
TIMSS	Trends in International Mathematics and Science Study
PISA	Programme for International Student Assessment
DMF	Değişen Madde Fonksiyonu
1PLM	Bir Parametrelili Lojistik Model
2PLM	İki Parametrelili Lojistik Model
3PLM	Üç Parametrelili Lojistik Model
GRM	Graded Response Model
GPCM	Generalized Partial Credit Model
RMSE	Root Mean Square Error
GLK	Genelleştirilmiş Lord Ki-kare
GLR-LRT	Genelleştirilmiş Lojistik Regresyon-Olabilirlik Oran Testi
GLR-Wald	Genelleştirilmiş Lojistik Regresyon-Wald Testi
GMH	Genelleştirilmiş Mantel-Haenszel
Lordif	Logistic Ordinal Regression Differential Item Functioning Using IRT
TB DMF	Yalnızca güçlük parametresinden kaynaklı DMF türü
TBO(a) DMF	Yalnızca ayırt edicilik parametresinden kaynaklı DMF türü
TBO(ab) DMF	Hem güçlük hem de ayırt edicilik parametresinden kaynaklı DMF türü

BÖLÜM 1

GİRİŞ

Bu bölümde araştırma problemine giriş yapılarak araştırmanın amacı, önemi ve sınırlılıkları ele alınmıştır. Çoklu grupların dikkate alındığı bu çalışmada, alanyazında daha önce kullanıldığı belirlenen farklı yaklaşımların sınırlılıkları açıklandıktan sonra grup sayısının ikiden fazla olması durumunda neden çoklu grup tekniklerinin kullanılması gerektiği detaylı bir şekilde anlatılmıştır.

Problem

Eğitim ve psikoloji alanında bireylere farklı amaçlar doğrultusunda ölçme araçları uygulanmakta ve bu araçlardan elde edilen puanlara dayalı olarak bireyler hakkında birtakım kararlar alınmaktadır. Bu kararların isabetli olması elde edilen puanların geçerli olmasıyla, diğer bir ifadeyle ölçülmek istenen özellikten başka özelliklerin ölçme işlemine karışmamasıyla, mümkündür. Puanların farklı gruplarda yer alan bireyler için eşitliğinin sağlanması, doğru ve adil bir karşılaştırma yapılabilmesi için önemli görüldüğünden, son yarım yüzyıldır eğitsel ve psikolojik testlerde adalet kavramı sıklıkla gündeme gelmektedir (Penfield ve Camilli, 2007).

Eğitim ve Psikolojik Test Standartlarında (Standards for Educational and Psychological Testing) da vurgulandığı üzere testlerin fırsat eşitsizliğine yol açmaması gerekmektedir (AERA, APA ve NCME, 1999). Ölçülen özellik bakımından aynı yetenek düzeyindeki bireylerin bazı demografik özelliklerine bağlı olarak ölçme aracındaki maddeleri doğru yanıtlama olasılıklarının farklılaşması, yani madde parametrelerinin çeşitli gruplar arasında değişiklik göstermesi, ilgili maddenin yanlı olabileceğine işaret etmektedir. Bu durumda ise, ölçülen özellik haricinde başka bir özelliğin ölçme işlemine karıştığı kabul edildiğinden hem geçerliğin tehdit edildiği hem de fırsat eşitsizliğinin olduğu düşünülmektedir (Clauser ve Mazor, 1998; Martinková vd., 2017; Osterlind ve Everson, 2009; Penfield ve Camilli, 2007).

Psikometri alanında 1960'lı yıllardan sonra yapılan çalışmalara dayalı olarak yanlılığın geçerlik üzerinde önemli bir sorun teşkil ettiği sonucuna ulaşılmakta ve bu

konuda kapsamlı çalışmaların yapıldığı belirtilmektedir (Camilli, 2006; Holland ve Wainer, 1993, 1994; Oshima, Wright ve White, 2015; Osterlind ve Everson, 2009). Madde düzeyindeki yanlılık, ölçme aracındaki bir maddenin belirli bir alt grup lehine veya aleyhine olması anlamına gelmektedir (Osterlind ve Everson, 2009). Madde yanlılığının puanların geçerliğini tehdit eden ve sistematik hataya neden olan önemli bir kavram olması nedeniyle (Clauser ve Mazor, 1998; Osterlind, 1983) eğitim ve psikometri alanındaki araştırmalarda yanlılık incelemelerinin yapılması önemli görülmektedir (Magis, Raîche, Béland ve Gérard, 2011).

Aynı yetenek düzeyindeki bireylerin madde düzeyindeki performanslarının cinsiyet veya kültür gibi özellikler açısından farklılaşması, bu maddelerin gruba bağımlı olduğunu gösterdiği için, bireylerin yanlı bir maddeye doğru yanıt verme olasılığı yetenek düzeyinden ziyade bireyin ait olduğu grubun özelliğine bağlı olmaktadır. Oysa farklı demografik özellik gösteren gruplarda yer alsalar dahi, yetenek düzeyi aynı olan bireylerin, ölçme araçlarındaki maddeleri doğru yanıtlama olasılıklarının eşit olması beklenmektedir (Osterlind, 1983; Zieky, 2006; Zumbo, 1999). Nitekim aynı yetenek düzeyinde olan iki veya daha fazla demografik grup için maddelerin yanlı olmadığına tespit edilmesi, grupların karşılaştırılabilir olduğunu ortaya koymaktadır (Ackerman, 1992; Salehi ve Tayebi, 2012; Walker, 2011; Walker ve Beretvas, 2001).

Araştırmalarda ölçme araçlarındaki yanlı maddelerin geçerlik sorunu oluşturması nedeniyle, ilgili maddelerin ölçekten çıkarılması veya gözden geçirilerek düzenlenmesi önerilmektedir (Cohen, Kim ve Wollack, 1996; Hope, Adamson, McManus, Chis ve Elder, 2018). Bir maddenin yanlı olması ölçme aracının da yanlı olabileceği (test yanlılığı) anlamına gelmekte, bu durum ise uygulama yapılan bazı gruplara avantaj veya dezavantaj sağlayabilmektedir (Penfield ve Camilli, 2007). Diğer taraftan bazen madde düzeyinde yanlılık bulunmaktayken ölçme aracının yanlı olmadığı durumlarla da karşılaşılabilir. Örneğin A ve B gibi iki grup dikkate alındığında bir maddenin A grubu, diğer bir maddenin ise B grubu lehine farklılaşması, yanlılığın etkisinin test düzeyinde ortadan kalkması ile sonuçlanabilmektedir. Böylesi bir durumda test düzeyinde yanlılık gözlenmese de madde düzeyindeki yanlılığın detaylı bir şekilde incelenmesi gerekmektedir (Li ve Zumbo, 2009). Öte yandan eğer bir madde testi alan herkes için tehdit oluşturuyorsa, bu durumun yanlılık olarak değerlendirilmediği belirtilmektedir (Zumbo, 1999). Yapılan bu açıklamalar; ölçme araçlarının geliştirildiği veya kullanıldığı çalışmalarda, maddelerin herhangi bir grup lehine yanlı olup olmadığının belirlenmesi gerektiğini göstermektedir.

Ölçme araçlarındaki bir maddenin bir alt gruba diğer grup veya gruplara kıyasla haksız bir fayda sağlaması, yani gruplardan birisi için avantajlı olması, madde parametrelerinin gruplar arasında farklılaştığını göstermektedir (Penfield ve Camilli, 2007). Lord (1977) çalışmasında madde parametrelerinin gruplar için değişiklik göstermesi sonucunda gruplara ilişkin Madde Karakteristik Eğrilerinin (MKE) farklılaşacağını ifade etmiş, bu durumun ise madde yanlılığının bir göstergesi olabileceğini açıklamıştır. Eğitim, davranış bilimleri ve psikoloji alanlarındaki çalışmalarda önemli bir konu alanı olan yanlılık (Camilli, 2006; Kulick ve Hu, 1989) istatistiklere dayalı olarak yapılan bir yorum olmasının yanı sıra, madde parametrelerinin gruplar arasında sistematik bir şekilde düşük veya yüksek kestirilmesi sebebiyle adaletsizlik kavramı ile de ilişkilendirilmektedir. Yanlı bir madde, ölçülmesi amaçlanan bilgi veya beceri haricinde birtakım özelliklere sahip olan grup/gruplar için avantaj, bu özelliklere sahip olmayan grup/gruplar için ise dezavantaj oluşturabilmektedir. Ölçme araçlarında yanlı maddelerin bulunması ise verilen kararların hatalı olmasına yol açabilmektedir. Bu nedenle madde ve test yanlılığının genellikle olumsuz bir anlamının olduğu vurgulanmaktadır (Ackerman, 1992; de Ayala, 2009; Ellis ve Raju, 2003).

Yanlılık üzerine yapılan ilk araştırmalarda, maddelerin azınlık gruptaki bireylere zor gelebileceği düşüncesiyle, farklı gruplardaki bireylerin performanslarına ilişkin ampirik kanıtların toplanması amaçlanmıştır. Bu anlamda aynı yetenek düzeyinde olan farklı gruplardaki bireylerin tek bir baskın özelliği ölçen (tek boyutlu) testlerdeki maddeleri doğru yanıtlama olasılıklarının farklılaşması, maddelerin Değişen Madde Fonksiyonu (DMF-Differential Item Functioning) gösterdiği şekilde yorumlanmış ve böyle bir durumda maddeler üzerinde birtakım incelemelerin yapılması gerektiği vurgulanmıştır (Hambleton, Swaminathan ve Rogers, 1991; Wen, 2014; Wood, 2011).

Madde yanlılığının tespit edilmesi amacıyla yapılan incelemeler, genel olarak istatistiksel bir süreç ile başlamakta ve bu süreç DMF belirleme çalışmaları olarak isimlendirilmektedir (Holland ve Thayer, 1988; Holland ve Wainer, 1993). DMF belirleme sürecinde yapılan işlemlerin amacı dikkate alındığında DMF'nin istatistiksel bir terim olduğu ifade edilebilmektedir. Türkçe alanyazındaki birçok çalışmada DMF olarak kullanılmasına rağmen, bu terimin bazı araştırmalarda Maddenin Farklı Fonksiyonlaşması (Acar, 2011; Acar ve Kelecioğlu, 2010), Diferansiyel Madde Fonksiyonu (Yurdugül ve Aşkar, 2004), Madde İşlev Fonksiyonu (Ünsal Özberk ve Koç, 2017), Madde İşlev Farklılığı (Bilican Demir, 2014; Korkmaz, 2005), Madde İşlevinin Farklılaşması (Suna, 2012) ve Farklılaşan Madde Fonksiyonu (Yalçın, 2015) gibi farklı

şekillerde de isimlendirildiği görülmüştür. Bu bağlamda çalışmaların çoğunda olduğu gibi mevcut araştırma kapsamında da Değişen Madde Fonksiyonu olarak kullanılmasına karar verilmiştir.

Alanyazındaki araştırmalarda yapılan ortak açıklamalara göre DMF; testle ölçülen yetenek düzeyi bakımından aynı olmasına rağmen ülke, coğrafi bölge, kültürel özellikler, anadil, etnik köken, cinsiyet, yaş, sosyoekonomik düzey, dini inanç, ekonomik farklılıklar ve testin uygulanış şekli gibi özellikler açısından farklı gruplarda bulunan bireylerin, test maddelerine doğru yanıt verme olasılıklarının istatistiksel olarak farklılaşması şeklinde tanımlanmaktadır (Gamerman, Gonçalves ve Soares, 2018; Goodman, Wilse, Allen ve Klaric, 2011; Osterlind, 1983; Raju, 1990; Zumbo, 1999). Diğer yandan Penfield ve Camilli (2007) tarafından yapılan araştırmada demografik, sosyal ve kültürel olarak farklı gruplarda bulunan aynı yetenek düzeylerindeki bireylerin bir maddeyi doğru yanıtlamalarına ilişkin koşullu olasılık dağılımının, grup üyeliğinden bağımsız olmaması durumunda DMF'nin varlığından söz edilebildiği açıklanmaktadır. Örneğin yetenek kontrol edildikten sonra erkek öğrencilerin bir problem çözme maddesini doğru yanıtlama olasılığı, kadınlara göre daha yüksek olduğunda bu maddenin DMF gösterebileceği belirtilmektedir (Wen, 2014). Benzer şekilde PISA (Programme for International Student Assessment) veya TIMSS (Trends in International Mathematics and Science Study) gibi uluslararası uygulamalarda, aynı yetenek düzeyinde fakat gelişmiş ve gelişmekte olan ülkelerde yer alan öğrencilerin bir maddeyi doğru yanıtlama olasılıklarının farklılaşması DMF'nin varlığına işaret etmektedir (Gamerman vd., 2018). Öğrencilerin aynı yetenek düzeyinde ancak azınlık ve çoğunluk olmak üzere farklı gruplarda yer aldığı başka bir örnek durum dikkate alındığında ise; sınava giren azınlık gruptaki öğrencilerin bazı test maddelerinin içeriğine aşina olmadıkları durumda, bu öğrencilerin çoğunluk gruba kıyasla maddelere yanlış yanıt verme olasılıklarının daha yüksek olabileceği ve azınlık grubundaki öğrencilerin daha düşük puan almalarının söz konusu olabileceği ifade edilmektedir (Walker, 2011).

DMF ile ilgili öncelikle iki grubun karşılaştırılması amacıyla yapılan tanımlar ve geliştirilen teknikler, alanyazındaki ihtiyaçlar ve kuramsal gelişmelere bağlı olarak grup sayısının ikiden fazla olduğu çoklu gruplar için de genelleştirilebilmektedir. Hidalgo ve López-Pina (2004) ile Wyse ve Mapuranga (2009) yaptıkları araştırmalarında bireylerin yetenekleri sabit tutulduktan sonra bir maddede iki veya daha fazla alt grubun farklı performans sergilemesinin, bu maddenin DMF göstermesine neden olacağını belirtmişlerdir. Bu açıklamaya göre yetenek düzeyi kontrol altına alındıktan sonra grup

sayısı ne olursa olsun örnekleme yer alan en az bir alt grubun performansındaki farklılaşmanın DMF'ye neden olacağı anlaşılmaktadır (Goodman vd., 2011; Hope vd., 2018).

Genel olarak DMF belirleme çalışmalarında bireylerin madde düzeyindeki performanslarının karşılaştırılabilmesi amacıyla örneklemin analiz öncesinde referans ve odak olmak üzere gruplara ayrılması gerekmektedir (Cheng, Chen, Qian ve Chang, 2013; Kim, Cohen ve Park, 1995; Zieky, 1993). Odak grup birincil olarak ilgilenilen grup iken, referans grup odak grubun performansının karşılaştırıldığı standart grup olarak tanımlanmaktadır. Örneğin bir sınava katılan Siyah öğrenciler odak grup, Beyaz öğrenciler ise referans grup olarak kabul edilmektedir. Cinsiyet açısından düşünüldüğünde ise erkekler genellikle referans grup, kadınlar ise odak grup olarak dikkate alınmaktadır (de Ayala, 2009; Holland ve Thayer, 1988; Rollins III, 2018; Taylor ve Lee, 2012; Walker, 2011). Bununla birlikte farklı özelliklere sahip olan örneklerde veya PISA ve TIMSS gibi çok sayıda ülkenin katıldığı uluslararası uygulamalarda referans grubunun belirlenmesinin kolay olmaması nedeniyle, referans ve odak grupların genellikle araştırmacının amacı ve hedeflerine bağlı olarak belirlendiği de gözlemlenmektedir (Kim vd., 1995; Magis vd., 2011; Svetina ve Rutkowski, 2014). Bu bilgiler doğrultusunda, DMF analizlerinin çoğunlukla referans ve odak olmak üzere iki farklı grup üzerinde gerçekleştirildiği ve gruplardaki tüm bireylerin özelliklerinin benzer olduğu varsayılmaktadır (de Ayala, Stapleton ve Dayton, 2002).

İki gözlenen grubun var olduğu DMF çalışmalarında genellikle referans grup çoğunluk veya avantajlı gruptur ve tipik olarak odak grubun karşılaştırıldığı temel grup olarak kabul edilmektedir. Buna karşın odak grup ise azınlık veya dezavantajlı gruptur ve tipik olarak adalet endişelerinin odak noktası olarak dikkate alınmaktadır (Camilli, 1992; Cohen vd., 1996; Desjardins ve Bulut, 2018; Guilera, Gómez-Benito, Hidalgo ve Sánchez-Meca, 2013; Mapuranga, Dorans ve Middleton, 2008). Bu endişelerin sebebi ise, aynı yetenek düzeyindeki iki gruptan birinde yer alan bireylerin, diğerindeki bireylere göre bir maddeye doğru yanıt verme olasılıklarının farklılaşma ihtimalidir (Ellis ve Raju, 2003). Örneğin aynı yetenek düzeyinde olmalarına rağmen odak gruptan sınava girenlerin referans gruptan sınava girenlere kıyasla testteki bir maddeye doğru yanıt verme olasılığının düşük olması adaletsizliğe neden olabilmektedir (Desjardins ve Bulut, 2018). Dolayısıyla grupların karşılaştırıldığı çalışmalarda bir maddeye doğru yanıt verme olasılığının daha düşük olabileceği öngörülen grubun odak grup, daha yüksek olabileceği öngörülen grubun ise referans grup olarak ele alındığı belirtilmektedir. Yapılan DMF

incelemeleri sonucunda ise gruplar arasında farklılaşma olup olmadığı belirlenmekte ve herhangi bir grup lehine avantaj/dezavantaj sağlayan olası DMF’li maddeler tespit edilebilmektedir (Cheng vd., 2013; Noortgate ve De Boeck, 2005; Walker, 2011; Wen, 2014).

Karami ve Nodoushan (2011) tarafından da belirtildiği üzere örnekleme her zaman dezavantajlı grup bulunmaması nedeniyle, grupların isimlendirilmesi keyfi olarak da yapılabilmektedir. Yani referans ve odak grupların etiketlenmesi için her çalışmada uyulması zorunlu olan kesin bir kuralın söz konusu olmadığı belirtilmektedir. Hatta referans ve odak grup tanımının araştırmacıya veya araştırmaya göre farklılık gösterdiği durumlar da mümkün olabilmektedir (Tay, Meade ve Cao, 2015). Örneğin genel olarak araştırmalarda kadınların odak grup olarak dikkate alındığı görülmekteyken, Glas ve Jehangir (2013) ile Geramipour ve Shahmirzadi (2019) yaptıkları çalışmalarında kadınların referans, erkeklerin ise odak grup olabileceğini belirtmişlerdir. Farklı bir örnekte ise aynı test farklı zamanlarda uygulandığında, farklı uygulamalar arasında DMF gösteren madde olup olmadığı incelenmiş ve uygulamalardan birinin “referans”, diğer uygulama/uygulamaların ise “odak/odaklar” olarak tanımlanabileceği ifade edilmiştir (Magis vd., 2011).

Farklı gruplar arasında yanlılık olup olmadığını belirlemek amacıyla alanyazında yapılan ilk çalışmalarda; DMF’nin kaynağı olarak cinsiyet (Kadın-Erkek) ve ırk (Siyahlar-Beyazlar) değişkenlerinin ele alındığı gözlemlenmektedir. Bu bağlamda örnekleme yer alan bireyler Siyahi, Hispanik, Asyalı gibi azınlık gruplar olarak veya engelli olma, kadın olma gibi bazı özellikler açısından toplumda dezavantajlı olan gruplar olarak dikkate alınmaktadır (Zieky, 1993; Zumbo, 2007). Alanyazındaki ilk DMF belirleme çalışmalarında, bir referans grubun (tipik olarak çoğunluk veya avantajlı grup) yalnızca bir odak grupla karşılaştırıldığı görülmektedir. Bu araştırmalarda genellikle odak gruptaki birey sayısının (örneklem büyüklüğünün), referans gruptaki birey sayısından daha az olabileceği düşünülmektedir. Örneğin, testin geliştirildiği anadili konuşan öğrencilerin referans grup, bu dili yeni öğrenen öğrencilerin ise odak grup olarak dikkate alınması durumunda, referans gruba ait örneklemin daha büyük olduğu kabul edilmektedir (Fidalgo ve Scalon, 2010; Martinková vd., 2017; Sinharay, Dorans, Grant ve Blew, 2009; Zieky, 1993).

Yapılan açıklamalardan da anlaşıldığı üzere DMF analizlerinde grup faktörü olarak en çok dikkate alınan değişkenlerin cinsiyet, ırk, etnik köken, anadil, yöresel ve kültürel özellikler olduğu ifade edilebilmektedir (Camilli, 2006, 2013; Iwata, Turner ve

Lloyd, 2002; Roussos ve Stout, 1996). DMF ile ilgili 51 çalışmanın incelendiği bir araştırmada, DMF'nin en fazla cinsiyet (%35.3) açısından araştırıldığı sonucuna ulaşılmaktadır (Tay vd., 2015). Bununla birlikte etnik köken, ırk, anadil ve kültürel özellikler gibi değişkenlerin birbirleri ile ilişkili olması nedeniyle, bu değişkenlerin farklı kombinasyonlarının da ele alındığı belirtilebilmektedir (Jin, Chen ve Wang, 2015). Bu anlamda ilk geliştirilen DMF belirleme teknikleri ile azınlık-çoğunluk veya referans-odak gibi iki grubun ele alınabilmesi nedeniyle (Dell-Ross, 2021; Kanjee, 2007; Penfield ve Lam, 2000; Zieky, 1993, Zumbo, 2007) birçok DMF belirleme çalışmasında yalnızca iki grubun karşılaştırılabildiği açıklanmıştır (Oshima vd., 2015). Nitekim Türkiye dışında (Cheema, 2019; Ellis ve Raju, 2003; Finch, 2016; Le, 2009; Lyons-Thomas, Sandilands ve Ercikan, 2014; Mendes-Barnett ve Ercikan, 2006; Taylor ve Lee, 2012; Zampetakis, Bakatsaki, Litos, Kafetsios ve Moustakis, 2017) ve Türkiye'de (Bakan Kalaycıoğlu ve Kelecioğlu, 2011; Çelik ve Özer Özkan, 2020; Çıkrıkçı Demirtaşlı ve Ulutaş, 2015; Doğan ve Öğretmen, 2008; Gök, Kelecioğlu ve Doğan, 2010; Kan, Sünbül ve Ömür, 2013; Karakaya, 2012; Kuzucu ve Sarıot Ertürk, 2020; Uzun ve Gelbal, 2017; Yurdugül ve Aşkar, 2004) yapılmış olan çok sayıda çalışmada referans ve odak olmak üzere yalnızca iki grubun karşılaştırılmasına odaklanıldığı belirlenmiştir.

İki grubun karşılaştırıldığı DMF belirleme çalışmalarının yanı sıra bazı araştırmalarda ya değişkenin doğasından kaynaklı ya da grupların tanımlanmasına bağlı olarak ikiden fazla grup arasında karşılaştırma yapılmasına ihtiyaç duyulduğu belirtilmektedir. Bu bağlamda sosyo-ekonomik düzey, etnik köken, ırk, yaş, sosyal sınıf, konuşulan dil, ülke, bölge veya kültür değişkenlerinin biri veya birkaçına (etnik köken ve sosyo-ekonomik düzey) göre bireylerin ikiden daha fazla sayıda alt gruba ayrılması gerekebilmektedir (Camilli, 2013; Finch, 2016; Jin vd., 2015; Kim vd., 1995; Langer, 2008; Magis vd., 2011; Magis vd., 2010; Morales-Ortiz ve Peña-Cardenas, 2019; Oshima vd., 2015; Penfield, 2001; Schmitt ve Dorans, 1990; Vaughn-Coaxum, Mair ve Weisz, 2016). Örneğin bir çalışmada aynı coğrafi bölgedeki üç şehirde veya bir ülkedeki dört farklı bölgede yaşayan aynı yetenek düzeyindeki öğrenciler için, bir testteki maddelerin DMF gösterip göstermediğini araştırmak amaçlanabilmektedir. Benzer şekilde Svetina ve Rutkowski (2014) tarafından yapılan çalışmada da ifade edildiği gibi; çok sayıda katılımcı ülkenin kendi dillerine çevrilmiş testlerdeki performanslarına dayalı olarak karşılaştırılmalarına imkân sağlayan PISA ve TIMSS gibi uygulamalarda, dil, kültür ve müfredat açısından farklı gruplarda yer alan aynı yetenek düzeyindeki öğrencilerin madde performanslarının araştırılması gerekebilmektedir. Bu kapsamda çok sayıda grup

arasında yapılan karşılaştırmaların geçerliğinin kanıtlanması amacıyla geniş ölçekli veri setleri üzerinde DMF belirleme çalışmalarının yapılması önemli görülmektedir (Glas ve Jehangir, 2013; Huang, Wilson ve Wang, 2016; Sandilands, Oliveri, Zumbo ve Ercikan, 2013; Svetina ve Rutkowski, 2014; Tay, Huang ve Vermunt, 2016; Wyse ve Mapuranga, 2009).

Penfield (2001) tarafından da belirtildiği üzere odak grup sayısının ikiden fazla olduğu durumlarda maddelerin yanlı olup olmadığının belirlenmesi sıklıkla arzulanan bir durumdur. Nitekim Ellis ve Raju'nun (2003) çalışmalarında da ifade edildiği gibi aslında çoklu gruplarda da DMF incelemeleri yapmak mümkün olmakta ve yirminci yüzyılın sonlarına kadar olan süreçte çoklu grupların farklı yaklaşımlara dayalı olarak ele alındığı görülmektedir. Yapılan araştırmalarda ikiden fazla sayıda grubun karşılaştırılması gerektiğinde; ikili karşılaştırma, bileşik grup ve odak grup birleştirme gibi üç yaklaşımdan birine başvurulduğu gözlemlenmektedir (Kanjee, 2007; Kim vd., 1995; Magis, Béland, Tuerlinckx ve De Boeck, 2010; Ong, 2012; Penfield, 2001).

Grup sayısının ikiden fazla olduğu durumlarda belirtilen üç farklı yaklaşımdan araştırmalarda en yaygın tercih edilen ikili karşılaştırma yaklaşımıdır (Dell-Ross, 2021; Svetina ve Rutkowski, 2014). Bu yaklaşımda referans grup ile her bir odak grup arasında ikiyeşerli olarak karşılaştırmalar yapılmaktadır (Ellis ve Kimmel, 1992; Sari, 2013; Schmitt ve Dorans, 1990). Örneğin katılımcılar Siyah, Beyaz ve Hispanik olarak tanımlandığında, referans grup olarak belirlenen grup ile odak grupların her biri doğrudan ve bağımsız olarak ayrı ayrı karşılaştırılmaktadır. Bu yaklaşım kullanılarak yapılan gruplar arası karşılaştırmalarda, referans olarak belirlenen grubun (örn. Beyaz) farklılaşabileceği belirtilmektedir (Magis vd., 2010). İkiiden daha fazla sayıda grubun dikkate alındığı DMF çalışmalarında ikili karşılaştırma yaklaşımına dayalı olarak Türkiye'de (Acar, 2008; Çepni, 2011; Demir ve Köse, 2014; Gök, Atalay Kabasakal ve Kelecioğlu, 2014; Ünsal Özberk ve Koç, 2017) ve Türkiye dışında (Ercikan ve Koh, 2005; Farrington ve Lonigan, 2015; Fleishman, Spector ve Altman, 2002; Huang vd., 2016; Iwata vd., 2002; Kim, 2003; Ong, 2012; Sandilands vd., 2013; Schmitt ve Dorans, 1990; Stark, Chernyshenko ve Drasgow, 2004) yapılmış olan çok sayıda araştırma bulunmaktadır.

Çoklu grupların yer alındığı DMF belirleme çalışmalarında ikili karşılaştırma yaklaşımına başvurulduğunda dikkat edilmesi gereken hususlar Ellis ve Mead (2000) tarafından yapılan araştırmada açıklanmış ve Penfield'in (2001) çalışmasında bu yaklaşımın pratikteki dezavantajlarından bahsedilmiştir. Gerçekleştirilen ikili karşılaştırmaların çok zaman alıcı olması nedeniyle kullanılmasının verimli olmaması, bu

yaklaşımın dezavantajlarından biri olarak gösterilmiştir. Örneğin altı farklı dilin konuşulduğu bir ülkede, aynı yetenek düzeyindeki öğrencilerin konuştukları dile göre performanslarının farklılaşıp farklılaşmadığı ikili karşılaştırma yaklaşımına dayalı olarak incelenmek istenildiğinde, 15 ayrı karşılaştırmanın yapılması gerekmektedir. Dolayısıyla çok sayıda grubun olduğu bazı gerçek uygulamalarda ikili karşılaştırma yaklaşımına dayalı olarak yapılacak DMF çalışmalarının çok karmaşık olacağı belirtilmektedir. Bu yaklaşımın diğer bir dezavantajı ise, ikiden fazla grup için birden fazla hipotezin test edilmesi nedeniyle tekrarlanan anlamlılık testleri sonucunda Tip I hatanın artmasıdır. İkili karşılaştırma yaklaşımına dayalı olarak yapılan her bir analizde Tip I hata yapma olasılığı α_i 'dir. Bu anlamda tekrarlanan hipotez testleri sonucunda Tip I hata yapma olasılığı kabul edilen (nominal) manidarlık düzeyini aşabilmektedir. Bonferroni düzeltmesi yapılarak manidarlık düzeyinin kontrol edilmesi ile bu problemin üstesinden gelinebileceği ifade edilmesine rağmen, grup sayısı arttıkça ikili karşılaştırma sayısı arttığından Bonferroni düzeltmesinin pratikteki kullanımının sınırlandığı belirtilmektedir (Carroll, 2015; Fidalgo ve Scalón, 2010; Kanjee, 2007; Kim vd., 1995; Magis vd., 2010; Magis vd., 2011; Oshima vd., 2015; Penfield, 2001; Rockoff, 2018; Rollins III, 2018). Penfield (2001) tarafından da açıklandığı üzere, Bonferroni düzeltmesinin kullanılması Tip I hata oranının kontrol edilmesini sağlasa da genel olarak elde edilen istatistiksel güç, tek bir hipotezin test edildiği araştırmalardan daha düşük olabilmektedir. Bu nedenle çoklu gruplar için DMF belirleme sürecinde ikili karşılaştırma yaklaşımı dikkate alınırken, bu yaklaşımın sınırlılıklarının farkında olunması gerektiği belirtilmektedir (Ellis ve Kimmel, 1992; Fidalgo, 2011; Kim vd., 1995; Sari, 2013).

Alanyazındaki çalışmalarda açıklandığı üzere ikili karşılaştırma yaklaşımının yanı sıra, analiz edilen grupların her birinin ayrı ayrı odak grup olarak, tüm bireylerden oluşan ortak bir grubun ise referans grup olarak dikkate alındığı bileşik (composite) grup yaklaşımının da bazı araştırmalarda kullanıldığı görülmüştür. DMF belirleme çalışmalarında daha az tercih edilen bileşik grup yaklaşımı ilk olarak Ellis ve Kimmel (1992) tarafından ortaya konulmuştur. Amerikan, Fransız ve Alman öğrencilerin odak grup olarak dikkate alındığı araştırmada, bütün örneklemin dahil olduğu bir bileşik grup referans olarak kabul edilmiştir. Analiz sürecinde DMF belirleme tekniği olarak Lord'un ki-kare tekniği kullanılmış ve odak gruplar ile bileşik grup arasında ikili karşılaştırmalar yapılmıştır (Ellis ve Kimmel, 1992). Böylelikle, bu yaklaşımda her bir odak grup, bileşik grup olarak da isimlendirilen referans grupla karşılaştırılmıştır. Örneğin Siyah, Beyaz ve Hispanik öğrencilerden oluşan bir örnekleme yalnızca bir grubun verileri kullanılarak

kalibre edilen madde parametrelerinin, bütün örneklemin verileri kullanılarak kalibre edilen (operasyonel) madde parametreleri ile karşılaştırıldığı belirtilmiştir (Sari ve Huggins, 2014). Sınırlı sayıda araştırmada tercih edilen bu yaklaşımın uygulamalardaki avantaj/dezavantajlı yönleri, az sayıda çalışmada incelenmiştir (Çelik ve Özer Özkan, 2020; Liu ve Dorans, 2013; Sari ve Huggins, 2014).

İkili karşılaştırma ve bileşik grup yaklaşımlarının karşılaştırıldığı Ellis ve Kimmel'in (1992) çalışmasında, belirtilen bu yaklaşımların avantajlı yönlerinin göreceli olarak değişiklik göstermesi nedeniyle, bu yaklaşımların uygulamalarda farklı sorulara yanıt bulmak için kullanılabileceği iddia edilmiştir. Bileşik grup yaklaşımının dikkate alındığı bir araştırmada Sari ve Huggins (2014), bu yaklaşımın avantajlı olduğu durumları kanıtlamak amacıyla bir çalışma yürütmüştür. Belirtilen araştırmada bileşik grup yaklaşımda da ayrı ayrı bütün gruplar ile bileşik grup arasında ikili karşılaştırmalar yapılmasına rağmen, grup sayısı dört veya daha fazlayken bileşik grup yaklaşımda ikili karşılaştırma yaklaşımına kıyasla daha az sayıda karşılaştırma yapılacak olması nedeniyle, Tip I hata yapma olasılığının düşük olacağı ifade edilmiştir. Bu anlamda bileşik grup yaklaşımda Tip I hata yapma probleminin üstesinden gelinebilir de grup sayısı arttığında ikili karşılaştırma yaklaşımına göre Tip I hata yapma olasılığının azaldığı belirtilmiştir. Diğer taraftan bileşik grup yaklaşımda belirli bir grubun bütün grupları içeren bileşik gruptan farklı olup olmadığı araştırıldığından, karşılaştırılan grupların birbirlerinden bağımsız olmaması bir sınırlılık olarak görülmüştür (Ellis ve Kimmel, 1992; Kim vd., 1995; Sari, 2013; Sari ve Huggins, 2014).

İkili karşılaştırma ve bileşik grup yaklaşımlarının yanı sıra, Kanjee, (2007) tarafından yapılan araştırmada da odak grup birleştirme yaklaşımından bahsedilmiştir. Bu yaklaşımda odak gruplar tek bir grup altında toplanarak, yine referans ve odak olmak üzere iki grup arasında karşılaştırma yapılmaktadır. Başka bir ifadeyle, odak grup birleştirme yaklaşımında tüm odak gruplar tek bir grup altında toplandıktan sonra iki grup için geliştirilen teknikler ile DMF analizleri gerçekleştirilmektedir. Örneğin Siyah, Beyaz ve Hispanik öğrencilerden oluşan bir örnekleme, Beyaz öğrenciler referans, diğer iki gruptaki (Siyah ve Hispanik) öğrencilerin tamamı ise odak grup olarak dikkate alınmaktadır. Bu yaklaşımda, grupların aynı yetenek dağılımlarına sahip ve gruplardaki DMF miktarının aynı düzeyde olduğu varsayımı kabul edilmektedir (Rollins III, 2018).

Alanyazında odak grup birleştirme yaklaşımının tercih edildiği sınırlı sayıda çalışmaya rastlanılmıştır (Bowe, 2017; Glas ve Jehangir, 2013; Zhang, Dorans ve Matthews-López, 2005). Zhang ve diğerleri (2005) tarafından yapılan araştırmada ikiden

fazla alt grubun varlığında, referans grup ile diğerleri olmak üzere iki grup oluşturulmuş ve bu iki grup arasında karşılaştırmalar yapılmıştır. Ancak odak grupların birleştirilmesi neticesinde, farklı alt gruplar arasındaki performans farklılıklarının (örn. erkek göçmenlere karşı göçmen olmayan kadınlar) birbirlerinden ayrılamaması nedeniyle, bu yaklaşımda bazı alt grupların göz ardı edildiği belirtilmiştir. Ayrıca bu yaklaşımda bir odak grubun avantajlı, başka bir odak grubun ise dezavantajlı olması durumunda, maddelerdeki mevcut DMF'nin belirlenemeyeceği açıklanmıştır (Langer, 2008; Mapuranga vd., 2008; Oshima vd., 2015).

Odak grup birleştirme yaklaşımda ikili karşılaştırmalardan kaçınılması nedeniyle Tip I hata yapma olasılığının düştüğü ifade edilmiştir. Öte yandan bu yaklaşımda bütün odak grupların tek bir grup altında toplanmasının, veri setinin yapısının bozulmasına ve bilgi kaybına neden olduğu belirtilmiştir. Böylece normal şartlarda referans ve odak grupların herhangi biri arasında DMF gösteren madde veya maddelerin, odak grupların birleştirilmesi sonucunda DMF'li olarak belirlenemeyebileceği açıklanmıştır. Ayrıca bu yaklaşım tercih edildiğinde, örneklemelerin büyüklüğüne bağlı olarak DMF'li maddelerin hatalı belirlenmesinin de mümkün olabileceği ifade edilmiştir (Kanjee, 2007; Magis vd., 2011). Bu bilgiler doğrultusunda odak grup birleştirme yaklaşımda Tip I hata yapma olasılığının düşmesi genellikle bir avantaj olarak görülürken bazı durumlarda DMF'li maddelerin gözden kaçabilmesinin dezavantaj olduğu vurgulanmıştır.

İkiden fazla grup için dilsel veya kültürel farklılıklar gibi özelliklerin incelendiği araştırmalarda DMF analizleri yapılırken başvuru ikili karşılaştırma yaklaşımda Tip I hata yapma olasılığının artması, bileşik grup yaklaşımda referans ve odak gruplar arasındaki bağımsızlığın ihlal edilmesi, odak grup birleştirme yaklaşımda ise odak gruplar arasındaki farklılıkların göz ardı edilmesi gibi sınırlılıklar sebebi ile, Ellis ve Kimmel (1992) ile Kim ve diğerleri (1995) tarafından ikiden fazla grubu eşzamanlı olarak ele alabilecek yeni bir yaklaşıma ihtiyaç duyulduğu belirtilmiştir. Bu araştırmacılar tarafından yapılan öneriye benzer şekilde Linn'in (1993) çalışmasında, Kübalı, Meksikalı, Porto Rikolu ve Hispanik grupların tek bir odak grup altında toplanmasının uygun olmadığı ve gruplar arasındaki olası farklılıkları eşzamanlı olarak belirleyebilecek yeni tekniklerin geliştirilmesi gerektiği ifade edilmiştir.

Çoklu grupların varlığında ikili karşılaştırma yaklaşımının tercih edilmesi, ikiden fazla grubun eş zamanlı olarak dikkate alınabildiği Tek Yönlü ANOVA yerine, ikişerli karşılaştırmalar içeren bağımsız örneklem t-testinin kullanılmasına benzemektedir (Jin vd., 2015). Tek Yönlü ANOVA'nın çok sayıda yapılan bağımsız örneklem t-testinden

istatistiksel olarak üstün olduğu gibi, ikiden çok grupta DMF analizlerini eş zamanlı olarak yapmanın da ikili karşılaştırma yaklaşımından daha etkili olduğu belirtilmektedir (Penfield, 2001). Sonuç olarak ikiden fazla alt grup için üç farklı yaklaşımdan herhangi biri tercih edildiğinde oluşan sınırlılıkların üstesinden gelmek amacıyla çoklu gruplarda eşzamanlı olarak analiz yapmaya olanak sağlayan DMF belirleme tekniklerinin geliştirildiği görülmektedir.

Alanyazındaki araştırmalarda iki grup için geliştirilen tekniklerin çoklu gruplar için uygun olmadığı belirtilmiş ve ancak istatistiksel tekniklerin kuramsal olarak ikiden fazla sayıda gruba genelleştirilebildiği taktirde çoklu gruplar için kullanılabilmesinin mümkün olduğu açıklanmıştır (Cao, Tay ve Liu, 2017; Gamerman vd., 2018; Kim, 2003; Langer, 2008; Magis vd., 2010; Millsap ve Everson, 1993; Muraki, 1999; Penfield, 2001; Rockoff, 2018; Soares, Gonçalves ve Gamerman, 2009; Svetina ve Rutkowski, 2014; Tay vd., 2015; Woods, 2009b; Woods, Cai ve Wang., 2013). İlk olarak Landis, Heyman ve Koch (1978) tarafından yapılan araştırmada MH tekniğinin geliştirilerek çoklu gruplarda da kullanılabileceği ortaya konulmuştur. Wang (1998) ise araştırmasında MTKOO tekniğinin geliştirilerek ikiden fazla grup için de kullanılabileceğini belirtmiştir. Ayrıca Woods'un (2009b) çalışmasında MIMIC modelin çoklu gruplarda DMF belirlemek amacıyla kullanılabileceği bildirilmiştir. Benzer şekilde iki grup için kullanılmak üzere geliştirilen LRL tekniğinin geliştirilerek çoklu gruplar için de kullanılabileceği Magis ve diğerleri (2015) tarafından ifade edilmiştir. Böylece gelecekteki çalışmalarda iki grup için geliştirilen bütün DMF belirleme tekniklerinin geliştirilerek çoklu gruplar için de kullanılabilmesi amacıyla kuramsal araştırmaların yapılabileceği açıklanmıştır (Socha, DeMars, Zilberberg ve Phan, 2015). Nitekim referans ve odak gibi iki grubun dikkate alındığı araştırmalarda geliştirilmiş çoklu grup tekniklerinin kullanılabilmesi mümkünken (León, Martínez-Huertas, Olmos, Moreno ve Escudero, 2019; Penfield, 2001), iki grup için geliştirilen tekniklerin çoklu gruplarda kullanılamaması nedeniyle, gerçek test uygulamalarında çoklu grup tekniklerinin tercih edilmesinin avantajlı olduğu sonucuna ulaşılmıştır.

Alanyazında yapılmış olan araştırmalarda tek bir faktörün dikkate alındığı çalışmaların haricinde, cinsiyet (iki düzey) ve etnik köken (iki düzey) gibi iki veya daha çok sayıdaki değişkenlerin de ele alındığı DMF çalışmalarına rastlamanın mümkün olduğu ifade edilmektedir (Tay vd., 2016). Bu tür araştırmalarda genellikle iki farklı yaklaşım benimsenmektedir. Bu yaklaşımların birincisinde cinsiyet ve etnik köken için ayrı ayrı DMF belirleme çalışması yapılırken (Schauberger ve Mair, 2020), ikinci

yaklaşımında bu iki değişkenin etkileşimi ile oluşan dört grup arasında DMF incelenmektedir. Böylece ilk yaklaşımda, DMF analizi iki faktör için ardışık olarak yürütülmekte ve önce cinsiyet değişkeni, daha sonra ise etnik köken değişkeni için DMF analizleri yapılmaktadır. İkinci yaklaşımda ise cinsiyet ve etnik köken değişkenlerinin düzeyleri çaprazlanıp, dört düzeyli sahte (pseudo) bir grup faktörü oluşturulduktan sonra çoklu grup teknikleri ile DMF analizleri gerçekleştirilmektedir. Belirtilen ikinci yaklaşım, faktöriyel DMF yaklaşımı olarak isimlendirilmekte ve bu yaklaşımın bazı araştırmacılar tarafından tercih edildiği görülmektedir (Jin vd., 2015; Wang, 2000).

Çoklu grup DMF belirleme teknikleri bir referans ve bir odak grup olmak üzere iki grup için kullanılabilen tekniklerin, odak grup sayısı birden fazla olacak şekilde genelleştirilen uzantılarıdır (Magis vd., 2010). İki grubun ele alındığı durumlara benzer şekilde çoklu gruplarda yapılan DMF belirleme çalışmalarında da gruplardan birinin avantajlı olduğu varsayılmakta ve bu grup referans olarak dikkate alınmaktadır. Birden fazla odak grubun bağımsız olarak ele alındığı bu yaklaşımda, bu grupların tamamının dezavantajlı oldukları kabul edilmektedir (Davidson, Wortzman, Ko ve Li, 2021; Gamerman vd., 2018; Penfield, 2001). Böylece çoklu grup teknikleri kullanılarak yapılan çalışmalarda örneklemdaki olası bütün alt gruplar incelenebilmekte (Magis vd., 2011; Rollins III, 2018) ve testteki maddelerin odak gruplardan herhangi birinin aleyhine farklı çalışıp çalışmadığı belirlenebilmektedir (Goodman vd., 2011; Noortgate ve De Boeck, 2005; Wyse ve Mapuranga, 2009).

Çoklu gruplarda DMF belirlemek amacıyla yapılan ilk araştırmada, Kim ve diğerleri (1995) Lord tarafından ortaya konulan ki-kare istatistiğini genelleştirerek, çoklu gruplar için kullanılabilen ilk DMF belirleme tekniği olan Genelleştirilmiş Lord Ki-kare (GLK) tekniğini alanyazına kazandırmışlardır. GLK tekniğini ilk kez çalışmalarında kullanan Kim ve diğerleri (1995), ikiden fazla grup için eşzamanlı olarak DMF belirlemenin ikili grup karşılaştırmalarından daha etkili ve kullanışlı olduğunu belirtmişlerdir. GLK tekniği ortaya konulduktan sonra çoklu gruplarda madde ve madde takımları düzeyinde DMF belirleyebilmek amacıyla parametrik ve parametrik olmayan teknik veya yaklaşımlar ortaya konulmuş olmasına rağmen (Battauz, 2019; Choi, Gibbons ve Crane, 2011; Gierl ve Bolt, 2001; Jin vd., 2015; Langer, 2008; Magis vd., 2011; Oshima vd., 2015; Penfield, 2001; Schauburger ve Tutz, 2016; Strobl, Kopf ve Zeileis, 2015; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) bu araştırma kapsamında yalnızca ikili puanlanan maddelerde DMF belirlemek amacıyla geliştirilen teknikler dikkate alınmıştır.

İkili puanlanan maddelerde iki grup için kullanılan DMF belirleme tekniklerinin matematiksel formüllerinin genelleştirilmesi sonucunda; Genelleştirilmiş Mantel Haenszel (GMH) (Penfield, 2001), Genelleştirilmiş Lojistik Regresyon (GLR) (Magis vd., 2011), Lordif (Choi vd., 2011), Langer Genelleştirilmiş Çoklu Grup Wald (LGÇGW) (Langer, 2008), Battauz Genelleştirilmiş Çoklu Grup Wald (BGÇGW) (Battauz, 2019), Raju Telafi Edici Olmayan Çoklu Grup DFIT (MG-NDIF) (Oshima vd., 2015), MIMIC model (Woods, 2009b) ve Lojistik Regresyon Lasso (LRL) (Tutz ve Schauburger, 2015) teknikleri ortaya konulmuştur. Ayrıca Faktöriyel Lojistik Regresyon (FLR) (Jin vd., 2015), DMFboost (Schauburger ve Tutz, 2016) ve çok değişkenli uç değer belirleme (Magis ve De Boeck, 2011) yaklaşımlarından da çoklu gruplarda DMF belirleme sürecinde yararlanılabileceği açıklanmıştır. Tek boyutlu modellerde ikili puanlanan maddelerin varlığında (Raju, van der Linden ve Fleer, 1995) kullanılmak üzere geliştirilmiş olan bu çoklu grup teknikleri ile analizlerin madde düzeyinde yapılabildiği belirtilmiştir. Yapılan bu araştırma kapsamında çoklu grup teknikleri arasından sıklıkla tercih edildiği görülen GLK, GLR-LRT, GLR-Wald, GMH ve Lordif dikkate alınmıştır.

İki grup için geliştirilen tekniklere benzer şekilde çoklu grup tekniklerinden GMH ve GLR'nin KTK'ya dayalı, GLK ve Lordif'in ise MTK'ya dayalı olduğu ifade edilebilmektedir. Öte yandan GMH ile yalnızca TB DMF gösteren maddeler belirlenebilirken, GLK, GLR ve Lordif teknikleri ile hem TB hem de TBO(a), TBO(ab) DMF'li maddeleri belirlemenin mümkün olduğu belirtilebilmektedir (Jin vd., 2015; Magis vd., 2011; Penfield, 2001; Svetina ve Rutkowski, 2014). Ayrıca Magis ve diğerleri (2011) tarafından geliştirilen GLR tekniğinde Wald ki-kare ve Olabilirlik Oran (Likelihood Ratio) olmak üzere iki istatistiksel testin kullanılabilmesi nedeniyle araştırmalarda bu iki testten elde edilen sonuçların karşılaştırılmasının yararlı olacağı açıklanmaktadır (Finch, 2016; Jin vd. 2015).

DMF ile ilgili Türkiye ve Türkiye dışında yapılan araştırmalarda çoğunlukla iki grup arasındaki DMF'yi belirlemek amacıyla farklı tekniklerin kullanıldığı belirtilmesine rağmen, eğitim ve psikometri gibi birçok alanda ikiden fazla grubun aynı anda analize dahil edilmesinin de gerekli olduğu vurgulanmaktadır (Carroll, 2015; Dell-Ross, 2021; Finch, 2016). Grup sayısı ikiden fazla olduğunda ikili karşılaştırma, bileşik grup ve odak grup birleştirme yaklaşımlarına dayalı olarak gerçekleştirilen DMF belirleme çalışmalarında; Tip I hata yapma olasılığı yükselmekte, çok sayıda analiz yapılması nedeniyle zaman kaybı yaşanmakta ve dikkate alınan gruplar için bilgi kaybı söz konusu olabilmektedir. Ancak 1995 yılından sonra çoklu gruplarda eş zamanlı olarak DMF

incelenmesine olanak sağlayan istatistiksel tekniklerin geliştirilmesiyle birlikte, örneklemdaki alt gruplar bağımsız olarak ayrı ayrı dikkate alınabilmekte, tek bir hipotez test edildiğinden Tip I hata oranı kontrol altına alınabilmekte ve zaman kaybı yaşanmayarak hesaplama verimliliği sağlanabilmektedir (Langer, 2008; Fidalgo, 2011; Penfield, 2001; Rollins III, 2018). Böylece ikili karşılaştırma, bileşik grup ve odak grup birleştirme yaklaşımlarında kullanılan ikili grup DMF belirleme tekniklerine göre daha geçerli, kullanışlı ve istatistiksel olarak etkili olan çoklu grup tekniklerinden yararlanılabilmektedir (Carroll, 2015; Kim vd., 1995; Magis vd., 2011; Penfield, 2001). Bu bilgiler doğrultusunda çoklu gruplarda eş zamanlı olarak DMF incelenmesinin gerekliliği ve daha iyi performans gösteren istatistiksel tekniklere duyulan ihtiyaç göz önüne alındığında (Dell-Ross, 2021; Langer, 2008; Oshima vd., 2015) çoklu grup tekniklerine ilişkin performansların bir simülasyon çalışması kapsamında karşılaştırılması son derece önemli görülmektedir.

İkili puanlanan maddelerde çoklu grup tekniklerinden GLK, GLR-LRT, GLR-Wald, GMH ve Lordif'in kullanıldığı araştırmalar incelendiğinde; belirtilen tekniklerinden yalnızca birinin dikkate alındığı çalışmaların (Battauz, 2019; Carroll, 2015; Dell-Ross, 2021; Fidalgo ve Scalón, 2010; Jin vd., 2015; Kim vd., 1995; Penfield, 2001; Svetina ve Rutkowski, 2014) yanı sıra, bu tekniklerden iki veya üçünü karşılaştıran araştırmaların da yapıldığı belirlenmiştir (Finch, 2016; Rollins III, 2018; Magis vd., 2011; Morales-Ortiz ve Peña-Cardenas, 2019; Uyar ve Kaya Uyanık, 2016). Yapılmış olan bu çalışmalarda ilk olarak, genelleştirilmiş DMF belirleme tekniklerinin kuramsal temellerine ilişkin bilgiler verilmiş, sonrasında ise ulusal veya uluslararası test uygulamalarından elde edilen gerçek veriler (Kim vd., 1995; Magis vd., 2011; Morales-Ortiz ve Peña-Cardenas, 2019; Uyar ve Kaya Uyanık, 2016) veya simülasyon çalışmaları kapsamında üretilen simülatif veriler (Battauz, 2019; Carroll, 2015; Dell-Ross, 2021; Fidalgo ve Scalón, 2010; Finch, 2016; Jin vd., 2015; Penfield, 2001; Rollins III, 2018; Svetina ve Rutkowski, 2014) kullanılarak, tekniklerin performansları karşılaştırılmıştır.

Alanyazında ikiden fazla grupta çoklu grup tekniklerinin kullanıldığı az sayıda araştırmanın var olması nedeniyle bazı araştırmacılar tarafından yeni çalışmalara ihtiyaç duyulduğu açıklanmıştır (Fidalgo ve Scalón, 2010; Finch, 2016; Penfield, 2001; Svetina ve Rutkowski, 2014). Penfield (2001) tarafından yapılan çalışmada, GMH ile GLK tekniklerinin, Langer'ın (2008) araştırmasında ise GLK tekniğinin çeşitli koşullardaki performanslarının belirlenmesi ve diğer tekniklerle karşılaştırılması amacıyla yeni simülasyon çalışmalarının yapılması gerektiği belirtilmiştir. Benzer şekilde Magis ve

diğerleri (2011) tarafından yapılan çalışmada GLR tekniğinin performansının GMH ve GLK tekniklerinin performansları ile karşılaştırılması gerektiği önerilmiştir. Finch (2016) ise hem TB hem de TBO DMF'yi belirlemek amacıyla kullanılabilen bazı çoklu grup DMF belirleme tekniklerinin üstünlük ve sınırlılıklarının farklı koşullar altında tespit edilmesinin çok önemli olduğunu vurgulamıştır. Böylece yapılan incelemeler sonucunda çoklu grup tekniklerine ilişkin performansların araştırıldığı çalışmalarda henüz birçok tekniğin birbirleri ile karşılaştırılmadığı görülmüş ve bu nedenle belirtilen teknikleri araştırmalarında kullanmak isteyen uygulayıcılar ve araştırmacılar için yol gösterici başka çalışmaların yapılması gerektiği anlaşılmıştır.

Araştırmacıların belirttikleri önerilerinden yola çıkarak, çoklu grup tekniklerinden hangisinin kullanılması durumunda daha doğru sonuçların elde edilebileceğine ilişkin bilimsel kanıtların ortaya konulmasının gerekli olduğu sonucuna ulaşılmaktadır. Gerçek veri kullanılarak yapılan analizler sonucunda DMF belirleme tekniklerinin karşılaştırılmasının zor olması nedeniyle, simülatif veriler üretilerek yeni çalışmaların yapılması gerekmektedir. Böylelikle çeşitli koşulların (örn, DMF'nin miktarı, örneklem büyüklüğü, yetenek dağılımı vb.) incelendiği simülasyon çalışmaları sonucunda, bazı çoklu grup teknikleri için etkili olan koşulların diğerleri için de etkili olup olmadığı belirlenebilmektedir (Fidalgo, Ferreres ve Muñiz, 2004a; Rollins III, 2018). Bu bağlamda farklı simülasyon koşulları dikkate alınarak; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinden hangisinin daha doğru sonuçlar ortaya koyacağını tespit edilmesine çalışıldığı mevcut araştırmada dikkate alınan beş tekniğin daha önce karşılaştırılmaması nedeniyle önemli olduğu ifade edilebilir.

Simülasyon çalışmalarında tekniklerin performansları karşılaştırılırken; hesaplanan istatistiksel güç, Tip I hata ve Tip III hata oranları değerlendirilerek istatistiksel tekniklerin hangi koşullar altında kullanılmasının daha yararlı olduğu belirlenebildiğinden (Macdonald, 1999), araştırmalardaki temel ilgi noktasının sırasıyla Tip I hata, istatistiksel güç ve Tip III hata olması gerekmektedir (Morris, White ve Crowther, 2019; Tureson ve Odland, 2018). Benzer durum DMF belirleme çalışmaları için de geçerlidir. DMF tekniklerinin performanslarının incelendiği bu çalışmalarda da tekniklerin; duyarlılık (İstatistiksel güç, "Sensitivity"), özgüllük (1-Tip I hata, "Specificity") (Camilli, 2013; Cuevas ve Cervantes, 2012) ve Tip III hata oranları (Lopez Rivas, 2012) açısından değerlendirilmeleri gerekmektedir. Mükemmel duyarlılığa sahip olan bir teknik, DMF'li tüm maddeleri doğru bir şekilde belirleyebilmekte, mükemmel özgüllüğe sahip olan bir teknik ise DMF içermeyen maddelerin doğru bir şekilde DMF'li

olmadığını tespit edebilmektedir (Camilli, 2013). Ayrıca bir teknik hangi DMF türünü tespit etmek amacıyla geliştirildi ise diğer DMF türündeki maddeleri DMF'li olarak işaretleyememesi beklenmektedir. Yani TBO DMF'yi tespit edebilmesi amacıyla geliştirilen tekniklerin, gerçekte TB DMF'li maddeleri DMF'li olarak işaretleyememesi amaçlanmakta, aksi taktirde gerçekte TB DMF gösteren bir maddenin yapılan analizler sonucunda TBO DMF'li olarak belirlenmesi DMF türü açısından hatalı yorumların yapılmasına sebep olabilmektedir (Lopez Rivas, 2012). Özetle gerçekte DMF içeren maddelerin, yine DMF'li olarak tespit edilmesi istatistiksel güce, gerçekte DMF göstermeyen maddelerin yapılan analizler sonucunda yanlışlıkla DMF'li olarak belirlenmesi Tip I hataya ve DMF türünün hatalı tespit edilmesi durumu ise Tip III hataya karşılık gelmektedir. Lopez Rivas (2012), tarafından da belirtildiği üzere istatistiksel güç ve Tip III hata arasında DMF türünü doğru sınıflama açısından bir ilişki vardır ve gerçekte DMF'li olan maddelerin DMF türünün doğru belirlenmesi istatistiksel güce, yanlış belirlenmesi ise Tip III hataya işaret etmektedir. Araştırmalarda DMF belirleme tekniklerine ilişkin hesaplanan istatistiksel güç oranlarının yüksek, Tip I hata ve Tip III hata oranlarının ise düşük olması istenilen durum olarak belirtilmektedir.

Alanyazındaki birçok araştırmada da belirtildiği gibi genellikle azınlık ve çoğunluk olmak üzere çeşitli özellikler açısından (örneğin, ırk/etnik köken, test uygulama yeri, ana dil, sosyo-ekonomik durum, müdahale yöntemi, tedavi koşulları vb.) farklı grupların dikkate alındığı gerçek test uygulamalarında, ikiden fazla grup arasında maddeler için DMF incelemelerinin yapılması gerekebilmektedir. Bu gereklilik üzerine önceleri ikili karşılaştırma, bileşik grup ve odak grup birleştirme yaklaşımlarına başvurulmuş, ancak bazı sınırlılıklar sebebi ile grupların eş zamanlı olarak incelenmesini sağlayan tekniklerin geliştirilerek gerçek uygulamalarda bu tekniklerin kullanılması gerektiği belirtilmiştir. Nitekim 1995 yılından günümüze kadar geçen süreçte çoklu grup teknikleri adı altında yeni teknikler geliştirilmesine rağmen, bazı tekniklere ilişkin performansların birbirleri ile karşılaştırılmadığı görülmüştür. Bu bağlamda ikili puanlanan maddelerde GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinden en az birini gerçek test uygulamalarında kullanmak isteyen araştırmacıların, bu teknikleri hangi koşullar altında tercih etmeleri gerektiğine ilişkin yol gösterecek bir simülasyon çalışması yapılmasının gerekli olduğu belirlenmiştir. İkili puanlanan maddeler için çoklu grup DMF belirleme tekniklerinin farklı simülasyon koşullarındaki performanslarının Tip I hata, istatistiksel güç ve Tip III hata oranları açısından karşılaştırılmasına duyulan bu gereklilik, araştırmanın problemi oluşturmuştur.

Alanyazındaki çalışmalarda DMF türü, DMF miktarı, DMF’li madde yüzdesi, madde sayısı, referans ve odak grupların örneklem büyüklükleri, odak grup sayısı, yetenek ve madde parametrelerinin dağılımları gibi çok sayıda faktörün DMF belirleme tekniklerinin performanslarını etkileyebildiği açıklanmasına rağmen (Battaüz, 2019; Carroll, 2015; Dell-Ross, 2021; Finch, 2016; Magis vd., 2010; Penfield, 2001; Svetina ve Rutkowski, 2014) yapılan bu simülasyon çalışmasında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin odak grup sayısından, referans grubun örneklem büyüklüğünden, referans ve odak grupların örneklem büyüklüğü oranından, DMF miktarından ve DMF türünden [TB, TBO(*a*), TBO(*ab*)] etkilenip etkilenmediği incelenmiştir. Böylece yapılan bu simülasyon çalışmasında GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin performansları Tip I hata, istatistiksel güç ve Tip III hata oranları açısından karşılaştırılarak gerçek uygulamalarda hangi tekniğin tercih edilmesinin daha uygun olacağı belirlenmeye çalışılmıştır.

Amaç

Araştırmanın genel amacı ikili puanlanan maddelerde çoklu gruplar için kullanılabilen DMF belirleme tekniklerinin performanslarını, çeşitli simülasyon koşullarında Tip I hata, istatistiksel güç ve Tip III hata oranları açısından karşılaştırmaktır. Bu amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

1. Grup sayısının 3 ve 4 olduğu durumda; referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3), DMF miktarı [*a*(0.0, 0.4, 0.7), *b*(0.0, 0.4, 0.7)] ve DMF türü [DMF Yok, TB, TBO(*a*), TBO(*ab*)] koşulları için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin;
 - a. Tip I hata oranları,
 - b. İstatistiksel güç oranları,
 - c. Tip III hata oranları nasıl değişmektedir?
2. Grup sayısı (3, 4), referans grubun örneklem büyüklüğü (600, 1800), DMF miktarı (DMF Yok, Düşük DMF, Orta DMF, Yüksek DMF) ve DMF türü [DMF Yok, TB, TBO(*a*), TBO(*ab*)] koşulları için, GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin;
 - a. Tip I hata oranları,
 - b. İstatistiksel güç oranları,
 - c. Tip III hata oranları nasıl değişmektedir?

Önem

İkili puanlanan maddelerin dikkate alındığı araştırmaların çoğunda, referans ve odak olmak üzere iki grup arasında DMF belirleme çalışmalarının yapıldığı bilinmektedir. Grup sayısının üç veya daha fazla olduğu çoklu gruplarda ise ikili karşılaştırma, bileşik grup ve odak grup birleştirme yaklaşımlarına dayalı olarak yapılmış olan çok sayıda araştırmanın var olduğu görülmektedir. Özellikle ulusal ve uluslararası düzeydeki geniş ölçekli test uygulamalarına çok sayıda grupta yer alan bireyler katılmasına rağmen geliştirilen DMF belirleme tekniklerinin birçoğu yalnızca iki grubun karşılaştırılmasına imkân tanımaktadır. DMF'li maddelerin doğru belirlenmesinin hayati önem taşıdığı bilinmesine rağmen, farklı demografik özelliklere sahip çoklu grupların varlığında ikili karşılaştırma, bileşik grup ve odak grup birleştirme yaklaşımlarının tercih edilmesi, çok sayıda hipotezin test edilmesinden kaynaklanan Tip I hata yapma olasılığının artmasına ve elde edilen sonuçların hatalı değerlendirilmesine neden olabilmektedir. Bu nedenle daha doğru sonuçlara ulaşabilmek amacıyla ikiden fazla grubun bağımsız ve eşzamanlı olarak ele alınabildiği çoklu grup tekniklerinin geliştirilerek araştırmalarda kullanılması gerektiği vurgulanmaktadır. Çoklu grup DMF belirleme tekniklerinin sınırlı sayıda araştırmada kullanılması nedeniyle yapılan bu simülasyon çalışmasının, çoklu grup teknikleri ile ilgilenecek olan araştırmacılar için kılavuz niteliğinde olduğu düşünülmektedir.

Çoklu gruplarda DMF tespit etmek amacıyla birçok farklı parametrik ve parametrik olmayan istatistiksel teknik geliştirilmesine rağmen, bu tekniklerin avantajlı veya dezavantajlı oldukları koşulların farklılaşabilmesi nedeniyle, çalışmalarda hangi tekniğin kullanılması gerektiği araştırmacıların temel endişesi olabilmektedir. İlgili alanyazındaki birçok araştırmada genelleştirilmiş teknikler kullanılmasına rağmen, bu tekniklerin farklı koşullar altındaki etkililiğinin incelendiği görece az sayıda simülasyon çalışmasının yapıldığı görülmektedir. Bu nedenle hem üç hem de dört grup için çoklu grup tekniklerinin karşılaştırıldığı bu simülasyon çalışmasının ilgili alana katkı sağlaması hedeflenmektedir.

Çoklu gruplar için DMF çalışmalarındaki önemli bir boşluğu doldurmak amacıyla yapılan bu araştırmada, önceki simülasyon çalışmalarından farklı olarak referans grubun örneklem büyüklüğü, odak grupların örneklem büyüklüklerinin referans gruba oranı, DMF miktarı ve DMF türü faktörlerine ilişkin farklı koşullar belirlenmiş ve üretilen simülatif veriler GLK, GLR-LRT, GLR-Wald, GMH, Lordif teknikleri ile analiz

edilmiştir. Yapılan bu simülasyon çalışması ile hem daha önce ele alınmış olanlardan farklı koşulların değişimlenmesi hem de farklı tekniklerin dikkate alınması nedeniyle, elde edilen sonuçların DMF ile ilgili alanyazındaki bilgi birikimine katkı sağlaması hedeflenmiştir. Bu kapsamda çoklu grup tekniklerinin farklı koşullar altındaki performanslarının incelendiği bu araştırmanın, tekniklerin etkili olduğu koşullara ilişkin önemli bilgiler sunması nedeniyle, gerçek veri üzerinde çalışma yapmayı planlayan ölçme ve değerlendirme uzmanlarına, psikometristlere, ölçek geliştiren ve uyarlayan araştırmacılara ve test uygulayıcılarına yol gösterici olacağı düşünülmüştür.

Çoklu gruplar üzerinde yapılan DMF araştırmalarında genellikle güçlük parametresinden kaynaklı olan TB ve ayırt edicilik parametresinden kaynaklı olan TBO DMF’li maddelerin incelendiği görülmektedir. Bununla birlikte ilgili alanyazında hem güçlük hem de ayırt edicilik parametrelerinin birlikte dikkate alındığı çok az sayıda çalışmaya rastlanılmaktadır. Yapılan bu araştırmada ise yalnızca güçlük ve yalnızca ayırt edicilik parametrelerinin yanı sıra hem güçlük hem de ayırt edicilik parametresinden kaynaklı olan üç farklı DMF türü [TB, TBO(*a*), TBO(*ab*)] dikkate alınmaktadır. Bu anlamda yapılan araştırmanın ilgili alanyazına öncülük etmesi açısından önemli olduğu belirtilebilmektedir. Ayrıca daha önceki çalışmalarda çeşitli çoklu grup DMF belirleme teknikleri ele alınmış olsa da belirtilen üç DMF türü için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin performanslarını Tip I hata, istatistiksel güç ve Tip III hata oranları açısından karşılaştıran hiçbir çalışma bulunmamaktadır. Bu nedenle araştırmanın özgün olduğu düşünülmekte ve elde edilen sonuçların alanyazındaki mevcut bilgi birikimine katkı sağlayacağına inanılmaktadır.

Genellikle simülasyona dayalı DMF belirleme çalışmalarında istatistiksel tekniklerin performansları değerlendirilirken Tip I hata ve istatistiksel gücün dikkate alındığı bilinmekte, Tip III hatanın ise göz ardı edildiği çoğu araştırmacı tarafından belirtilmektedir. İstatistiksel tekniklerin performanslarının üç farklı değerlendirme kriterine göre karşılaştırılmasının daha doğru sonuçlara ulaşmak açısından önemli olduğunun vurgulanması nedeniyle, yapılan bu simülasyon çalışmasında çoklu grup tekniklerinin performanslarını değerlendirmek için Tip I hata ve istatistiksel güç oranlarının yanı sıra Tip III hata oranları da hesaplanmıştır. Tip III hatanın daha önce yapılan sınırlı sayıda DMF çalışmasında dikkate alınmış olması nedeniyle, bu araştırma sonuçlarının hem Türkiye hem de Türkiye dışındaki DMF ile ilgili alanyazına katkı sağladığı düşünülmektedir. Sonuç olarak; belirtilen çoklu grup tekniklerinin dikkate alınan koşullar altındaki performanslarının üç farklı kritere göre karşılaştırıldığı ilk

alıřma niteliğinde olan bu arařtırmanın, DMF ile ilgilenen arařtırmacılara derin bir bakıř aısı sunması aısından önemli olduđu kabul edilmektedir.

Gerek veri üzerinde alıřan arařtırmacılar ve uygulayıcıların testteki maddelerin oklu gruplar arasında DMF gsterip gstermediđini ve olası DMF’li maddelerin DMF trn analiz ncesinde bilememeleri nedeniyle, zellikle test geliřtirmenin ilk deneme uygulamalarında hangi oklu grup tekniđini tercih etmeleri gerektiđi konusunda kararsız kaldıkları bilinmektedir. Bu nedenle arařtırmacıların hem kendi arařtırmalarındaki veri setlerine uygun hem de yksek istatistiksel g ile dřk hata (Tip I, Tip III) oranlarına sahip oklu grup tekniđini kullanabilmeleri amacıyla simlasyon sonularından yararlanmalarının önemli olduđu vurgulanmaktadır. Bu bađlamda gerek uygulamalardaki DMF’li maddeleri ve bu maddelerdeki DMF trn dođru ve etkili bir řekilde belirlemek isteyen arařtırmacılara GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinden hangisi veya hangilerini, hangi kořullar altında kullanmaları gerektiđi hususunda yol gsterici olan bu simlasyon alıřmasının, DMF ile ilgilenen arařtırmacıların bařvuracakları bir kaynak olduđu dřnlmektedir.

Sınırlılıklar

Her arařtırmada olduđu gibi, bu alıřmanın da simlasyon deseni ve kullanılan DMF teknikleri aısından sınırlılıkları vardır. oklu gruplarda DMF belirlemek amacıyla ikili ve oklu puanlanan maddeler iin ok sayıda istatistiksel teknik geliřtirilmesine rađmen, bu arařtırma yalnızca ikili puanlanan maddelerde kullanılabilen beř oklu grup DMF belirleme tekniđi (GLK, GLR-LRT, GLR-Wald, GMH ve Lordif) ile sınırlı tutulmaktadır. Ayrıca arařtırma kapsamında simlatif veri retilmesi nedeniyle elde edilen sonuların genelleřtirilebilirliđi, deđiřimlenen ve sabit tutulan simlasyon kořulları ile sınırlı kalmaktadır.

alıřmada 2PLM, 20 madde ve %20 DMF’li madde yzdesi dikkate alınarak veri retilmiř ve deđiřimlenen kořulların gerek uygulamalara benzer olması amalansa dahi, gerek uygulamaların kapsamı erevesinde deđiřimlenebilecek btn faktrlerin kapsam ve zaman sınırlaması nedeniyle bu alıřmada dikkate alınamamıř olması, bu arařtırmanın bir diđer sınırlılıđı olarak belirtilebilmektedir.

oklu gruplarda kullanılabilen DMF belirleme tekniklerinin performanslarının arařtırıldıđı bu alıřmada geniř lekli test uygulamalarının kapsamının ele alınması amalansa da ilgili arařtırmalardaki grup sayıları ve bu alıřmanın simlasyon

desenindeki olası toplam koşul sayılarına bağılı olarak, gerekli hesaplamaların sorunsuz bir şekilde gerçekleştirilebilmesi için yalnızca üç ve dört grup ile sınırlı tutulmaktadır.

Yapılan bu çalışmada gerçek uygulamalarda DMF'li maddelerin testten çıkarılmasının mümkün olmadığı özel durumlar dikkate alınmış ve ek bir saflaştırma (purification) işlemi yapılmadan çoklu grup tekniklerine ilişkin performansların karşılaştırılmasına karar verilmiştir. Ayrıca yapılan bu simülasyon çalışmasında grup sayısının artmasına bağılı olarak, koşul sayısı ve tekrar sayısının çarpımından oluşan satır sayısına ulaşıncaya kadar veri üretimi devam ettiğinden, gerekli hesaplama süreleri göz önünde bulundurularak tekrar sayısı 100 olarak sınırlandırılmıştır.

Psikometri alanında genel olarak bazı kavramların Türkçe kaynaklardaki karşılıklarının farklılaşmasının yanı sıra bu kavramların İngilizcelerine olan aşinalık nedeniyle alanyazında sıklıkla kullanılan RMSE, ML, MML, GPCM, GRM, GLR-LRT gibi birçok kavramın yeniden Türkçe'ye çevrilmesinin zor ve okuyucular için kafa karışıklığı yaratabilmesi nedeniyle, yapılan araştırma kapsamında belirtilen kavramların İngilizce kısaltmalarının kullanılmasına karar verilmiştir. İlgili kavramların belirtilen nedenlerle Türkçe'ye çevrilememiş olması bir sınırlılık olarak kabul edilmiştir.

BÖLÜM 2

KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Ölçme araçlarındaki maddelerin psikometrik özelliklerinin çeşitli gruplar arasında değişmez olup olmadığının belirlenmesi için DMF belirleme teknikleri kullanılarak birtakım analizlerin yapılması gerekmektedir (Clauser ve Mazor, 1998; Finch ve French, 2007; Gómez-Benito, Sireci, Padilla, Hidalgo ve Benítez, 2018; Osterlind ve Everson, 2009). Başka bir ifadeyle yapılan DMF çalışmalarında gruplar arasındaki farklı madde yanıt örüntülerini belirlemek amacıyla istatistiksel tekniklerden yararlanılabildiği belirtilebilmektedir. Yanıt örüntüleri değerlendirilirken ilk önce kadın-erkek, gelişmiş-gelişmekte olan, azınlık-çoğunluk gibi karşılaştırma grupları, ilgilenilen temel özelliğe (örn., sözel yetenek veya matematik başarısı) göre eşleştirilmekte, sonrasında ise maddelere ilişkin yanıt örüntüleri DMF belirleme teknikleri ile analiz edilmektedir (Zumbo ve Gelin, 2005).

Zumbo ve Gelin (2005) tarafından belirtildiği gibi DMF belirleme çalışmaları (a) testte eşitlik ve adaleti sağlamak, (b) iç geçerliğe yönelik olası bir tehditle başa çıkmak ve (c) farklı grupların madde yanıtlama davranışlarını anlamaya çalışmak, olmak üzere üç amaç doğrultusunda yapılabilmektedir. DMF içeren bir maddenin yanlı olma olasılığı bulunmakta, yanlı maddelerden oluşan testler ise bireylerin yeteneklerinin yanlı belirlenmesine (Ackerman, 1992; Kamata ve Vaughn, 2004) ve ölçme değişmezliğinin ihlal edilmesine neden olmaktadır. Bu nedenle DMF belirleme çalışmaları, madde yanlılığını tespit etme sürecinin ilk aşaması olarak kabul görmektedir (Dorans ve Holland, 1993).

İdeal olarak bir maddeyi doğru yanıtlama olasılığı, yalnızca bireylerin madde tarafından ölçülen yapı üzerindeki yetenek seviyesine bağlı olmalıdır (Zieky, 2006). Aksi takdirde aynı yetenek düzeyindeki alt grupların testte yer alan bir maddeyi doğru yanıtlama olasılıkları arasında fark olması, ilgili maddenin DMF içerdiğini ortaya koymaktadır (Clauser ve Mazor, 1998; Osterlind ve Everson, 2009; Sireci, Patsula ve Hambleton, çev. 2017). DMF'li maddeler test puanlarına ölçülmek istenen yapıdan bağımsız varyans ekleyebilmekte, bu durum ise test puanları üzerinde sistematik hata oluşmasına yol açarak, gruplardan birine avantaj sağlayabilmektedir (Jin vd., 2015, 2018;

Walker, 2011; Zieky, 2006; Zumbo, 2007). Demografik özellikler açısından farklı gruplarda yer alan bireyler için sistematik hatanın varlığı ise, bu bireylerin puanlarında çarpıklık oluştuğunu göstermektedir (Korkmaz, 2005). Bu bilgiler doğrultusunda DMF çalışmaları ile; birtakım demografik özelliklere göre gruplar arasında sistematik hata oluşmasına engel olmak, grup performanslarının haksız bir şekilde etkilenmesini önlemek ve bireyler hakkında daha doğru kararlar vermek amaçlanmaktadır (Camilli ve Shepard, 1994; Penfield, 2001).

Hem Klasik Test Kuramı (KTK) hem de tek boyutlu Madde Tepki Kuramı (MTK) modellerinde bir bireyin test performansının altında yatan yeteneğin tek boyutlu bir gizil özellik olduğu varsayılmaktadır. Başka bir ifadeyle iki temel kuram altında ele alınan tek boyutlu modeller için bireylerin testteki performanslarının yalnızca bir gizil yetenekten etkilendiği kabul edilmektedir (Ackerman, 1992; Camilli, 1992; Ellis ve Raju, 2003; Hambleton ve Swaminathan, 1985; Potenza ve Dorans, 1995). Ayrıca boyutluluk kavramı madde düzeyinde de dikkate alınmakta ve bir madde yalnızca bir özelliği ölçüyor ise o maddenin tek boyutlu, aynı anda iki ya da daha fazla sayıda özelliği ölçmeye çalışıyor ise, o maddenin çok boyutlu olduğu ifade edilmektedir (Ackerman, 1992; DeMars, 2010).

Belirli test maddelerinde bireylerin performanslarını etkileyen ikinci boyutun birinci boyutla ilişki veya ilişkisiz olabilmesinin mümkün olduğu, yapılmış olan birçok araştırmada (Ackerman, 1992; Camilli, 1992; Shealy ve Stout, 1993; Taylor ve Lee, 2012; Walker, 2011) açıklanmaktadır. Roussos ve Stout (1996) tarafından yapılan çalışmada testle ölçülen ikinci boyut; a) ölçülmesi amaçlanıyorsa yardımcı (auxiliary) boyut, b) ölçülmesi amaçlanmıyorsa rahatsız edici (nuisance) boyut olarak iki şekilde tanımlanmaktadır. Böylelikle ikinci boyutun yardımcı (ilişkili) veya rahatsız edici (ilişkisiz) olup olmadığına bakılmaksızın DMF oluşabildiği belirtilmektedir (Walker ve Beretvas, 2001). Başka bir anlatımla test maddeleri ile ölçülen ikinci bir boyutun var olması durumunda, ilk boyutta aynı yetenek düzeyinde olan bireylerin ikinci boyuttaki performanslarının farklılaşması, DMF oluşmasına yol açabilmektedir (Ackerman, 1992; Camilli, 1992; Roussos ve Stout, 1996; Taylor ve Lee, 2012; Walker ve Beretvas, 2001). Bu anlamda yardımcı boyutun neden olduğu DMF, tehlikesiz (benign) DMF olarak, rahatsız edici boyutun neden olduğu DMF ise olumsuz (adverse) DMF olarak isimlendirilmektedir (Roussos ve Stout, 1996; Walker ve Beretvas, 2001).

İkinci boyutun hem yardımcı hem de rahatsız edici olması durumunda, bu boyuttan kaynaklı olarak grupların yetenek dağılımlarının değişmesi nedeniyle madde karakteristik eğrileri de farklılaşmaktadır. Özetle maddeye ilişkin özelliklerin, grup

üyeliği ile etkileşimi sonucunda DMF'nin oluşabildiği ifade edilebilmekte (de Gruijter ve van der Kamp, 2008; Kulick ve Hu, 1989) ve DMF'nin temel nedeni olan ikinci boyutun, ölçülmek istenen yapıyla hem ilişkili hem de ilişkisiz olması durumunda testin geçerliği etkilenebilmektedir (Walker ve Beretvas, 2001). Örneğin dört işlem becerisini ölçtüğü iddia edilen bir çoktan seçmeli testin, bu beceriye yönelik maddelerden oluşması gerekmesine rağmen, maddelerin ölçmeyi amaçladığı yetenekten başka bir yeteneği (ilişki veya ilişkisiz) ölçmesi durumunda yapı geçerliğinin tehdit edildiği belirtilebilmektedir (Ackerman, 1992). Başka bir örnek ile açıklamak gerekirse; futbol tarihi ile ilgili bir metinde öğrencilerin okuduğunu anlama becerilerine ek olarak futbol tarihi hakkındaki bilgileri de ölçüldüğünden, maddelerin çok boyutlu olduğu ifade edilebilmektedir. Öğrencilerin okuduklarını anlama becerileri sabit tutulduğunda erkeklerin futbol tarihi hakkında daha bilgili olabilmeleri nedeniyle aynı beceri düzeyindeki kadınlara kıyasla bu metin ile ilgili maddeleri doğru yanıtlama olasılıklarının daha yüksek olabileceği düşünülmektedir. Bu durumda test ile ölçülen kuramsal yapıyla ilişkisiz olan dışsal bir boyuta bağlı olarak DMF oluşabilmektedir (Walker, 2011).

Taylor ve Lee (2012) ile Walker (2011) tarafından belirtildiği üzere ölçülmek istenen yapıyla ilişkili olan bir boyutun (örn., bazı bilişsel faktörler) var olması da maddelerin DMF göstermesine neden olabilmektedir. Örneğin öğrencilerin yanıtlarını açıklamaları ve/veya gerekçelendirmeleri gereken açık uçlu bir matematik maddesinde, sınava girenlerin matematik becerileri eşit olsa dahi, bilgileri organize etme ve sunma, düşünce ve fikirlerini yazılı olarak ifade ederek iletişim kurma becerileri iyi olan öğrencilerin bu maddelere doğru yanıt verme olasılıklarının daha yüksek olabileceği varsayılmaktadır. Uygulanan test ile öğrencilerin matematik becerilerine ek olarak yazılı iletişim kurma becerileri de ölçüldüğünden, bu testteki maddelerin çok boyutlu olduğu belirtilebilmekte ve yazılı iletişim açısından yetkin olan ve olmayan gruplar arasında bir farklılığın söz konusu olup olmadığının incelenmesi gerekebilmektedir.

Ulusal Matematik Öğretmenleri Konseyi (2000) tarafından belirtildiği üzere yazılı olarak matematiksel iletişim kurma, matematik becerisinin önemli bir bileşenidir. Bu nedenle test ile ölçülmek istenen yapıyla ilişkili olan dışsal bir boyut oluştuğunda da DMF söz konusu olabilmektedir (Walker, 2011). Öte yandan, bireylerin yetenek dağılımları hem birinci hem de ikinci boyutta farklılık göstermiyorsa madde çok boyutlu olsa bile DMF tespit edilmeyebilmektedir (Ackerman, 1992; Walker, 2011; Walker ve Sahin, 2017). Böyle bir durumda testle ölçülen boyutlar için karşılaştırılan gruplardaki bireylerin yetenekleri aynı seviyede olmakta ve DMF belirlenmemektedir (Walker, 2011). Bu

açıklamalar ışığında maddelerin boyut sayısının DMF'yi etkileyebileceği yönünde genel bir görüş birliği olması nedeniyle (Roussos ve Stout, 1996; Shealy ve Stout, 1993; Walker, 2011) DMF analizlerinden önce veri setlerinde anlamlı ikinci boyutların var olup olmadığının belirlenmesi amacıyla boyutluluk incelemelerinin yapılması önemli görülmektedir (Ackerman, Gierl ve Walker, 2003).

Psikometri alanyazınındaki araştırmalarda belirtildiği üzere DMF, madde yanlılığı ve madde etkisi terimleri farklı anlamlara gelmektedir. Bu nedenle araştırmalarda yapılmış olan tanımlara dayalı olarak (Camilli, 2006; Holland ve Wainer, 1993; Taylor ve Lee, 2012; Zumbo, 2007; Zumbo ve Gelin, 2005) bu üç kavram hakkında detaylı açıklamaların yapılmasında yarar görülmektedir. Daha önce anlatıldığı gibi testlerde DMF'li madde olup olmadığına istatistiksel yollarla karar verilmekte ve bu işlem yanlılık çalışmalarında standart bir uygulama olarak kabul edilmektedir. Yani bir maddenin yanlı olduğunu söyleyebilmek için ön koşul olarak o maddenin istatistiksel yollarla DMF'li olarak tespit edilmesi gerekmektedir. Öte yandan bir maddenin yapılan analizler sonucunda DMF'li olması, o maddenin yanlı olduğunu iddia etmek için ilk şart olmasına rağmen tek başına yeterli değildir (Ackerman, 1992; Goodman vd., 2011; Martinková vd., 2017; Wen, 2014; Zumbo, 1999, 2007; Zumbo ve Gelin, 2005). Bu nedenle DMF'nin tespit edilmesinden sonraki süreçte, madde veya test özelliklerinin DMF'ye neden olup olmadığının uzmanlar tarafından incelenmesi gerekmektedir (Taylor ve Lee, 2012).

Zumbo (1999) tarafından yapılan araştırmada DMF, madde yanlılığı ve madde etkisi terimleri arasındaki ilişkiler ortaya konularak analiz sürecinden sonra tespit edilen DMF'nin, madde etkisinden mi yoksa yanlılıktan mı kaynaklandığı açıklanmaya çalışılmaktadır (Taylor ve Lee, 2012; Zhu ve Aryadost, 2020; Zumbo ve Gelin, 2005). Madde yanlılığı ve madde etkisi genel olarak, DMF'nin nedeninin testle ilgili veya ilgisiz olmasına göre birbirlerinden ayrılmaktadır. Bir maddenin DMF'li olduğu belirlendikten sonra, testin ölçmeyi amaçladığı yetenekle ilgili olmayan bazı özellikleri nedeniyle, çeşitli gruplar için farklı çalışması maddenin yanlı olduğu anlamına gelmektedir (Camilli, 1992; Wen, 2014; Zumbo, 2007; Zumbo ve Gelin, 2005). Örneğin okuma becerisini ölçen bir testin, anadili testte kullanılan dil olan ve bu dili yeni öğrenen göçmen çocuklara uygulandığı düşünüldüğünde, aynı yetenek düzeyindeki çocukların uzun metin içeren maddelerdeki performanslarının farklılaşması madde yanlılığı şüphesini doğurmaktadır. Zhu ve Aryadost (2020) tarafından belirtildiği üzere DMF sonucunun yanlılık olarak değerlendirilmesi için iki duruma özellikle dikkat edilmelidir: (1) Madde içeriğinde

ölçülmek istenen yapı ile ilgisiz olan bir durum DMF oluşmasına neden olmalıdır, (2) DMF’li madde belirli bir grubu diğerlerine göre sistematik olarak avantajlı hale getirmelidir. Örnek vermek gerekirse; matematik başarı testindeki bir madde eğer kadınların aşına olmadığı bir spor terimi içermesi nedeniyle aynı yetenek düzeyindeki öğrenciler arasından erkeklere avantaj sağlıyorsa, spor terimleri hakkında bilgi sahibi olmak testin amacı ile ilgili olmadığından, maddenin yanlı olduğu ifade edilebilmektedir (Zumbo ve Gelin, 2005).

Madde yanlılığına ek olarak, DMF’nin var olduğu durumda çeşitli gruplardaki bireylerin bir maddeye doğru yanıt verme olasılığının farklılaşmasının nedeni, madde tarafından ölçülen özellikteki gerçek farklılık olduğunda madde etkisi söz konusu olmaktadır (Zumbo, 1999, 2007). Başka bir anlatımla farklı gruplardaki yanıtlayıcıların testteki bir maddeyi doğru veya yanlış yanıtlama olasılıklarındaki gerçek farklılık, madde etkisine işaret etmektedir. Bu gerçek fark belirli bir yetenek düzeyinde var olan farklılıklardır ve bazı grupların sahip olduğu bilgi, beceri veya tecrübesi sonucunda oluşabilmektedir. Örneğin eğitim öğretim fırsatlarının eşit olmadığı gruplar için bir maddenin DMF’li olması, genellikle madde yanlılığından ziyade madde etkisinden kaynaklı olabilmektedir. Dolayısıyla ölçmeye konu olan özellik nedeniyle testin uygulandığı çeşitli grupların performansları arasında farklılık oluşması madde etkisinin varlığına işaret etmektedir (Camilli, 2006; Clauser ve Mazor, 1998; Gök vd., 2010). Bununla birlikte matematik becerileri açısından eşleştirilmiş olan kadın ve erkeklerin performansları karşılaştırıldığında, öğrencilerin matematik becerilerinin ölçüldüğü bir testteki bazı maddeleri kadınların doğru yanıtlama olasılıklarının erkeklere kıyasla düşük olmasına rağmen bu farklılığın nedeni çoğu zaman açıklanamamaktadır. Böylesi durumlarda DMF’li maddeleri inceleyen uzmanlar maddelerin yanlı olabilmesi için genellikle bir neden gösteremeyip, DMF’nin madde etkisinden kaynaklı olabileceğini ifade etmektedirler. Böylece ya erkek ve kadınların matematik becerilerinin farklı değişkenlerden etkilendiği ya da gerçekten bu iki grubun matematik becerilerinin farklı olduğu (madde etkisi) sonucuna ulaşılmaktadır (McNamara ve Roevers, 2006; Zumbo ve Gelin, 2005; Zumbo vd., 2015).

Öğrencilerin performansları arasındaki gerçek farklılıklar belirli bir alandaki geçmiş yaşantı, tecrübe ve kültürlerinden kaynaklı da oluşabilmektedir. Örneğin uygulanan bir testte dayalı olarak düşük performans gösteren öğrenciler genel matematik derslerine, yüksek performans gösteren öğrenciler ise ileri matematik derslerine katıldığında, farklı öğrenme deneyimleri oluşmakta ve dolayısıyla farklı gruplardaki

öğrencilerin bilgi ve becerilerinde gerçek farklılıklar görülebilmektedir (Ong, 2012; Taylor ve Lee, 2012). Böylelikle bazı öğrencilerin sahip oldukları bilgi, beceri, tecrübe ve kültürel özelliklerin gruplar arasında gerçek farklılıkların oluşmasına neden olabileceği ifade edilebilmektedir.

Yapılan açıklamalardan anlaşıldığı üzere DMF'nin tespit edilmesi, test sonuçlarına ilişkin adaletin sağlanması için gerekli ancak yeterli olmayan bir ön koşuldur (Ackerman, 1992). Bu sebeple DMF belirlendikten sonraki adım, DMF'nin gerçekleşme nedenlerini araştırmak olmalıdır (Scott, 2014; Zhu ve Aryadost, 2020). Yapılan çalışmalarda belirtildiği üzere madde formatı (Scott, 2014; Zenisky, Hambleton ve Robin, 2003), madde içeriği/özelliği (Mendes-Barnett ve Ercikan, 2006; Sandilands vd., 2013; Scott, 2014), çeviri problemleri (Kristjansson, Desrochers ve Zumbo, 2003; Stubbe, 2011), kültürel farklılıklar (Huang vd., 2016; Li, Cohen ve Ibarra, 2004) ve testin hız testi olup olmadığı (Scott, 2014) gibi nedenler maddelerin farklı gruplardaki bireyler için DMF göstermesine yol açabilmektedir.

DMF'li olarak belirlenen maddelerin herhangi bir grup lehine gerçekten yanlı olup olmadığına karar verme sürecinde uzmanlardan görüş alınarak, içerik analizi veya ampirik incelemeler yapılması gerekmektedir (Wu, Tam ve Jen, 2016; Zumbo, 1999). DMF gösteren maddelerin ilgili alandaki uzmanlar tarafından incelenmesi sonucunda öncelikle yanlı olup olmadıkları belirlenmeli, sonrasında ise bu karara bağlı olarak maddelerin gelecekte kullanılıp kullanılamayacağına veya ölçme aracından çıkarılıp çıkarılmayacağına karar verilmelidir (Oliveri, Ercikan ve Zumbo, 2013; Zumbo ve Gelin, 2005). DMF'li maddeler belirlendikten sonra, bu maddeleri nihai testten çıkarmak veya gerekli durumlarda gözden geçirerek düzenlemek, geçerliğin artması açısından önemlidir (Magis vd., 2011; Schauburger ve Tutz, 2016; Walker, 2011).

DMF'li maddeler tespit edildikten sonra ne yapılacağına ilişkin karar verme sürecinde genel olarak farklı yaklaşımların izlendiği belirtilmektedir (Wu vd., 2016). Bir yaklaşıma göre DMF'li maddeler doğrudan testten çıkarılmalıdır. İkinci yaklaşıma göre ise eğer testteki DMF'li madde sayısı az, maddelerdeki DMF miktarı düşük ve maddelerin ölçmeyi amaçladığı özellik önemli ise bu maddelere ilişkin gerekli açıklamalar yapılarak ilgili maddelerin testte tutulmasına karar verilmelidir (Gamerman vd., 2018). Ancak Cho, Suh ve Lee'nin (2016) çalışmalarında da belirtildiği üzere DMF'li madde sayısının fazla ve DMF miktarının büyük olduğu durumda, DMF'nin göz ardı edilmesinin testin geçerliğini etkilemesi muhtemeldir. Öte yandan bütün DMF'li maddelerin testten çıkarılmasından sonra yapılan ikinci analizler neticesinde, önceden

DMF'li olmayan maddelerin de DMF göstermesinin mümkün olabileceğinin araştırmacılar tarafından hesaba katılması gerekmektedir (Bechger ve Maris, 2015; Wu vd., 2016). Bu çerçevede yapılan analizler sonucunda DMF'li olarak belirlenen bütün maddelerin testten çıkarılmasının hatalı olabilmesi nedeniyle (Gelin ve Zumbo, 2003), DMF'li madde sayısı ile bu maddelerdeki DMF miktarının da karar verme sürecinde dikkate alınması gerektiği belirtilebilmektedir (Kim ve Cohen, 1992).

Araştırmalarda belirtildiği üzere DMF belirleme çalışmalarının yalnızca madde yazarları veya test geliştiriciler tarafından değil, aynı zamanda test uyarlayanlar, uygulayıcılar, eğitimciler ve her disiplinden araştırmacılar tarafından da yapılması gerekmektedir (Guilera vd., 2013; Roussos ve Stout, 1996). Bu nedenle eğitim ve psikometri alanlarının (Hope vd., 2018; Li ve Zumbo, 2009) yanı sıra, psikoloji (Dodeen ve Johanson, 2003; Tan, Cai, Li, Zhang ve Tu, 2018), sağlık bilimleri (Gelin ve Zumbo, 2003; Panter ve Reeve, 2002) ve sosyal bilimler (Hong vd., 2018; Opariuc-Dan vd., 2017) gibi birçok alanda ölçek geliştirmek, uyarlamak ve ölçme aracının farklı bir örneklemdeki geçerliğini incelemek amacıyla yapılan madde analizleri sürecinde DMF çalışmalarının yürütülmesi gerektiği belirtilmektedir (Walker, 2011; Zumbo, 2007; Zumbo vd., 2015). Genellikle araştırmacıların istatistiksel yazılımlara erişememeleri, bu yazılımları kullanacak deneyime sahip olmamaları veya DMF belirleme teknikleri hakkında bilgi sahibi olmamaları gibi çeşitli sebeplerin DMF çalışmalarının göz ardı edilmesine neden olduğu ifade edilmektedir (Li ve Zumbo, 2009; Zumbo, 2007). Buna rağmen yine de bazı test geliştirme veya uyarlama çalışmalarında, maddelerin DMF içerip içermediğinin incelendiği görülebilmektedir (Croom ve Svetina, 2021; Demir, 2018; Hong vd., 2018; Paz, Spritzer, Reise ve Hays, 2017; Rimehaug, Kaat, Nordvik, Klokkeud ve Robinson, 2022; Tan vd., 2018; Zhong, Gelaye, Fann, Sanchez ve Williams, 2014).

Başarı ve yetenek testlerinde veya seçme ve yerleştirme amacıyla yapılan yüksek riskli sınavlarda adalet ve eşitliğin sağlanmasının önemli olması nedeniyle, madde yanlılığı ile ilgili endişeler daha da artmaktadır (Zumbo, 2007). Özellikle bazı kurumlar tarafından seçme ve yerleştirme amacıyla kullanılan testlerin sonuçlarına dayalı olarak öğrencilerin gelecekleri hakkında ciddi kararlar verilmesi nedeniyle testlerde yanlı madde bulunmasının, test puanlarına dayalı olarak yapılan değerlendirmelerin geçerliğini etkileyebileceği belirtilmektedir (Acar, 2008; Çepni, 2011; Kan vd., 2013; Toprak ve Yakar, 2017). Benzer durum standart testler için de geçerlidir. Kültürlerarası çalışmalar yürütmek amacıyla geliştirilen standart testler farklı dillere çevrilerek uyarlandıktan sonra bireylere uygulanmaktadır. Standart testlerin kullanıldığı eğitim ve psikoloji alanındaki

alışmalarda da len zellik bakımından eřit, ancak farklı lkelerde yer alan bireylerin sonularının aynı veya benzer olması gerekmektedir (Guilera vd., 2013). Katılımcıların puanlarının farklı kltrlerde karřılařtırılabilir olduėundan emin olmak iin ise, test geliřtirme veya uyarlama srecinde DMF belirleme alışmalarının yapılmasının nemli olduėu vurgulanmaktadır (Ellis ve Raju, 2003; Guilera vd., 2013; Revicki, Chen ve Tucker, 2015; Sireci, Patsula ve Hambleton, ev. 2017; Walker, 2011). Bu kapsamda Penfield'ın (2001) alışmasında bir lme aracındaki maddelerde yanlılık incelemesi yapmayı amalayan arařtırmacıların pilot, saha ve gerek uygulama olmak zere test geliřtirmenin tm ařamalarında DMF analizlerini yrtmesi gerektiėi belirtilmektedir.

PISA ve TIMSS gibi birok geniř lekli test uygulamasında da madde dzeyinde yapılan DMF incelemeleri, lme deėiřmezliėi alışmalarının rutin bir parası olarak kabul grmektedir (Jin vd., 2015; Svetina ve Rutkowski, 2014). Nitekim farklı kltr veya anadile sahip olan eřitli lkeler iin; madde evirilerinin zayıf olması, kavramların karřılıklarının farklı olması, madde formatının ėrencilere ařına gelmemesi veya madde ieriėinin uygun olmaması gibi nedenlerden dolayı, uluslararası geniř lekli uygulamalara farklı lkelerden katılan aynı yetenek dzeyindeki ėrencilerin performansları deėiřebilmektedir (Arım ve Ercikan, 2014; Ellis ve Mead, 2000; Ercikan ve Koh, 2005; Gırl ve Khaliq, 2001; Muñiz, Hambleton ve Xing, 2001; Walker, 2011; Zenisky vd., 2003). Bu anlamda uluslararası uygulamalara katılan lkeler hakkında doėru kararlar verilebilmesi amaıyla elde edilen veriler zerinde DMF alışmalarının yapılması nemli grlmektedir (Cheng vd., 2013).

Alanyazındaki ilk yanlılık alışmaları test puanlarındaki farklılıkları anlamak iin yapılmıř olsa da, yaklaşık olarak son 50 yıldır yapılan DMF belirleme alışmalarında bireylerin madde dzeyindeki performans farklılıklarının incelenmesi amalandıėından (Walker, 2011), gerekleřtirilen DMF analizleri ile test dzeyinden ziyade madde veya madde takımları dzeyindeki performans farklılıklarının tespit edilmesine odaklanıldıėı grlmektedir (Finch, 2012; Martinkov vd., 2017; Mucherah, Finch ve Keaitse, 2012; Ong, Williams ve Lamprianou, 2011). Yapılan arařtırmalar incelendiėinde, ncelikle madde dzeyinde, sonrasında ise madde takımı dzeyinde analiz yapabilmek amaıyla DMF belirleme tekniklerinin geliřtirildiėi grlmektedir (Millsap ve Everson, 1993; Potenza ve Dorans, 1995; Roussos ve Stout, 1996).

Alanyazındaki ikinci nesil DMF belirleme alışmalarında istatistiksel tekniklerin geliřtirilmesine odaklanıldıėı (Zumbo, 2007) ve istatistiksel tekniklerin geliřtirilmesine 1970'li yıllarda bařlanmasına raėmen (Gelin, 2001; Millsap ve Everson, 1993; Penfield

ve Camilli, 2007) özellikle 1990'lı yıllardan sonra çok sayıda yeni tekniğin geliştirildiği belirtilmektedir (Goodman vd., 2011; Lee ve Geisinger, 2016). Böylece çok sayıda araştırmacı tarafından üzerinde çalışılan DMF belirleme teknikleri; ölçme kuramları, maddeleri puanlama şekli, DMF türü ve karşılaştırılan grup sayısı gibi özellikler açısından sürekli geliştirilerek ortak özelliklerine göre farklı şekillerde sınıflandırılmaktadır (Mapuranga vd., 2008).

Genel olarak DMF belirleme teknikleri; dikkate alınan kurama göre KTK'ya ve MTK'ya dayalı teknikler (Ellis ve Raju, 2003; Drabinová ve Martinková, 2017); gözlenen puan ve örtük (gizil) puana dayalı teknikler (Elosua ve Wells, 2013); parametrik ve parametrik olmayan teknikler (Magis vd., 2010; Potenza ve Dorans, 1995); maddeleri puanlama şekillerine göre ikili veya çoklu puanlanan maddeler için geliştirilen teknikler (Penfield ve Lam, 2000); DMF türüne göre Tek Biçimli (TB) veya Tek Biçimli Olmayan (TBO) DMF'yi belirleyebilen teknikler (Mellenbergh, 1982) ve grup sayısına göre ikili veya çoklu gruplar için geliştirilen teknikler (Magis vd., 2010) olmak üzere farklı şekillerde sınıflandırılmıştır (Gamerman vd., 2018). Bu araştırma kapsamında öncelikle KTK ve MTK'nın özellikleri açıklanarak her iki kurama dayalı geliştirilen DMF belirleme teknikleri sırasıyla tanıtılmış, sonrasında ise çoklu gruplar için kullanılabilen teknikler kapsamlı bir şekilde ele alınmıştır.

KTK'nın 1970'li yıllardan sonra bilimsel araştırmalarda yaygın kullanılmasına bağlı olarak ilk DMF belirleme çalışmaları bu kuram çerçevesinde yapılmaya başlanmış (Gelin, 2001) ve MTK'ya dayalı teknikler geliştirilinceye kadar, KTK'ya dayalı teknikler araştırmalarda sıklıkla kullanılmıştır (Mapuranga vd., 2008). KTK'ya dayalı tekniklerde öncelikle bireylerin toplam puanları hesaplanmakta ve toplam puanlar açısından eşleştirilen farklı alt gruplardaki bireylerin madde düzeyindeki performansları ile gözlenen puanları arasındaki ilişkiye odaklanılmaktadır (Dorans ve Holland, 1993; Elosua ve Wells, 2013; Magis vd., 2010). Büyük örneklemeler gerektirmemesi ve hesaplamaların nispeten basit olması nedeniyle, KTK'ya dayalı tekniklerin pratikteki uygulamalarda kullanılmasının göreceli olarak daha kolay olduğu belirtilmektedir (Jin vd., 2018; Narayanan ve Swaminathan, 1996).

Bireylerin maddelere verdikleri tepkileri açıklamak amacıyla geliştirilmiş bir model ailesini oluşturan MTK'da ise, bireylerin yetenek düzeyleri ile madde özellikleri arasındaki ilişki incelenmektedir (Smith, 2002). Bu kurama dayalı DMF belirleme teknikleri iki grubun madde parametrelerini karşılaştırma veya iki grubun madde karakteristik eğrileri arasındaki alanı hesaplama olmak üzere iki temel yaklaşıma

dayanmaktadır (Clauser ve Mazor, 1998; Kim ve Cohen, 1995; Kim, Cohen ve Park, 1995; Lord, 1980; Raju, 1988). Bu yaklaşımlara göre; grupların madde parametreleri birbirine eşit veya bu parametrelere dayalı olarak çizilen madde karakteristik eğrileri arasında kalan alan sıfıra eşit ise ilgili maddenin DMF göstermediği ifade edilmektedir. Aksi taktirde bir maddenin parametre kestirimlerinin gruplar arasında önemli ölçüde farklı olması, bu maddenin madde karakteristik eğrisinin farklılaşmasına ve dolayısıyla ilgili maddenin DMF göstermesine neden olmaktadır (Jin vd., 2018; Kim vd., 1995).

Yapılan açıklamalardan yola çıkarak her iki kuram karşılaştırıldığında; bireylerin eşleştirilme sürecinin farklılaştığı görülmektedir. KTK'ya dayalı teknikler bireyleri toplam puana dayalı olarak eşleştirirken, MTK'ya dayalı tekniklerde bireylerin maddelere verdikleri yanıtlardan yararlanılarak gizil yetenek kestirilmekte ve eşleşme değişkeni olarak gözlenen puanlardan ziyade gizil yeteneğe ilişkin kestirimler kullanılmaktadır (Clauser ve Mazor, 1998; Martinková vd., 2017). Bu çerçevede MTK'ya dayalı teknikler, DMF analizi sırasında madde ve yetenek parametrelerini kestirmek için bir MTK modelinin kullanılmasını gerektirdiğinden, parametrik teknikler olarak kabul edilmektedir. KTK'ya dayalı tekniklerde ise DMF analizleri sırasında madde ve test düzeyinde ham puanların kullanılması gerektiğinden, parametrik olmayan teknikler olarak dikkate alındıkları belirtilmektedir (Desjardins ve Bulut, 2018).

Hem KTK hem de MTK'ya dayalı geliştirilen teknikler kullanılarak gerçekleştirilen analizler, test performansının altında yatan tek boyutlu bir gizil özelliğin olduğu varsayımına dayanmaktadır (Cohen ve Kim, 1993; Swaminathan ve Rogers, 1990). DMF için kullanışlı bir çerçeve oluşturan MTK'ya dayalı modeller, tek boyutluluk varsayımının yanı sıra yerel bağımsızlık ve parametre değişmezliği gibi varsayımların karşılanmasını ve model-veri uyumunun sağlanmasını da gerektirmektedir. Bu anlamda yapılan çalışmalarda daha fazla varsayım gerektirmesi nedeniyle KTK'ya dayalı tekniklere kıyasla MTK'ya dayalı DMF belirleme tekniklerinin kullanılmasının göreceli olarak zor olduğu ifade edilebilmektedir (Holland ve Wainer, 1993). MTK modellerinde özellikle geniş örneklem büyüklüğüne ihtiyaç duyulması ve belirtilen modellere aşına olmayan araştırmacıların MTK'ya dayalı teknikleri kullanmalarının daha zor olması gibi nedenler, pratik uygulamalarda bu tekniklerin kullanılmasını sınırlandırmaktadır (Clauser ve Mazor, 1998; Jin vd., 2018; Osterlind, 1983). Model varsayımları karşılandığı sürece MTK'ya dayalı tekniklerin daha fazla güce sahip olma eğiliminde oldukları açıklanmaktadır (Sireci ve Rios, 2013). Başka bir ifadeyle, eşleştirme değişkeni olarak kestirilen yetenek parametrelerinin kullanılması nedeniyle MTK'ya dayalı tekniklerin

diğerlerine kıyasla daha doğru sonuçlar ortaya koydukları ifade edilmektedir (Martinková vd., 2017). Bu bağlamda KTK ve MTK'ya dayalı tekniklerin kuramsal özellikleri ve gerektirdikleri varsayımlar farklılık göstermesine rağmen, her iki kurama dayalı teknikler kullanılarak DMF tespit edilmesinin gerekli olduğu belirtilmektedir (Mucherah vd., 2012).

Alanyazında açıklandığı üzere DMF belirlemek amacıyla yapılan ilk araştırmalarda ikili gruplar ve ikili puanlanan maddelere odaklanılmış (Penfield ve Lam, 2000) ve öncelikle madde güçlük indeksine, kontinjensi tablolarına ve varyans kaynaklarını belirlemeye dayalı tekniklerden yararlanılmıştır (Camilli ve Shepard, 1994). Bilimsel gelişmelerle birlikte yine ikili grup ve ikili puanlanan maddelerde kullanılmak üzere KTK'ya dayalı, Delta (Angoff ve Ford, 1973); Standartlaştırma (Dorans ve Kulick, 1986); Mantel-Haenszel (MH; Holland ve Thayer, 1988); Lojistik Regresyon (LR; Swaminathan ve Rogers, 1990); Lojistik Regresyon Lasso (LRL; Magis, Tuerlinckx ve De Boeck, 2015); SIBTEST (Simultaneous Item Bias Test; Shealy ve Stout, 1993; Roussos ve Stout, 1996); Diskriminant Lojistik Regresyon (DLR; French ve Miller, 1996) ve yapısal eşitlik modeli yaklaşımına dayalı MIMIC model (Multiple Indicators Multiple Causes; Finch, 2005) teknikleri alanyazına kazandırılmıştır. Benzer şekilde ikili grup ve ikili puanlanan maddelerde DMF belirlemek amacıyla MTK'ya dayalı, Lord ki-kare (LK; Lord, 1980); MTK Olabilirlik Oran (MTKOO) (Thissen, Steinberg ve Wainer, 1993); Raju'nun İşaretli ve İşaretsiz Alan İndeksleri testi (Raju, 1988, 1990); madde ile testin farklı fonksiyonlaşması (DFIT; Raju vd., 1995); güçlük parametresinin farklılaşması testi (Lord, 1980) ve Lordif (Logistic ordinal regression differential item functioning using IRT; Choi vd., 2011) geliştirilmiştir. Görüldüğü üzere DMF tespit etmek amacıyla çok sayıda istatistiksel teknik var olmasına rağmen, Narayanan ve Swaminathan (1996) tarafından yapılan çalışmada MH ve LR gibi tekniklerin DMF belirlemek amacıyla da kullanılabilen genel istatistiksel teknikler oldukları belirtilmiştir.

Araştırmalarda ikili puanlanan maddeler için geliştirilen bazı tekniklerin, geliştirilerek çoklu puanlanan maddeler için de kullanılabilmesinin mümkün olduğu belirtilmiştir (Oshima vd., 2015; Zumbo, 1999). Çoklu puanlanan maddelerden eğitim alanında duyuşsal özellikleri belirlemek amacıyla yararlanıldığından, araştırmalarda bu tür maddeler için kullanılabilen DMF belirleme tekniklerinin de geliştirilmesi gerektiği açıklanmıştır (Elosua ve Wells, 2013). Genelleştirilmiş MH (Zwick, Donoghue ve Grima, 1993), Sıralı Lojistik Regresyon (Wang ve Lane, 1996; Zumbo, 1999), SIBTEST (Chang, Mazzeo ve Roussos, 1996), Lordif (Choi vd., 2011) ve MIMIC model (Wang ve Shih,

2010) gibi teknikler kuramsal olarak çoklu puanlanan maddeler için geliştirilen tekniklerden bazılarıdır. Hem ikili hem de çoklu puanlanan maddelerde KTK ve MTK'ya dayalı DMF belirleme tekniklerinin özelliklerine ilişkin detaylı tartışmalar Clauser ve Mazor'un (1998), Millsap ve Everson'ın (1993) ve Penfield ve Camilli'nin (2007) çalışmalarında yer aldığından, kuramsal açıklamalar için bu araştırmaların incelenmesi tavsiye edilebilmektedir.

Alanyazındaki DMF araştırmalarının çoğunda, genellikle ikili puanlanan maddelere odaklanıldığı belirtilmektedir (Kim, Cohen, Alagoz ve Kim, 2007). İkili puanlanan maddeler için madde parametrelerinin gruplar arasında farklılaşması DMF'ye neden olmakta ve parametreler arasındaki farkın miktarının veya yönünün gruplar arasında değişmesi ise farklı DMF türlerinin oluşmasına yol açmaktadır (Li ve Stout, 1996; Walker, 2011). İlk olarak Mellenbergh (1982) tarafından DMF'nin TB ve TBO olmak üzere iki türü olduğu açıklanmakta ve birçok araştırmada belirtilen bu DMF türleri dikkate alınmaktadır (Clauser ve Mazor, 1998; Swaminathan ve Rogers, 1990). DMF analizlerinden elde edilen sonuçların doğru bir şekilde yorumlanabilmesi için DMF türleri arasındaki farkın iyi bilinmesi kritik bir önem taşımaktadır (Penfield ve Camilli, 2007).

MTK ile ilgili yapılan araştırmalarda belirtilen farklı MTK modellerinin kuramsal özelliklerinden yola çıkarak, TB ve TBO DMF arasındaki fark ortaya konulabilmektedir (Camilli, 2006). MTK'da madde tepki fonksiyonları (item response function) ve madde parametreleri arasında doğrudan bir ilişki olduğu bilinmekte ve güçlük ile ayırt edicilik parametrelerinden en az birinin gruplar arasında değişiklik göstermesi durumunda madde tepki fonksiyonlarındaki farklılaşmanın DMF'nin varlığına işaret ettiği ifade edilmektedir (Gao, 2019; Jin vd., 2015; Lord, 1980; Steinberg ve Thissen, 2006).

MTK çerçevesinde yetenek, θ (teta) olarak tanımlanmakta ve genel olarak -3 ile +3 aralığında yer almaktadır (Swaminathan ve Rogers, 1990). Bireyin bir maddeye doğru yanıt verme olasılığı ise, madde parametrelerine ve yeteneğe (örtük özelliğe) bağlı monoton bir fonksiyondur (Walker, 2011). Bu araştırmada belirli bir yetenek düzeyindeki bireylerin bir maddeye doğru yanıt verme olasılıkları 3 Parametrelili Lojistik Model (3PLM) kapsamında ele alınmış ve olasılık fonksiyonu Eşitlik 1'de verilmiştir (Magis vd., 2010).

$$P(x_{ij} = 1 | \theta_i, a_j, b_j, c_j) = c_j + (1 - c_j) \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}} = c_j + (1 - c_j) \frac{\exp a_j(\theta_i - b_j)}{1 + \exp a_j(\theta_i - b_j)} \quad (1)$$

Eşitlik 1’de x_{ij} , i . bireyin j . maddedeki yanıtını; θ_i , i . bireyin yeteneğini ve b_j , a_j ve c_j , j . madde için sırasıyla güçlük, ayırt edicilik ve şans parametrelerini göstermektedir. Güçlük parametresi kuramsal olarak, $-\infty$ ile $+\infty$ aralığında değer alabilse de, gerçek uygulamalarda çok kolay ve çok zor madde istenilmediğinden genellikle -2 ile +2 aralığında yer almaktadır. 2 Parametrelili Lojistik Model’de (2PLM) Eşitlik 1’de verilen c_j sıfıra sabitlenerek, düşük yetenek düzeyindeki bireylerin ilgili maddeyi şansa doğru yanıtlamasının mümkün olmadığı kabul edilmektedir. 2PLM’de madde ayırt edicilik parametresi, madde tepki fonksiyonunun eğimini, yani MKE’nin dikliğini belirlemektedir. Rasch modelde ise tüm maddelerin ayırt edicilik parametrelerinin 1.00’a eşit olduğu kabul edilmektedir. Ayrıca eğer testteki maddelerin ayırt edicilik parametrelerinin ortalaması, bütün maddeler için ayırt edicilik değeri olarak kabul edilirse 1 Parametrelili Lojistik Model (1PLM) oluşmaktadır (DeMars, 2010; Gamerman vd., 2018; de Gruijter ve van der Kamp, 2008; Magis vd., 2010; Rollins III, 2018). Eşitlik 1’de verilen olasılık fonksiyonu MKE olarak isimlendirilen bir grafik şeklinde de gösterilebilmektedir. MKE’nin y ekseninde bir maddeyi doğru yanıtlama olasılığı ($P(\theta)$), x ekseninde ise yetenek düzeyi (θ) yer almaktadır (Walker, 2011). Genel olarak bireylerin bir maddeye doğru yanıt verme olasılığı, yalnızca yetenek parametresi ile madde parametrelerine bağlıdır ve testi alan bireylerin hangi grupta olduğuna bakılmaksızın bir maddeyi doğru yanıtlama olasılıklarının aynı olması gerekmektedir (Embretson ve Reise, 2000).

İki grubun madde parametreleri aynı olduğunda, iki madde tepki fonksiyonu arasında fark oluşmadığından, ilgilenilen madde DMF göstermemekte ve grupların MKE’leri çakışmaktadır. Madde parametrelerinin gruplar arasında farklılaşması durumunda ise grupların MKE’leri farklılaşmakta ve incelenen maddelerin iki grup için farklı özellik gösterdiği anlaşılmaktadır (Clark, 2010; Clauser ve Mazor, 1998; de Ayala, 2009; Fischer ve Karl, 2019; Hladká ve Martinková, 2020). Bu bağlamda DMF türü ile ilişkisi nedeniyle dikkate alınan MTK modelinin önemli olduğu belirtilmektedir (Martinková vd., 2017; Penfield ve Camilli, 2007; Steinberg ve Thissen, 2006).

Maddelerin ayırt edicilik parametrelerinin aynı olduğu Rasch model ve 1PLM’de yalnızca güçlük parametresi farklılaştığından, tüm yetenek düzeylerinde olasılıklar arasındaki farklar aynı kalacak şekilde, avantajlı gruba ait olan MKE dezavantajlı gruba ait MKE’nin üzerinde yer almakta ve böylece TB DMF oluşmaktadır (Clauser ve Mazor, 1998; Smith, 2002; Walker, Beretvas ve Ackerman, 2001). Bu durumda bir madde iki

grup için eşit derecede ayırt edici iken, aynı madde gruplardan birine daha zor ($b_r < b_o$ ve $a_r = a_o$) gelebilmekte ve gruplara ilişkin MKE'lerin paralel olduğu görülmektedir (Walker, 2011). Birçok araştırmada da (Desjardins ve Bulut, 2018; Finch, 2016; Finch ve French, 2007; Lee, Bulut ve Suh, 2016; Smith, 2002) belirtildiği üzere TB DMF içeren maddeler bütün yetenek düzeylerinde gruplardan birine daha kolay gelerek, bu grupta yer alan bireylere ciddi ölçüde avantaj sağlamaktadır. Aynı zamanda TB DMF'de ayırt edicilik parametresinin gruplar arasında farklılaşmaması nedeniyle, yetenek düzeyi ile bir gruba ait olma arasında etkileşim olmadığı ifade edilebilmektedir (Finch ve French, 2007, 2008; Magis vd., 2010; Penfield ve Lam, 2000; Smith, 2002; Swaminathan ve Rogers, 1990; Zumbo, 2007).

MTK'ya dayalı 2PLM ve 3PLM'de, madde güçlük parametresine ilaveten sırasıyla madde ayırt edicilik ve şans parametrelerinin de gruplar arasında değişebileceği belirtilmektedir (Clauser ve Mazor, 1998). Şans parametresinin gruplar arasında sabit olduğu 2PLM'de hem güçlük hem de ayırt edicilik parametreleri birlikte farklılaşabileceğinden TBO DMF oluşabilmektedir (de Ayala, 2009). Yalnızca madde ayırt edicilik parametrelerinin gruplar arasında farklılaşması durumunda MKE'leri kesişmektedir. Bu nedenle bazı araştırmalarda kesişen (crossing) DMF olarak da belirtilen bu DMF türünde, paralelliğin bozulması nedeniyle grupların MKE'lerinin kesişme noktasının sol tarafında bir grup, sağ tarafında ise diğer grup avantajlı konuma geçmektedir (de Gruijter ve van der Kamp, 2008; Finch ve French, 2007; Shealy ve Stout 1993; Swaminathan ve Rogers, 1990; Woods, 2008). Başka bir ifadeyle; incelenen maddenin MKE'lerin kesiştiği noktanın altındaki yetenek düzeylerinde gruplardan birinin lehine, bu noktanın üstündeki yetenek düzeylerinde ise diğer grubun lehine çalıştığı ve böylelikle DMF'nin yönünün değiştiği belirtilebilmektedir (Hambleton vd., 1991; Lee vd., 2016; Smith ve Reise, 1998).

Alanyazındaki çalışmalarda açıklandığı üzere TBO DMF'li bir madde için yetenek düzeyleri ile bir gruba ait olma arasında etkileşimden söz edilebilmektedir (Finch ve French, 2007, 2008; Magis vd., 2010; Penfield ve Lam, 2000; Smith, 2002; Swaminathan ve Rogers, 1990; Zumbo, 2007). Örnek vermek gerekirse; yetenek seviyesine göre θ ölçeğinin bir ucunda A grubunun, diğer ucunda ise B grubunun bir maddeyi doğru yanıtlama olasılığının göreceli olarak yüksek olması durumunda, maddenin doğru yanıtlanması açısından gruplar arasında tutarsızlık olduğundan, grup üyeliği ile bireylerin yetenek düzeyleri arasında etkileşim olduğu anlaşılmaktadır (Osterlind ve Everson, 2009). Böylelikle gruplardaki bireylerin bir maddeyi doğru

yanıtlanma olasılığı arasındaki fark tüm yetenek seviyelerinde aynı olmayıp, ilgilenilen madde bazı yetenek düzeylerinde bir gruptaki bireylere kolay başka gruptaki bireylere ise zor gelmektedir (Finch, 2016; Jodoin ve Gierl, 2001; Mazor, Clauser ve Hambleton, 1994). Bu bağlamda bir maddeye ilişkin DMF türünün, yetenek sürekliliği boyunca grupların göreceli performanslarına bağlı olarak değişebildiği belirtilebilmektedir (Penfield, 2003; Penfield ve Camilli, 2007).

Gerçek uygulamalarda TBO DMF'ye TB DMF'den daha düşük bir oranda rastlanıldığı açıklanmasına rağmen, testlerde her iki DMF türüne sahip maddelerin bulunabileceği ifade edilmektedir (Finch ve French, 2007; Geraipour ve Shahmirzadi, 2019). Dorans ve Holland (1993) tarafından da belirtildiği üzere TBO DMF'li bir maddenin genellikle uygun bir şekilde düzenlenmesinin zor olması nedeniyle, bu tür maddelerin testten çıkarılmasının daha doğru bir karar olacağı düşünülmektedir.

Alanyazında TB ve TBO DMF türü, hem gruplara ilişkin MKE'lerin şekline hem de madde parametrelerine göre tanımlanmasına rağmen (Camilli, 2006; Clauser ve Mazor, 1998; Güler ve Penfield, 2009; Hanson, 1998; Hladká ve Martinková, 2020; Hope vd., 2018; Li ve Stout, 1996; Penfield, 2003; Shealy ve Stout, 1993; Smith, 2002; Swaminathan ve Rogers, 1990) yapılmış olan açıklamalar incelendiğinde, araştırmacılar tarafından kesin kabul görmüş bir tanımlamanın var olmadığı anlaşılmaktadır. Çok sayıda araştırmacı tarafından yapılan açıklamalarda, farklı gruplara ilişkin MKE'lerin -3 ile +3 aralığındaki belirli bir yetenek düzeyinde kesişebileceği ya da kesişmenin bu aralığın dışında olabileceği belirtilmektedir. Bu anlamda belirtilen yetenek düzeyi aralığında MKE'lerin kesiştiği veya kesişmediği durumlarda da TBO DMF oluşabileceğinden, bu durumun kafa karıştırıcı olabileceği vurgulanmaktadır (Li ve Stout, 1996; Lopez Rivas, 2012; Swaminathan ve Rogers, 1990). Bu karışıklığı gidermek amacıyla Walker ve diğerleri (2001) tarafından farklı bir tanımlama yapılmış ve TB (uniform), sıralı (ordinal) TBO ve sıralı olmayan (disordinal) TBO olmak üzere üç farklı DMF türünün varlığından bahsedilmiştir. Bu araştırmacılar, TB DMF'de madde ayırt edicilik parametresinin gruplar için eşit olduğunu, madde güçlük parametresinin ise gruplar arasında farklılaştığını ($a_R = a_O$ ve $b_R \neq b_O$) ifade etmiştir. TB DMF'de maddelerin doğru yanıtlanma olasılığındaki farklılık bütün yetenek aralıkları için tutarlı bir şekilde aynı grubun lehine veya aleyhine olduğundan, iki grup için MKE'lerin -3 ile +3 yetenek aralığında birbirlerini kesmediği belirtilmiştir (Lai, Teresi ve Gershon, 2005; Penfield ve Camilli, 2007). Böylelikle doğru yanıtlanma olasılığındaki farklılık yetenek ölçeğinin uç noktalarında daha az olsa da tüm yetenek düzeylerinde ilgilenilen maddeyi gruplardan

birindeki bireylerin doğru yanıtlama olasılıklarının yüksek olması nedeniyle belirtilen gruba daha kolay geldiği açıklanmıştır (Martinková vd., 2017). TB DMF için EK 1’de örnek olarak verilen MKE’ler incelendiğinde grupların madde ayırt edicilik parametrelerinin eşit olduğu, güçlük parametrelerinin ise değiştiği görülebilmektedir. Ayrıca referans grubun güçlük parametresi ile odak grubun güçlük parametresi arasındaki farka, yani DMF miktarına, göre MKE’lerin farklılaştığı ifade edilebilmektedir.

Walker ve diğerleri (2001) tarafından yapılan bir diğer tanımlamaya göre madde güçlük parametresi gruplar arasında eşit iken madde ayırt edicilik parametresi farklılık gösterdiğinde ($a_R \neq a_O$ ve $b_R = b_O$) sıralı olmayan TBO DMF oluşmaktadır. Bu DMF türünde bir madde bazı yetenek düzeylerinde bir grup için daha zor iken, diğer yetenek seviyelerinde aynı grup için daha kolaydır. Böylece sıralı olmayan TBO DMF oluştuğunda iki grup arasındaki farklılık negatif veya pozitif olmakta ve DMF’nin yönü değiştiğinden iki grubun MKE’leri ilgili yetenek aralığında kesişmektedir. EK 2’de örnek olarak verilen MKE’ler incelendiğinde, grupların güçlük parametrelerinin eşit olduğu ayırt edicilik parametrelerinin ise değiştiği görülebilmektedir. Bununla birlikte referans grubun ayırt edicilik parametresi ile odak grubun ayırt edicilik parametresi arasındaki farka, yani DMF miktarına, göre MKE’lerin farklılaştığı belirtilebilmektedir. Son olarak aynı araştırmacılar hem ayırt edicilik hem de güçlük parametrelerinin gruplar arasında farklılaşması durumunu ($a_R \neq a_O$ ve $b_R \neq b_O$) ise sıralı TBO DMF olarak tanımlamışlardır. Sıralı TBO DMF’de ilgilenilen madde gruplardan birine daha zor geldiği için MKE’lerin konumu değişmekte, ancak gruplar arasındaki bu fark yetenek aralığının tamamında sabit kalmamaktadır. Böylece bu DMF türünde bir grup diğerinden daha avantajlı olabilirken, avantajın miktarı yetenek düzeyine bağlı olarak farklılaşmaktadır. Başka bir ifadeyle bu DMF türünde güçlük ve ayırt edicilik parametrelerinin birlikte değişmesi nedeniyle, grupların madde parametrelerine göre MKE’ler de farklılaşabilmektedir. Sıralı TBO DMF ile TB DMF karşılaştırıldığında; her ikisinde de ortak olarak güçlük parametresi gruplar arasında farklılaşmakta, ancak sıralı TBO DMF’de ayırt edicilik parametresinin de gruplar arasında farklılaşması nedeniyle MKE’lerin düşük veya yüksek yetenek düzeylerinde kesişmesi mümkün olabilmektedir (Camilli, 2006; Güler ve Penfield, 2009; Walker, 2011; Walker vd., 2001).

Alanyazındaki çok sayıda çalışmada genel olarak yalnızca ayırt edicilik parametresinden kaynaklı olan (sıralı olmayan TBO) DMF’nin dikkate alındığı belirtilmiştir (Carroll, 2015; Jin vd., 2015; Li ve Stout, 1996; Svetina ve Rutkowski, 2014; Woods, 2008). Öte yandan hem güçlük hem de ayırt edicilik parametresinden kaynaklı

DMF'nin (sıralı TBO DMF) de dikkate alınması gerektiği bazı araştırmacılar tarafından vurgulanmasına rağmen (Güler ve Penfield, 2009; Smith, 2002) sıralı TBO DMF türünün ilgili alanyazındaki sınırlı sayıda araştırmada incelendiği görülmüştür (Clark ve LaHuis, 2012; Dell-Ross, 2021; Elosua ve Wells, 2013; Hidalgo ve López-Pina, 2004; Lopez Rivas, 2012; Lopez Rivas vd., 2009; Oshima vd., 2015; Seybert ve Stark, 2012; Seybert, Stark ve Chernyshenko, 2014; Stark, Chernyshenko ve Drasgow, 2006; Tay vd., 2015; Tay, Vermunt ve Wang, 2013; Woods, 2009a). Smith (2002) yapmış olduğu çalışmasında sıralı TBO DMF'nin daha kapsamlı inceleme gerektiren bir DMF türü olduğunu belirtmiştir. Çünkü sıralı TBO DMF'de gruplara ilişkin MKE'lerin farklı şekillere sahip olabilmesi nedeniyle, bazı durumlarda bu DMF türünün yorumlanmasının zor olabileceği ifade edilmiştir (Crane, Gibbons, Jolley ve Belle, 2006; Lopez Rivas, Stark ve Chernyshenko, 2009; Martinková vd., 2017). EK 3'te grupların hem güçlük hem de ayırt edicilik parametrelerinin farklılaştığı (sıralı TBO DMF) durumlardaki MKE'leri incelendiğinde, MKE'lerin her zaman -3 ile +3 aralığında kesişmeyebileceği sonucuna ulaşılmıştır. Dolayısıyla sıralı TBO DMF için çizilen MKE'lerin bazılarının sıralı olmayan TBO DMF'ye, bazılarının ise TB DMF'ye benzediği görülmüştür. Sonuç olarak alanyazında ortaya konulan bilgiler bağlamında yapılan bu araştırma kapsamında da DMF'ye neden olan madde parametrelerine göre; yalnızca b parametresinden (TB), yalnızca a parametresinden [TBO(a)] ve hem a hem de b parametresinden [TBO(ab)] kaynaklı olmak üzere üç farklı DMF türü dikkate alınmıştır.

DMF belirlemek için çok sayıda farklı istatistiksel teknik geliştirilmiş olmasına rağmen (Millsap ve Everson, 1993) bu tekniklerden bazılarının özel olarak TB, bazılarının ise TBO DMF'yi tespit edebilmek amacıyla kullanılabildiği ifade edilmiştir (Mellenbergh, 1982). Bütün tekniklerin her DMF türüne duyarlı olmaması nedeniyle, TB DMF'yi belirlemek amacıyla geliştirilmiş tekniklerin, genel olarak TBO DMF'yi tespit etmekte yetersiz kalabileceği kabul edilmiştir. Nitekim bu kapsamda Penfield (2003) tarafından yapılan çalışmada birleştirilmiş karar kuralı adı verilen bir yaklaşım önerilmiştir. Daha doğru sonuçlara ulaşabilmek için her iki DMF türünü belirleyebilen tekniklerin birlikte kullanılmasının gerekliliğinin vurgulanması nedeniyle, yapılan araştırmalarda TB ve TBO DMF'yi tespit etmek amacıyla geliştirilen tekniklerin birlikte kullanılmasının önemli olduğu sonucuna ulaşılmıştır (Penfield ve Camilli, 2007; Penfield ve Lam, 2000; Swaminathan ve Rogers, 1990; Walker, 2011).

Geliştirilmiş olan bütün DMF belirleme tekniklerinin genel olarak birbirlerine göre avantajlı ve dezavantajlı özelliklerinin olduğu çeşitli araştırmacılar tarafından

vurgulanmasına rağmen (Martinková vd., 2017; Rogers ve Swaminathan, 2016) DMF belirleme çalışmalarında kullanılan istatistiksel tekniklerin özelliklerinin de bu anlamda önemli olması sebebiyle (Fidalgo vd., 2004a; Hambleton, 2006; Higaldo ve Lopez-Pina, 2004; Revicki vd., 2015; Wyse ve Mapuranga, 2009) çoklu grup DMF belirleme teknikleri arasından araştırma kapsamında dikkate alınan GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerini daha iyi tanımak amacıyla belirtilen tekniklerin kuramsal özelliklerine ilişkin detaylı açıklamalar yapılmıştır.

Lord Ki-Kare ve Genelleştirilmiş Lord Ki-Kare

Lord ki-kare tekniği genel olarak MTK'ya dayalı kestirilen madde parametrelerinin vektörlerini gruplar arasında karşılaştırmak amacıyla geliştirilmiştir (Kim vd., 1995). Lord (1980) tarafından önerilen bu teknikte iki grup için MTK modelleri ile kestirilen parametre değerlerinin, gruplar arasında farklı olup olmadığını belirlemek amacıyla hipotezlerin test edildiği belirtilmiştir. Bazı kaynaklarda (Carroll, 2015; Langer, 2008) Wald test veya Lord'un Wald ki-kare testi olarak da adlandırılan bu teknikte, her iki gruptaki madde parametrelerinin eşit olup olmadığına ilişkin yokluk hipotezi test edilmekte ve bu test istatistiğinde serbestlik derecesi kestirilen parametre sayısına eşit olan ki-kare dağılımı dikkate alınmaktadır. Lord'un ki-kare istatistiği test edilirken alt gruplar için madde parametreleri ile kovaryansları hesaplanıp, kestirilen parametreler önce ortak bir metriğe dönüştürülmektedir. Başka bir ifadeyle iki veya daha fazla grubun olduğu durumlarda odak grup/grupların kalibrasyonundan elde edilen kestirimlerin referans grubun metriğine dönüştürülmesi gerekmektedir. Böylelikle ölçeklenmiş parametre ve kovaryans değerleri kullanılarak Lord'un ki-kare istatistiği hesaplanabilmektedir. Sonrasında ise elde edilen değerler serbestlik derecesine göre kritik ki-kare tablo değeri ile kıyaslanarak DMF'nin varlığı incelenmektedir (Battauz, 2019; Camilli ve Shepard, 1994; Carroll, 2015; Finch, 2016; Gao, 2019; Kim vd., 1995; Langer, 2008; Magis vd., 2010; Ong, 2012).

Lord (1980) tarafından belirtildiği üzere Lord'un ki-kare tekniğinde 2PLM için iki serbestlik dereceli, 3PLM için ise üç serbestlik dereceli ki-kare dağılımı dikkate alınmaktadır. Bu nedenle Lord ki-kare tekniği için serbestlik derecesi parametre kestirimlerinin yapıldığı MTK modeline göre farklılık göstermektedir (Kim vd., 1995; Woods vd., 2013). Lord ki-kare tekniğinin, MTKOO tekniğine (Thissen vd., 1993) asimptotik olarak benzer olduğu ifade edilmesine rağmen, yapılan araştırmalarda bu iki

teknikğin farklı performans gösterdikleri belirtilmektedir (Kim, Cohen ve Kim, 1994; Langer, 2008). Lord ki-kare testinde; Eşitlik 2’de verilmiş olan Q_j istatistiği ve vektörler kullanılmaktadır.

$$\chi_j^2 = Q_j = (V_{jR} - V_{jF})'(\Sigma_{jR} - \Sigma_{jF})^{-1}(V_{jR} - V_{jF}) \quad (2)$$

$$V_{jR} = (a_{jR}, b_{jR}, c_{jR}) \text{ ve } V_{jF} = (a_{jF}, b_{jF}, c_{jF})$$

$$Q_j = \frac{(b_{jR} - b_{jF})^2}{\hat{\sigma}_{jR}^2 + \hat{\sigma}_{jF}^2}$$

Eşitlik 2’de j . madde için referans ve odak gruplarındaki ayırt edicilik, güçlük ve şans parametrelerinden oluşan vektörler ile varyans-kovaryans matrisleri yer almaktadır. $V_{jR} = (a_{jR}, b_{jR}, c_{jR})$ ve $V_{jF} = (a_{jF}, b_{jF}, c_{jF})$ sırasıyla referans ve odak grup için j maddesinin ayırt edicilik, güçlük ve şans parametresine ilişkin kestirimlerin vektörleridir. Σ_{jR} ve Σ_{jF} referans ve odak gruplar için varyans-kovaryans matrisleridir. Eşitlik 2’de Q_j istatistiğinin 1PLM için basit formu verilmiştir. Bu eşitlikte $\hat{\sigma}_{jR}^2$ ve $\hat{\sigma}_{jF}^2$ sırasıyla referans ve odak gruptaki madde güçlük parametresinin kestirilen standart hatalarıdır. Asimptotik ki-kare dağılımına dayalı olan ve serbestlik derecesi modeldeki parametre sayısına karşılık gelen Q_j istatistiğinde, maximum likelihood (ML) yöntemi ile parametreler kestirilmektedir (Cao vd., 2017; Kim vd., 1995; Lord, 1980; Magis vd., 2010).

Battauz (2019) tarafından belirtildiği üzere iki grup için madde parametreleri ayrı ayrı kestirildiğinde farklı grupların parametre kestirimlerini karşılaştırmadan önce aynı metriğe dönüştürülmesi gerekmektedir. İkili grup için aynı parametre dönüştürme yaklaşımı (linking) kullanıldığında hem Q_j istatistiği hem de Lord ki-kare tekniği birbirlerine yakın sonuçlar ortaya koymaktadır. Q_j istatistiğinin manidar olması, çeşitli gruplardan kestirilen madde parametrelerinin farklılaşarak, dikkate alınan maddenin iki grupta farklı işlev gördüğünü ortaya koymaktadır (Kim vd., 1995). Langer’ın (2008) çalışmasında açıklandığı üzere Lord tarafından ortaya konulan ki-kare tekniği önce 1PLM, sonra ise 2PLM için geliştirilerek, 1PLM’de yalnızca güçlük, 2PLM’de ise hem güçlük hem de ayırt edicilik parametreleri birlikte dikkate alınmaktadır. Bununla birlikte Lord ki-kare tekniğinde gruplar arasındaki farkın yalnızca 1PLM’ye dayalı olarak incelenmesi durumunda, DMF’li maddeleri belirlemek amacıyla yapılan analiz sonuçlarının yetersiz olması nedeniyle Lord ki-kare tekniğinde 2PLM veya 3PLM’ye dayalı olarak parametre kestirilmesi gerektiği önerilmektedir. Sonuç olarak hem TB hem de TBO DMF’nin belirlenmesi için kullanılabilen Lord ki-kare’nin etkili bir teknik olduğu belirtilebilmektedir (Kim, 2010; Kim vd., 1995; Magis vd., 2010).

Yapılan çalışmalarda ikiden fazla grup için eş zamanlı olarak analiz yapmaya ihtiyaç duyulması nedeni ile Kim ve diğerleri (1995) tarafından Lord ki-kare tekniği ikiden fazla grupta kullanılmak üzere genelleştirilerek, GLK (Genelleştirilmiş Lord ki-kare) olarak adlandırılan yeni bir teknik alanyazına kazandırılmıştır. Bu teknik çoklu gruplarda eşzamanlı olarak analiz yapma olanağı sağlayan ilk teknik olma özelliğine sahiptir ve Q_j istatistiğine dayalıdır. Bu anlamda GLK tekniğinin üç veya daha fazla grup için madde parametre vektörlerini karşılaştıran Q_j istatistiği olarak isimlendirildiği görülmektedir (Kim vd., 1995; Woods vd., 2013).

Çoklu gruplar için analiz sürecinde iki gruplu teknikler kullanıldığında alt grupların ikili karşılaştırılmasının Tip I hatanın yükselmesine yol açması nedeniyle, bu sınırlılığın üstesinden gelmek için tek bir hipotezin test edildiği GLK tekniği önerilmektedir. Bu teknikte maddelerin parametre vektörleri gruplar arasında manidar bir farklılık gösterdiğinden, madde karakteristik eğrileri de gruplar arasında farklılaşmakta ve ilgili maddelerin DMF içerdiği anlaşılmaktadır. DMF gösteren bu maddelerin ise dikkate alınan gruplar için farklı çalıştığı ifade edilmektedir (Jin vd., 2015; Kim vd., 1995; Langer, 2008; Wang, 2016). GLK için yokluk hipotezi altındaki Q_j istatistiği Eşitlik 3'te, v_j vektörü ise Eşitlik 4'te verilmiştir.

$$Q_j = (Cv_j)'(C\sum_j C')^{-1} (Cv_j) \quad (3)$$

$$v_j = (a_1b_1, a_2b_2, \dots, a_gb_g) \quad (4)$$

Eşitlik 4'te a_g ve b_g sırasıyla g grubu için madde ayırt edicilik ve madde güçlükle parametreleridir. j . madde için referans grup ve odak gruplarından kestirilen madde parametre vektörlerinin birleştirilmesinden oluşan v_j vektörü ve aynı eşitlikte madde parametrelerinin varyans-kovaryans matrisinin köşegen bloğu olan $\sum_j$ matrisi yer almaktadır. Denklemden bulunan C matrisi ise gruplar arası karşılaştırma yapmak amacıyla kullanılan madde parametrelerini gösteren bir matristir. Bu bağlamda GLK, serbestlik derecesi $2C$ olan ki-kare dağılımının test edilmesine dayanmaktadır (Battaaz, 2019; Kim vd., 1995). GLK tekniğinde de Q_j istatistiği hesaplanmadan önce v_j vektöründeki bütün parametre kestirimlerinin tüm gruplar için ortak bir metriğe dönüştürülmesi gerekmekte ve dönüştürme yapıldıktan sonra $\sum_j$ tahmin edilmektedir (Magis vd., 2010; Woods vd., 2013). Gruplara göre madde parametrelerine ilişkin kovaryans matrisi Eşitlik 5'teki gibidir (Finch, 2016).

$$\Sigma_i = \begin{pmatrix} var(a_{i1}) & cov(a_{i1}, b_{i1}) & \dots & 0 & 0 \\ cov(a_{i1}, b_{i1}) & var(b_{i1}) & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & var(a_{ig}) & cov(a_{ig}, b_{ig}) \\ 0 & 0 & \dots & cov(a_{ig}, b_{ig}) & var(b_{ig}) \end{pmatrix} \quad (5)$$

Eşitlik 5’te verilen $var(a_g)$ g grubu için ayırt edicilik parametresinin varyansı, $var(b_g)$ g grubu için güçlük parametresinin varyansı, $cov(a_g, b_g)$ g grubu için güçlük ve ayırt edicilik parametrelerinin kovaryansıdır. Gruplar arasında uygun karşılaştırmalar yapılabilmesi için Lord ki-kare tekniğine benzer şekilde Q_j istatistiği hesaplanmadan önce bütün odak grup madde parametrelerinin referans grup temel alınarak ölçeklenmesi önemlidir. Çünkü gruplar arasında uygun karşılaştırmaların yapılabilmesi için referans grup temel alınarak tüm odak grup madde parametrelerinin kalibre edilmesi, kestirimlerin farklı gruplar için aynı ölçekte olmasını sağlamaktadır. Ayrıca yine Lord ki-kare tekniğinde olduğu gibi 1PLM, 2PLM ve 3PLM’ye dayalı olarak madde parametreleri kestirilebilmesi nedeniyle GLK tekniği ile de hem TB hem de TBO DMF belirlenebilmektedir (Finch, 2016; Magis vd., 2010; Magis vd., 2011).

GLK tekniğindeki Q_j istatistiği ile üç veya daha fazla grup için eş zamanlı olarak DMF belirlenmesine rağmen, Q_j istatistiğinin geniş örneklem büyüklüğü gerektirmesi gibi önemli bir gerekliliği bulunmaktadır. Çünkü MTK tabanlı bir teknik olan Q_j istatistiği madde parametre kestirimlerinin kararlılığı için büyük örneklem gerektirmektedir. Bu nedenle özellikle azınlık grupların, yani odak grupların küçük örnekleme sahip olması GLK’nın kullanımını sınırlandırmaktadır. Bu bilgiler doğrultusunda Q_j istatistiğinin bu sınırlılığı Lord ki-kare ve GLK tekniklerinin kullanılması esnasında dikkatli olunmasını gerektirmektedir (Camilli ve Shepard, 1994; Penfield, 2001). Ayrıca Q_j istatistiğinin manidar olması durumunda, herhangi iki grup arasında DMF olup olmadığının belirlenmesi amacıyla ikili gruplar için Lord ki-kare tekniğinin kullanılması ve düzeltilmiş alfa düzeyine göre post-hoc karşılaştırmalarının yapılması önerilmektedir (Kim vd., 1995). İkili gruplar arasındaki DMF’li maddeleri belirleme sürecinde grupları karşılaştırmak için C matrisi kullanılarak, hangi gruplar arasında DMF’nin mevcut olduğu tespit edilebilmektedir (Finch, 2016).

Alanyazında GLK tekniğinin geliştirilerek çoklu gruplarda DMF belirlemek amacıyla kullanıldığı ilk çalışmada, ilgili tekniğin farklı koşullardaki performansının belirlenmesi amacıyla çeşitli simülasyon çalışmalarının yürütülmesi gerektiği önerilmesine rağmen (Kim vd., 1995) bu tekniğin performansının incelendiği sınırlı

sayıda simülasyon çalışmasının yapıldığı belirtilmiştir (Woods vd., 2013). Kim ve diğerleri (1995) tarafından yapılan çalışmada MTK'ya dayalı GLK tekniğinin özellikle örneklem büyüklüğünden, test uzunluğundan ve DMF miktarından etkilenebilecek olması nedeniyle küçük örneklem, kısa test ve düşük DMF miktarı gibi faktörler dikkate alınarak çeşitli simülasyon çalışmalarının yapılmasının gerekli olduğu ifade edilmiştir. Benzer şekilde Langer'ın (2008) araştırmasında, GLK tekniğinin çeşitli faktörler altındaki performansının belirlenmesi amacıyla farklı simülasyon çalışmalarının yürütülmesi gerektiği belirtilmiştir. Böylece GLK tekniğinin farklı koşullar altındaki performansının ayrıntılı analizlere dayalı olarak incelenmesinin önemli olduğu vurgulanmıştır.

DMF ile ilgili çalışmalarda Lord ki-kare (LK) ve GLK tekniklerinin kullanıldığı araştırmalara rastlanılmasına rağmen, Langer (2008) tarafından yapılan çalışmada LK tekniğinin bazı sınırlılıklarının olduğu ileri sürülmüş ve Langer Genelleştirilmiş Wald (LGW) testi adında yeni bir teknik daha önerilmiştir. Bu teknik öncelikle iki grup için geliştirilmiş, sonrasında ise çoklu gruplara (LGÇGW) genelleştirilmiştir. İkili ve çoklu gruplarda kullanılabilen bu tekniğin hem ikili hem de çoklu puanlanan maddeler için de uygun olduğu ifade edilmiştir. LGW ve LGÇGW tekniklerinin hem TB hem de TBO DMF'yi tespit edebilmek amacıyla kullanılmasına rağmen, MTK'ya dayalı olması bu teknikler için de büyük örneklemelerin gerektiği belirtilmiştir (Carroll, 2015).

Mantel Haenszel ve Genelleştirilmiş Mantel Haenszel

Sağlık alanında çalışan araştırmacılar tarafından geliştirilen Mantel-Haenszel (MH) tekniği, ki-kare istatistiğinin tabakalara ayrılmış örneklemeler için kullanılabilmesi amacıyla 1950'li yıllarda yapılan araştırmalar sonucunda ortaya konulmuştur (Mantel ve Haenszel, 1959). Daha sonra Holland ve Thayer (1988) tarafından MH tekniğinin DMF belirlemek amacıyla kullanılabileceği belirtilmiştir. Düzeltilmiş ki-kare istatistiğine dayalı olan MH tekniğinin DMF belirlemek amacıyla geliştirilmiş popüler bir teknik olduğu çok sayıda araştırmacı tarafından ifade edilmiştir (Desjardins ve Bulut, 2018; Fidalgo, 2011; Holland ve Thayer, 1988). MH tekniğinde hesaplamaların kolay ve DMF belirlemek amacıyla geliştirilen birçok paket program yardımıyla gerçekleştirilebilmesi, bu tekniğin daha fazla tercih edilme nedeni olarak açıklanmıştır (Goodman vd., 2011; Tay vd., 2013). Ayrıca bazı araştırmalarda, MH tekniğinin birçok farklı koşulda diğer tekniklerden daha iyi performans gösterdiğinden bahsedilmiştir (Fidalgo vd., 2004a).

Alanyazında MH ki-kare olarak adlandırılan bu teknik parametrik olmayan teknikler arasında yer almaktadır. Bu teknikte referans ve odak gruptaki bireyler gözlemlenen bir yetenek puanına göre eşleştirilmekte ve bu eşleşmeye bağlı olarak hedef maddenin iki gruptaki doğru yanıtlanma olasılıkları karşılaştırılmaktadır (Finch ve French, 2008). Böylelikle eşleştirme kriterine göre belirlenen yetenek düzeyleri için grup üyeliği ile maddeye verilen yanıtlar arasında bir ilişki olup olmadığı incelenmektedir (Magis vd., 2010; Penfield, 2001). Eşleştirme kriteri olarak kullanılan toplam test puanı yetenek düzeylerine göre genellikle dört veya beş kategoriye (tabakaya) ayrılmakta ve böylece yetenek düzeyi açısından eşit veya benzer olan referans ve odak gruplar oluşturulmaktadır (Holland ve Wainer, 1993; Osterlind ve Everson, 2009).

Cochran (1954) istatistiğinin bir çeşidi olarak geliştirilen MH tekniğinde genel olarak $Q: R \times C$ olarak gösterilen kontinjensi tabloları oluşturulmaktadır. Bu tablolar eşleştirme kriteri ($h = 1, \dots, Q$), grup faktör düzeyi ($i = 1, \dots, R = 2$) ve yanıt düzeyi ($j = 1, \dots, C = 2$) olmak üzere üç farklı değişken içermektedir. MH tekniği ile DMF belirlemek amacıyla testteki maddeler için hipotezler ayrı ayrı test edilmektedir. Bir maddeyi doğru ve yanlış yanıtlayan birey sayılarından oluşan $2 \times 2 \times k$ (k =yetenek düzeyi) büyüklüğünde bir matris oluşturularak, her bir yetenek düzeyindeki grup üyeliği ile madde performansı arasındaki ilişki hesaplanmaktadır (Gamerman vd., 2018; Osterlind ve Everson, 2009; Penfield, 2001).

MH tekniği kullanılarak yapılan DMF belirleme çalışmalarında bireyler toplam test puanlarına göre gruplanmakta ve her bir yetenek düzeyi için ayrı ayrı olmak üzere grup üyeliği (referans ve odak) ile madde performansının (doğru-yanlış) etkileşimini içeren tablolar oluşturulmaktadır (Penfield, 2001). Örneğin J testteki madde sayısı, j testten alınan puan (sıfır ile J arasında), T_j sınava girenlerin sayısı olsun. Herhangi bir madde için maddeye verilen yanıtlara (doğru-yanlış) ve grup üyeliğine göre birey sayısı 2×2 kontinjensi tablosunda frekans olarak verilmektedir. Frekans sayılarından oluşan A_j , B_j , C_j ve D_j 'nin bu tablonun dört hücrelerini oluşturduğu düşünüldüğünde, A_j ve B_j hücreleri referans gruptaki, C_j ve D_j hücreleri ise odak gruptaki bireylerin bir maddeye verdikleri doğru ve yanlış yanıt sayılarını göstermektedir. Ayrıca frekansların toplanması sonucunda $N_{Rj} (A_j + B_j)$ ve $N_{Fj} (C_j + D_j)$ sırasıyla referans ve odak gruptaki birey sayılarını, $M_{1j} (A_j + C_j)$ ve $M_{0j} (B_j + D_j)$ ise sırasıyla maddeyi doğru ve yanlış yanıtlayan birey sayılarını belirtmektedir (Desjardins ve Bulut, 2018; Magis vd., 2010; Socha vd., 2015; Taylor ve Lee, 2012). Bu bilgilere dayalı olarak MH ki-kare istatistiği Eşitlik 6'da verildiği şekilde gösterilebilmektedir.

$$MH \chi^2 = \frac{(|\sum_j A_j - \sum_j E(A_j)| - 0.5)^2}{\sum_j Var(A_j)} \quad (6)$$

Eşitlik 6’da verilen $E(A_j)$ ve $Var(A_j)$ ’nin açılımları aşağıdaki Eşitlik 7’de verilmiştir.

$$E(A_j) = \frac{n_{Rj}m_{1j}}{T_j} \quad \text{ve} \quad Var(A_j) = \frac{n_{Rj}n_{Fj}m_{1j}m_{0j}}{T_j^2 (T_j - 1)} \quad (7)$$

Eşitlik 6’deki düzeltme katsayısı (-0.5) özellikle frekans sayısının az olduğu durumlarda gerekli olan ve ki-kare dağılımının yakınsamasını iyileştirmek için kullanılan bir sürekli düzeltme katsayısıdır. Eşitlik 7’deki değerler yerine yazılıp hesaplandıktan sonra bulunan sonuçlar Eşitlik 6’da verilen MH ki-kare formülünde kullanılmakta ve böylece her bir yetenek düzeyi için oluşturulan kontinjensi tablosuna dayalı olarak MH ki-kare istatistiği hesaplanmaktadır (Magis vd., 2010; Rollins III, 2018).

MH tekniğinde “toplam puana dayalı olarak oluşturulan her yetenek düzeyinde maddelere verilen yanıt ile grup üyeliği arasında ilişki yoktur” olarak belirtilen yokluk hipotezi test edilmektedir. Yani MH istatistiği maddeye verilen yanıt ile grup üyeliği arasında hiçbir koşullu ilişkinin olmadığı hipotezi altında, asimptotik olarak 1 serbestlik derecesine sahip bir ki-kare dağılımını kullanmaktadır. MH istatistik değeri ki-kare dağılımı asimptotik dağılıma dayalı olan kritik değerden daha büyükse, bu madde DMF’li olarak belirlenmektedir (Magis vd., 2010; Penfield, 2001; Rollins III, 2018).

MH tekniği kullanılarak yapılan DMF belirleme çalışmalarında birkaç varsayımın sağlanması gerekmektedir. Bu teknikte eşleştirme kriteri için kullanılan toplam puanın tek boyutlu ve eşleştirme puanının tüm seviyelerinde iki grup arasındaki DMF yönünün tek biçimli olduğu varsayılmaktadır. Bunlara ilaveten incelenen iki grup arasında kültür, sosyoekonomik statü gibi değişkenler açısından ortaklık olsa da bu grupların birbirinden bağımsız oldukları kabul edilmektedir. Son olarak ki-kare istatistiği için beklenen değerlerin hesaplanmasında, marjinal toplamların belirli bir toplam puanda sabit olduğu varsayılmaktadır (Rollins III, 2018).

Alanyazında MH tekniğinin referans ve odak olmak üzere iki grup için TB DMF’yi belirlemek amacıyla kullanılabilen en güçlü tekniklerden biri olduğu belirtilmesine rağmen (Finch, Immekus ve French, 2016; Holland ve Thayer, 1988), TBO DMF’yi tespit etmede etkisiz kaldığı ifade edilmiştir (Swaminathan ve Rogers, 1990). Böylece MH tekniği ile yapılan analizlerin istatistiksel olarak manidar olması, incelenen maddenin TB DMF gösterdiği anlamına gelmekte, ilgilenilen maddenin gerçekte TBO DMF göstermesi durumunda ise MH tekniğinin kullanılmasının uygun bir yaklaşım olmayabileceği bildirilmiştir (Desjardins ve Bulut, 2018). MH tekniğinin küçük

örneklemelerde kullanılabilmesinin yanı sıra hem hesaplanmasının hem de yorumlanmasının kolay olması nedeniyle son derece avantajlı olduğu açıklanmıştır (Desjardins ve Bulut, 2018; Holland ve Thayer, 1988; Swaminathan ve Rogers, 1990).

MH tekniğinin ikili puanlanan maddelerin yanı sıra çoklu puanlanan maddeler için de kullanılabilmesi amacıyla Mantel (1963) tarafından Mantel test ile Mantel ve Haenszel (1959) tarafından da MH tekniğinin geliştirilmiş bir versiyonu geliştirilmiştir (Finch vd., 2016). Genelleştirilmiş Mantel-Haenszel olarak isimlendirilen bu yeni tekniğin çoklu puanlanan kategorik veriler için, Mantel testin ise çoklu puanlanan sıralı veriler için kullanılabildiği çok sayıda araştırmada belirtilmiştir (Roussos, Schnipke ve Pashley, 1999; Wang ve Su, 2004; Zwick vd., 1993).

Alanyazında belirtildiği üzere MH tekniği ikili ve çoklu puanlanan maddelerde DMF belirlemek amacıyla geliştirilmiş olmasına rağmen referans ve odak olmak üzere yalnızca iki grubun olduğu durumda kullanılmak üzere sınırlandırılmıştır (Fidalgo ve Madeira, 2008; Fidalgo ve Scalón, 2010). Hem ikili hem de çoklu puanlanan maddeler için MH tekniği ile yalnızca iki grupta DMF belirlenebilmesi nedeniyle, ilk defa Landis ve diğerleri (1978) tarafından Mantel-Haenszel istatistiğinin geliştirilerek ikiden fazla grupta DMF belirlemek amacıyla kullanılması gerektiği önerilmiştir. Bu araştırma kapsamında R grup sayısı, C maddenin kategori sayısı olmak üzere $Q: R \times C$ kontinjensi tablosunun oluşturulabileceği açıklanarak, bu tekniğin hem çoklu puanlanan maddelerde hem de çoklu gruplarda kullanılabileceği ifade edilmiştir (Fidalgo, 2011; Fidalgo ve Madeira, 2008; Fidalgo ve Scalón, 2010). Zwick ve diğerleri (1993) çoklu puanlanan maddeler için, Penfield (2001) ise çoklu gruplar için geliştirilen DMF belirleme tekniklerinin geliştirilmiş teknikler olarak adlandırıldığını ifade etmişlerdir. Böylece ikili puanlanan maddeler ve iki grup arasında DMF belirlemek amacıyla geliştirilmiş olan MH tekniğinin, çoklu puanlanan maddeler ve çoklu gruplar için de kullanılabileceği belirtildikten sonra (Fidalgo ve Madeira, 2008; Finch ve French, 2019; Gamerman vd., 2018) ikiden fazla grup için eşzamanlı olarak analiz yapmak amacıyla da kullanılmaya başlanmıştır (Penfield, 2001; Somes, 1986).

Alanyazında çok sayıda araştırmada tercih edilen MH tekniğinin geliştirilerek çoklu gruplarda da kullanılabilmesi, araştırmacılar tarafından önemli bir gelişme olarak kabul edilmiştir (Langer, 2008). Landis ve diğerleri (1978) tarafından yapılan çalışmada GMH tekniği için $Q: R \times C$ olarak belirtilen kontinjensi tablosunun analiz edilebileceği açıklanmış ve matrislere dayalı Q_{GMH} formülü Eşitlik 8’de verilmiştir.

$$Q_{GMH} = \{\sum_{h=1}^Q (n_h - m_h)' A_h'\} \{\sum_{h=1}^Q A_h V_h A_h'\}^{-1} \{\sum_{h=1}^Q A_h (n_h - m_h)\} \quad (8)$$

Eşitlik 8’de verilen n_h , m_h , V_h ve A_h sırasıyla gözlenen frekansların vektörü, beklenen frekansların vektörü, kovaryans matrisi ve alternatif hipoteze göre tanımlanan doğrusal fonksiyonların bir matrisidir. Bu formül ile referans ve odak grupların her biri için maddeye verilen yanıtların rastgele dağıldığını kabul eden H_0 yokluk hipotezi test edilmektedir (Fidalgo ve Scalón, 2010).

Penfield (2001) tarafından belirtildiği gibi GMH tekniği MH ki-kare tekniğinin çoklu gruplarda da kullanılabilmesi amacıyla genelleştirilen yeni ve etkili bir versiyonudur. Bu teknikte bir maddenin ikiden fazla grup için DMF gösterip göstermediği tek bir analizle tespit edilebilmektedir (Carroll, 2015). İkili puanlanan maddelerde çoklu gruplar için geliştirilen GMH tekniğinde gruplardan biri referans grup olarak, diğer gruplar ise odak gruplar olarak tanımlanmaktadır. Referans ve odak grupların tanımlanması tamamen araştırmacının inisiyatifindedir. Örnekleme bir referans ve yalnızca bir odak grup olduğu durumda GMH ile MH tekniklerinin hemen hemen (-0.5 süreklilik düzeltmesi haricinde) aynı sonucu verdiği belirtilmektedir. GMH tekniğinde de MH tekniğine benzer şekilde her bir gruptaki bireylerin maddelere vermiş oldukları doğru ve yanlış cevap sayıları belirlenmekte ve eşleştirme kriteri olan toplam test puanlarına göre bireyler gruplandırılarak test istatistikleri hesaplanmaktadır. Ayrıca gruplar arasında madde yanıtlarının karşılaştırıldığı GMH tekniğinde de MH tekniğindeki gibi ki-kare dağılımı dikkate alınmaktadır (Finch, 2016; Finch ve French, 2019; Penfield, 2001; Rollins III, 2018). GMH tekniğini hesaplayabilmek için formül Eşitlik 9’da verilmiştir.

$$GMH = (n_j - \mu_j)' V^{-1} (n_j - \mu_j) \quad (9)$$

Eşitlik 9’da n_j ; referans grubundaki bireylerin toplam doğru cevaplarının gözlenen vektörü, μ_j ; DMF’nin olmadığı durumda referans grubundaki bireylerin toplam doğru cevaplarının beklenen sayısı, V ise n ’nin kovaryans matrisini göstermektedir. GMH tekniğinde J grup sayısı olmak üzere yokluk hipotezi $J-1$ serbestlik dereceli ki-kare dağılımı ile test edilmektedir. Sonucun istatistiksel olarak manidar olması yokluk hipotezinin reddedildiğine ve bu durum ise ilgili maddenin DMF’li olduğuna işaret etmektedir (Finch, 2016; Finch ve French, 2019; Penfield, 2001; Rollins III, 2018). GMH tekniğinde karşılaştırılan grup sayısına bakılmaksızın tek bir anlamlılık testi gerçekleştirildiğinden, ikili karşılaştırmalara dayalı yaklaşımlara göre performansının daha iyi olduğu ifade edilmektedir. Bu anlamda GMH tekniğinde Bonferroni veya

Benjamini-Hochberg gibi düzeltme yaklaşımlarının kullanılmasına gerek duyulmadan Tip I hata oranlarının iyi kontrol edilebildiği belirtilebilmektedir (Penfield, 2001).

Çoklu gruplar için MTK'ya dayalı olarak geliştirilen tekniklerin aksine (Kim vd., 1995) GMH tekniği hem MTK hem de KTK'ya dayalı olarak geliştirilen testlerden elde edilen veri setlerini analiz etmek amacıyla kullanılabilir (Fidalgo, 2011; Tay vd., 2013). Bu tekniğin MTK'ya dayalı olmaması nedeniyle geniş örneklem büyüklüğü gerektirmemesine bağlı olarak GLK tekniğine göre avantajlı olduğu belirtilmiştir (Fidalgo ve Scalón, 2010). Ayrıca GMH tekniğinin kullanıldığı araştırmalarda referans ve odak grupların yetenek dağılımlarının eşit olması durumunda Tip I hata oranının anlamlılık düzeyi olan .05 değerinden yüksek olmayacağı ortaya konulmuştur (Penfield, 2001). Öte yandan Penfield (2001) tarafından GMH tekniği için iki önemli husustan bahsedilmiştir. Bu hususlardan birincisi henüz GMH tekniği için etki büyüklüğü ölçüsü hakkında herhangi bir öneride bulunulmadığıdır. İkincisi ise çoklu gruplar için test edilen hipotezin anlamlı olması durumunda, DMF'nin hangi gruplar arasında olduğunu tespit etmek amacıyla ikili gruplar için MH tekniğinin kullanılması ve düzeltilmiş alfa düzeyine göre çoklu (post-hoc) karşılaştırmaların yapılmasıdır (Fidalgo ve Scalón, 2010; Finch, 2016; Finch ve French, 2019; Kim vd., 1995; Penfield, 2001).

Lojistik Regresyon ve Genelleştirilmiş Lojistik Regresyon

İstatistiksel bir teknik olan Lojistik Regresyon (LR) tekniğinin geliştirilmesi 1900'lü yılların başına denk gelmesine rağmen, bu teknik ilk defa Swaminathan ve Rogers (1990) tarafından DMF belirlemek amacıyla kullanılmıştır. Swaminathan ve Rogers (1990) çalışmalarında LR tekniğinin TB DMF'yi değerlendirmek amacıyla kullanılabileceğini göstermiş, Narayanan ve Swaminathan (1996) ise bu tekniğin TBO DMF'li maddeler için de uygulanabileceğini ortaya koymuştur. LR tekniği kullanılarak yapılan DMF belirleme çalışmalarında, ölçme aracının tek boyutlu olduğu varsayılmakta (Fischer ve Karl, 2019) ve bireyin yeteneğinin bir göstergesi olarak tanımlanan toplam test puanı, eşleştirme kriteri olarak ele alınmaktadır (Osterlind ve Everson, 2009). LR tekniği ile DMF belirlemek için toplam puan, grup üyeliği, toplam puan ile grup ilişkisinin etkileşimi ve bir maddeye doğru cevap verme olasılığı dikkate alınmaktadır (de Ayala, 2009). Böylece toplam test puanı, grup üyeliği ve bu ikisi arasındaki etkileşime dayalı olarak, incelenen maddeye doğru yanıt verme olasılığı için bir lojistik model kurulmaktadır (Magis vd., 2010; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016).

Araştırmalarda LR tekniğinin ikili puanlanan maddeler için iki grup arasında TB ve TBO DMF'yi belirlemek amacıyla sıklıkla kullanıldığı ifade edilmektedir. TB DMF incelendiğinde LR tekniğinin MH tekniği ile benzer performans sergilemesine rağmen (Finch ve French, 2008; Narayanan ve Swaminathan, 1996; Swaminathan ve Rogers, 1990) grupların ortalama yetenek düzeyleri farklılaştığında LR tekniğinin MTK'ya dayalı tekniklere göre yüksek Tip I hata oranı gösterdiği ifade edilmektedir (Güler ve Penfield, 2009). Hesaplama basitliği ve her iki DMF türünü belirleyebilme yeteneği nedeniyle alanyazındaki popüler tekniklerden biri olan LR için (Desjardins ve Bulut, 2018; Kim ve Oshima, 2013) Eşitlik 10'daki lojistik regresyon modeli dikkate alınmaktadır.

$$P(Y = 1|\pi) = \frac{e^{\pi}}{1+e^{\pi}} \quad (10)$$

Eşitlik 10'da verilen $P(Y = 1|\pi)$ DMF'li olup olmadığı araştırılan bir maddeye doğru yanıt verme olasılığı iken, π bir veya daha fazla yordayıcıya dayalı doğrusal bir denklemdir. LR tekniğinde temel model (Model 1) toplam test puanı olmak üzere (S) yalnızca tek bir yordayıcı içermektedir. Sonraki modelde (Model 2) temel modeldeki toplam test puanı değişkenine ilaveten TB DMF'yi test etmek için gerekli olan grup üyeliği (G) değişkeni dahil edilmektedir. Genellikle odak ($g = 1$) ve referans ($g = 0$) için farklı sayılar kullanılarak gruplar tanımlanmaktadır. Tam model olarak da belirtilen son modelde (Model 3) ise TBO DMF'yi incelemek için Model 2'deki değişkenlere ilaveten grup üyeliği ile toplam test puanının etkileşimi (SG) modele eklenmektedir. Model 1 ve Model 3 arasındaki anlamlı bir R^2 farkı, incelenen maddenin DMF gösterdiğini işaret etmekle birlikte, maddedeki DMF türünü tanımlamak için ek analizlerin gerekli olduğu belirtilmektedir. Bu bağlamda yapılan analizler sonucunda Model 1 ve Model 2 arasındaki sahte R^2 farkı anlamlı bulunduysa veya β_2 sıfırdan anlamlı bir şekilde farklıysa maddenin TB DMF gösterdiği belirtilebilmektedir. Benzer şekilde Model 2 ve Model 3 arasındaki sahte R^2 farkı anlamlı bulunduysa veya β_3 sıfırdan anlamlı bir şekilde farklıysa (β_2 'nin de anlamlı olup olmadığına bakılmaksızın) maddenin TBO DMF gösterdiği ifade edilebilmektedir (Desjardins ve Bulut, 2018; Jin vd., 2018; Tutz ve Berger, 2016). Tam model için oluşturulan lojistik regresyon modeli Eşitlik 11'de verilmiştir (Magis vd., 2010).

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 S_i + \beta_2 G_i + \beta_3 (SG)_i \quad (11)$$

Eşitlik 11'de π_i , i bireyinin bir maddeye doğru yanıt verme olasılığını göstermektedir. S_i toplam test puanı, G_i grup üyeliği (referans veya odak), $(SG)_i$ ise

etkileşimdir. Model parametreleri ise ($\beta_0, \beta_1, \beta_2, \beta_3$) istatistiksel testler (Wald test veya Olabilirlik Oran testi) kullanılarak test edilmektedir. $\beta_2 = \beta_3 = 0$ olduğu durumda DMF'nin olmadığı anlaşılmaktadır. TB DMF için β_2 katsayısına ($\beta_2 \neq 0; \beta_3 = 0$), TBO DMF için ise β_3 katsayısına ($\beta_2 = 0; \beta_3 \neq 0$) dayalı olarak yokluk hipotezi reddedilir. Dolayısıyla grubun ana etkisi dikkate alındığında TB DMF, grup ve toplam puanın etkileşimi dikkate alındığında ise TBO DMF'nin test edildiği belirtilebilmektedir. Bu bilgiler doğrultusunda referans ve odak grup arasında madde performansı açısından fark oluşmasının DMF'nin varlığına işaret etmesi nedeniyle hem β_2 hem de β_3 katsayılarının test edilmesi gerektiği ifade edilmektedir (Fischer ve Karl, 2019; Gamerman vd., 2018; Magis vd., 2010; Rollins III, 2018; Tutz ve Schauburger, 2015).

LR tekniğinde kurulan farklı modeller arasında log olabilirlik değerleri karşılaştırılarak TB ve TBO DMF test edilmektedir. Bu teknikte manidarlık ki-kare farkına dayalı olarak belirlendiğinden, örneklem büyüklüğünden etkilenmekte ve örneklem büyüklüğü arttıkça iki model için karşılaştırılan log olabilirlik değerlerindeki küçük farklar bile manidar sonuçlara yol açabilmektedir (Fischer ve Karl, 2019; Zumbo, 1999). Bu çerçevede Zumbo (1999) yaptığı araştırmasında Swaminathan ve Rogers (1990) tarafından önerilen analiz sürecine bir adım daha eklemiş ve istatistiksel anlamlılığın yanı sıra hesaplanan sahte R^2 etki büyüklüğü katsayılarından en az birinin yorumlanması gerektiğini belirtmiştir. Böylece istatistiksel olarak manidar bulunan test sonucunda DMF büyüklüğünü tespit etmek için McFadden's sahte R^2 , Nagelkerke ve Cox & Snell etki büyüklüklerinin yorumlanması gerektiği önerilmiş (Fischer ve Karl, 2019; Zumbo, 1999) ve McFadden's sahte R^2 etki büyüklüğünün dikkate alınmasının, Tip I hatayı azalttığı açıklanmıştır (Jodoin ve Gierl, 2001).

LR tekniğinin çoklu puanlanan maddelerde kullanılabilmesi amacıyla Sıralı (Ordinal) Lojistik Regresyon (SLR) tekniği geliştirilmiş ve bu tekniğin çoklu puanlanan maddeler için iki grup arasında DMF belirlemek amacıyla da kullanıldığı belirtilmiştir (Miller ve Spray, 1993). LR tekniğinin genelleştirilerek çoklu puanlanan maddelerde kullanılabilmesi, farklı madde türlerinden oluşan testlerin analiz sürecinde de bu tekniğin tercih edilebileceğini göstermiştir (Penfield ve Lam, 2000; Zumbo, 1999).

Alanyazındaki ilk DMF belirleme çalışmalarında LR tekniği kullanılarak referans ve odak grup olmak üzere iki grup karşılaştırmaları üzerine odaklanılmıştır. Ayrıca grup sayısı ikiden fazla olduğu durumlarda da ikili karşılaştırma veya grup birleştirme yaklaşımları dikkate alınarak iki grup için geliştirilen LR tekniğinin kullanılabilirdiği ifade edilmiştir (Magis vd., 2010). Bununla birlikte yapılan araştırmalarda grup sayısının

ikiden fazla olması durumunda, bu grupların ikiyeşerli karşılaştırılmasına dayalı yaklaşımların birtakım sınırlılıklarının olduğu belirtilmiştir. Belirtilen bu sınırlılıklardan yola çıkarak LR tekniğinin matematiksel formüllerinin geliştirilmesiyle birlikte ikiden fazla grup için de kullanılabileceği çok sayıda araştırmacı tarafından açıklanmıştır (Agresti, 2007; Kanjee, 2007; Miller ve Spray, 1993; Millsap ve Everson, 1993; Penfield, 2001). Yapılan öneriler doğrultusunda LR tekniği Magis ve diğerleri (2011) tarafından çoklu gruplarda kullanılmak üzere geliştirilmiş ve Genelleştirilmiş Lojistik Regresyon (GLR) olarak adlandırılan bu yeni teknik alanyazına kazandırılmıştır.

GLR tekniği iki grup arasındaki DMF'yi tespit etmek amacıyla önerilen LR'nin çoklu gruplar için de uygulanabilen bir genelleştirilmiş bir versiyonu olarak kabul görmektedir. GLR tekniğinde ikili karşılaştırmalara veya odak grupların birleştirilmesine gerek duyulmayıp bütün alt gruplar bağımsız olarak dikkate alınabilmektedir. Alt grupların ikili karşılaştırılmasının Tip I hatanın artmasına yol açması nedeniyle, tek bir hipotezin test edildiği GLR tekniği avantaj sağlamaktadır. Yani ikili (referans-odak) karşılaştırma yaklaşımına göre daha kullanışlı olan bu teknikte, örneklemdeki bütün alt grupları bağımsız olarak incelemek mümkün olmakta ve böylece daha doğru sonuçlara ulaşılabilmektedir. LR tekniğinde olduğu gibi GLR tekniğinde de hem TB hem de TBO DMF belirlenebilmektedir (Magis vd., 2011; Rollins III, 2018; Svetina ve Rutkowski, 2014). DMF belirleme sürecindeki karmaşıklık dikkate alındığında her iki DMF türünün de tespit edilebilmesi nedeniyle GLR tekniğinin birçok araştırmacı tarafından tercih edildiği ifade edilebilmektedir (Finch, 2016; Rollins III, 2018; Svetina ve Rutkowski, 2014). Çoklu grup tekniklerinden GLR için denklem Eşitlik 12'de verilmiştir.

$$\ln\left(\frac{\pi_{ig}}{1-\pi_{ig}}\right) = \text{logit}(\pi_{ig}) = \alpha + \beta S_i + \alpha_g + \beta_g S_i \quad (12)$$

Eşitlik 12'de π_{ig} , g grubundaki i bireyinin bir maddeye doğru yanıt verme olasılığını göstermektedir. α ve β , bütün gruplar için ortak kesişim (güçlük) ve eğim (ayırt edicilik) parametreleri iken α_g ve β_g , gruba özgü kesişim ve eğim parametreleridir. S_i , bireylerin yetenek düzeylerini gösteren bir eşleştirme kriteri olarak i bireyinin toplam test puanını göstermektedir. Referans grup için $g = 0$, odak gruplar için ise sırasıyla $g = 1, 2, \dots, F$ olarak tanımlandıktan sonra gruplar analize dahil edilmektedir. Tanımlama sorunlarını önlemek için gruba özgü parametrelerin sınırlandırılması gerekmektedir. Bu nedenle referans grubunun gruba özgü parametreleri olan α_0 ve β_0 , ilgili modeli tanımlamak amacıyla sıfıra eşitlenir. Dolayısıyla oluşan modelde $g = 0$ olduğunda (12)

eşitliğindeki formül $\text{logit}(\pi_{ig}) = \alpha + \beta S_i$ şekline dönüşür. Diğer taraftan $g \neq 0$ ise $g = 1, 2, \dots, F$ odak grup için (12) eşitliğindeki formül $\text{logit}(\pi_{ig}) = (\alpha + \alpha_g) + (\beta + \beta_g) S_i$ şeklinde gösterilebilir. Gruba özgü tüm parametreler sıfıra eşit olduğunda incelenen maddede DMF'nin olmadığı ifade edilebilmektedir. Bununla birlikte bir madde için π_{ig} olasılığı gruplar arasında değişirse DMF'nin varlığından söz edilebilmektedir. Yani alternatif hipotezde test edilen gruba özgü parametrelerden (α_g ve β_g) en az birinin sıfırdan farklı olması gerekmektedir. Bu durumda $\alpha_g \neq 0$ olduğunda bir veya daha fazla grupta TB DMF, $\beta_g \neq 0$ olduğunda ise bir veya daha fazla grupta TBO DMF tespit edildiği sonucuna ulaşılmaktadır (Finch, 2016; Magis vd., 2011; Svetina ve Rutkowski, 2014). Magis ve diğerleri (2011) tarafından yapılan çalışmada GLR tekniği için grupların parametrelerinin eş zamanlı olarak kestirildiği ifade edilmektedir. Ayrıca GLR tekniğinde grupların örneklemeleri ne kadar büyükse, parametre kestiriminin o kadar iyi olacağı ve dolayısıyla tekniğin DMF'li maddeleri doğru belirleme olasılığının artacağı açıklanmaktadır. Başka bir ifadeyle geniş örneklem büyüklüklerinde parametre kestirimleri iyileşmekte ve GLR'nin istatistiksel gücü artmaktadır. Bu nedenle GLR'nin gruplardaki katılımcı sayısının fazla olduğu geniş ölçekli testlerden elde edilen veriler üzerinde yapılan DMF belirleme çalışmalarında kullanılmasının daha uygun olduğu belirtilmektedir (Finch, 2016; Magis vd., 2011; Svetina ve Rutkowski, 2014).

GLR tekniğinde üç farklı yokluk hipotezi test edilmektedir. Magis ve diğerleri (2011) ile Rollins III (2018) tarafından yapılan çalışmada bir maddenin DMF'li olup olmadığına ve DMF türüne ilişkin yokluk hipotezleri sırasıyla Eşitlik 13, 14 ve 15'te verilmiştir.

$$H_0: \alpha_1 = \dots = \alpha_F = 0 = \beta_1 = \dots = \beta_F = 0 \quad (DMF) \quad (13)$$

$$H_0: \alpha_1 = \dots = \alpha_F = 0 | \beta_1 = \dots = \beta_F = 0 \quad (TB DMF) \quad (14)$$

$$H_0: \beta_1 = \dots = \beta_F = 0 \quad (TBO DMF) \quad (15)$$

Eşitlik 13, 14 ve 15'ten de görüldüğü üzere bir maddenin DMF'li olmaması her iki madde parametresinin gruplar arasında farklılaşmamasını gerektirmektedir. Bununla birlikte tüm eğim parametreleri sıfıra eşitken, en az bir kesişim parametresi sıfırdan manidar bir şekilde farklıysa TB DMF ve en az bir eğim parametresi sıfırdan manidar bir şekilde farklıysa TBO DMF oluşmaktadır. DMF istatistiksel olarak model parametrelerinin incelenmesi ile değerlendirilmektedir. τ model parametrelerinin bir vektörünü temsil etmektedir. TB DMF için $\tau = (\alpha, \alpha_1 \dots \alpha_F, \beta)^T$ ve TBO DMF için $\tau = (\alpha, \alpha_1 \dots \alpha_F, \beta, \beta_1 \dots \beta_F)^T$ olarak tanımlanmaktadır. GLR tekniğinde ML kestirim yöntemi

kullanılmakta ve yokluk hipotezini test edebilmek amacıyla Wald ki-kare ve Olabilirlik Oran testlerinden birini seçme olanağı sunulmaktadır (Finch, 2016; Jin vd., 2015; Rollins III, 2018; Svetina ve Rutkowski, 2014). Her iki testin de iç içe model karşılaştırmalarına dayalı olması ve aynı asimptotik ki-kare dağılımını kullanması bakımından benzer oldukları belirtilebilmektedir (Rollins III, 2018).

Bu araştırma kapsamında GLR-Wald olarak isimlendirilen Wald ki-kare testi, parametrelerin sıfırdan manidar bir şekilde farklı olup olmadığını incelemek amacıyla kullanılan bir parametre vektörünün anlamlılık testidir (Rollins III, 2018). Wald ki-karede test istatistiği sonucu, ki-kare dağılımının odak grup sayısına eşit olduğu serbestlik derecesindeki kritik puanı aştığında yokluk hipotezi reddedilmekte ve DMF'nin istatistiksel olarak var olduğu sonucuna ulaşılmaktadır. Diğer taraftan bu araştırma kapsamında GLR-LRT olarak isimlendirilen Olabilirlik Oran testinde ise, yokluk hipotezi ile alternatif hipotez için oluşturulan iki iç içe geçmiş model karşılaştırılmakta ve en uygun modelin belirlenmesi için Wilks tarafından ortaya konulan lamda istatistiğinden yararlanılmaktadır (Finch, 2016; Jin vd., 2015; Svetina ve Rutkowski, 2014). Olabilirlik Oran testinde sırasıyla yokluk ve alternatif hipotezlerin test edilmesi için kurulan modeller (M_0 ve M_1) (Svetina ve Rutkowski, 2014) Eşitlik 16 ve Eşitlik 17'de verilmiştir.

$$M_0 \equiv \text{logit}(\pi_{ig}) = \left\{ \begin{array}{c} \alpha + \beta S_i \\ \alpha + \beta S_i + \alpha_g \end{array} \right\} \quad (16)$$

$$M_1 \equiv \text{logit}(\pi_{ig}) = \left\{ \begin{array}{c} \alpha + \beta S_i + \alpha_g \\ \alpha + \beta S_i + \alpha_g + \beta_g S_i \end{array} \right\} \quad (17)$$

Eşitlik 16 ve 17'de verilen π_{ig} , g grubundaki i bireyinin ikili puanlanan bir maddeye doğru yanıt verme olasılığını göstermektedir. Referans grubu için $g = 0$ 'dır ve odak gruplar ise $g = 1, 2, 3, \dots, F$ olarak temsil edilmektedir. M_0 ve M_1 modellerinde parantezlerin üst kısmı TB DMF, alt kısmı ise TBO DMF test edilirken kullanılmaktadır. Kurulan bu modeller ile maddelerin DMF'li olup olmadığı lamda istatistiği ile test edilmektedir. Lamda istatistiğinde $\Delta = -2\log(\frac{L_0}{L_1})$, L_0 ve L_1 sırasıyla M_0 ve M_1 için en yüksek olabilirlik değerleridir ve boş model ve alternatif modele karşılık gelmektedir. Bu istatistik için yokluk hipotezi, M_0 ve M_1 modelleri arasındaki parametre sayısına eşit serbestlik derecesindeki asimptotik ki-kare dağılımına dayalı olarak test edilmektedir. Lamda istatistiğinin yeterince büyük olması durumunda incelenen maddelerin DMF'li olduğu sonucuna ulaşılmaktadır (Magis vd., 2011; Rollins III, 2018).

Hem GLR-LRT hem de GLR-Wald teknikleri aynı asimptotik dağılıma dayalıdır ve bu iki teknik de büyük örneklemelerde benzer sonuçlar ortaya koymaktadır (Magis vd., 2011). Ancak yine de bu tekniklerden biri çok sınırdaki p değerine sahipse, diğer tekniğin manidar bulunmama ihtimali göz önünde bulundurulmalıdır. Magis ve diğerleri (2011) tarafından belirtildiği üzere her iki tekniğin de avantajları bulunmaktadır. Yapılan bir araştırmada küçük örneklemelerde Olabilirlik Oran testine dayalı tekniğin (GLR-LRT), Wald ki-kare testine dayalı (GLR-Wald) tekniğe göre daha güvenilir olduğu ifade edilmesine rağmen (Agresti, 2007), DMF'nin hangi iki grup arasında olduğunu belirlemek için Wald ki-kare testinin kullanılmasının gerektiği belirtilmektedir. Ancak yine de genel olarak her iki testten elde edilen sonuçlar arasında kabul edilebilir bir benzerlik olması beklenmekte ve bu tekniklerin sonuçları arasında önemli farklılıkların bulunması, gruplardaki örneklemelerin yeterince büyük olmadığı şeklinde yorumlanabilmektedir. Bu bağlamda araştırmalarda bu iki tekniğin birlikte kullanılarak elde edilen sonuçların karşılaştırılması gerektiği önerilmektedir (Magis vd., 2011). Alanyazında LR tekniğinin kullanıldığı iki gruplu DMF belirleme çalışmalarında Olabilirlik Oran ve Wald ki-kare testlerine ilişkin sonuçların karşılaştırıldığı görülmektedir (Gao, 2019). Benzer şekilde GLR tekniğinin tercih edildiği çoklu grup DMF belirleme çalışmalarında ise yalnızca GLR-LRT (Rollins III, 2018; Svetina ve Rutkowski, 2014) ve her iki tekniğin (GLR-LRT, GLR-Wald) birlikte kullanıldığı araştırmalara da rastlamak mümkün olmaktadır (Magis vd., 2011; Teymoori vd., 2020).

Magis ve diğerleri (2011) yaptıkları çalışmalarında GLR tekniğindeki Olabilirlik Oran testinin Thissen ve diğerleri (1993) tarafından ortaya konulan MTK'ya dayalı Olabilirlik Oran tekniği ile karıştırılmaması gerektiğini ifade etmektedirler. Ayrıca Magis ve diğerleri (2011), Svetina ve Rutkowski (2014) ve Rollins III (2018) tarafından belirtildiği üzere GLR tekniğinin kullanıldığı çalışmalarda grup sayısı arttığında grupların lojistik eğrileri arasındaki küçük farklılıkların dahi ilgilenilen maddelerin DMF'li olarak belirlenmesine yol açabileceği belirtilmektedir. Böylece gerçekte DMF içermeyen maddelerin GLR tarafından DMF'li olarak işaretlenmesinin (Tip I hata) söz konusu olabileceği ifade edilmektedir. Son olarak iki grup için önerilen etki büyüklüğü katsayılarının çoklu gruplar için de geliştirilerek kullanılabileceği araştırmacılar tarafından vurgulanmaktadır.

Madde Tepki Kuramı'na Dayalı Lojistik Regresyon (Lordif)

İkili puanlanan maddelerde DMF belirlemek amacıyla LR tekniği geliştirildikten sonra (Swaminathan ve Rogers, 1990) Sıralı LR (SLR) tekniği de alanyazına kazandırılarak çoklu puanlanan maddeler için de DMF analizlerinin yapıldığı belirtilmektedir. Hem LR hem de SLR tekniklerinde bireyin yeteneğinin bir göstergesi olan toplam test puanı eşleştirme kriteri olarak ele alınmaktadır (Osterlind ve Everson, 2009; Zumbo, 1999). DMF çalışmalarında aynı yetenek düzeylerindeki bireylerin performansları karşılaştırıldığından, eşleştirme kriteri son derece önemlidir ve KTK ve MTK'ya dayalı tekniklerde eşleştirme kriterlerinin farklılık gösterdiği açıklanmaktadır (Choi vd., 2011; Crane, Gibbons vd., 2006).

MTK temelli tekniklere göre hesaplama kolaylığı sunan LR ve SLR tekniklerinde (Bowe, 2017) eşleştirme kriteri için toplam puanın dikkate alınması önemli bir sınırlılık olarak görülmektedir (Swaminathan ve Rogers, 1990). Millsap ve Everson (1993) yaptıkları çalışmalarında toplam puanın eşleştirme kriteri olarak kullanılmasının oluşturduğu dezavantajlardan bahsederek, eşleştirme kriterinin belirlenmesi sürecinde MTK yaklaşımına başvurmanın daha iyi bir seçenek olabileceğini belirtmişlerdir. Camilli ve Shepard (1994) ise çalışmaları kapsamında LR tekniği için ilgili testin dışındaki farklı yetenek tahminlerinin veya ilgili test dikkate alındığında MTK'ya dayalı olarak kestirilen yetenek parametrelerinin eşleştirme kriteri olarak kullanılabileceğini ifade etmişlerdir. Bu bağlamda oluşan genel fikir birliği neticesinde bir yandan LR tekniğinin diğer yandan ise MTK'nın özelliklerini barındıran yeni bir tekniğin geliştirilmesinin önemli olduğu düşünülmüştür. Bu anlamda esnek yapısı nedeniyle geliştirilmesi mümkün olan LR tekniğinin, MTK çerçevesinde yeniden ele alınmasının yararlı olabileceği vurgulanmıştır.

Çok sayıda araştırmacı tarafından LR tekniği ile ilgili yapılan öneriler bağlamında, eşleştirme kriteri olarak toplam puan yerine MTK temelli yetenek kestirimlerinin dikkate alındığı yeni bir teknik geliştirilmiştir. Böylelikle hem LR tekniğinin hem de MTK'nın özellikleri birleştirilerek DMF belirleme sürecinde kullanılabilecek yeni ve daha gelişmiş bir teknik alanyazına kazandırılmıştır. Bu yeni teknikte toplam puan yerine MTK'ya dayalı kestirilen yetenek parametreleri kullanıldığından daha doğru sonuçlara ulaşılmasının mümkün olduğu belirtilmiştir (Choi vd., 2011; Crane, Gibbons vd., 2006; Feuerherd, Knuth, Muehlan ve Schmidt, 2014).

Alanyazında ilk önce “difwithpar” olarak adlandırılan bu tekniğin hem ikili puanlanan hem de çoklu puanlanan maddeler için kullanılabildiği belirtilmektedir. Bu

teknikten her iki puan türüne sahip olan maddeler için de yararlanılabilmesi nedeniyle, ilgili tekniğin DMF çalışmaları için son derece kullanışlı olduğu ifade edilmektedir (Crane, Gibbons vd., 2006). Yapılan sonraki çalışmalarda MTK'ya dayalı sıralı lojistik regresyon (Logistic ordinal regression differential item functioning using IRT) tekniği olarak açıklanan bu teknik, kısaca "Lordif" olarak isimlendirilmektedir (Choi vd., 2011).

Son derece kapsamlı ve esnek bir teknik olan Lordif'in, ikili puanlanan maddeler için LR (Swaminathan ve Rogers, 1990) ve çoklu puanlanan maddeler için SLR (Miller ve Spray, 1993; Zumbo, 1999) teknikleri temel alınarak geliştirildiği belirtilmektedir. Bu nedenle Lordif tekniği, her iki puan türü için de kullanılabilmekte ve bu durum önemli bir avantaj olarak görülmektedir. Lordif tekniğinde KTK'ya dayalı olarak geliştirilen MH, LR, SIBTEST gibi tekniklerden farklı olarak bireyleri eşleştirmek amacıyla MTK'ya dayalı kestirilen yetenek parametrelerinin kullanılması nedeniyle bu tekniğin de MTK'ya dayalı teknikler arasında yer alması gerektiği belirtilmektedir (Choi vd., 2011; Crane, Gibbons vd., 2006). LR tekniğinde olduğu gibi hem TB hem de TBO DMF'nin belirlenmesine olanak sunan bu tekniğin, bir başka avantajı da çoklu gruplarda kullanılabilmesidir. Özetle hem KTK'nın hem de MTK'nın özelliklerine sahip, yani hibrit bir yaklaşıma dayalı olan bu tekniğin, ikili ve çoklu puanlanan maddelerde ve grup sayısı iki veya daha fazla olduğunda kullanılabilmesi nedeniyle son derece kullanışlı olduğunu söylemek mümkündür (Choi vd., 2011; Crane, Gibbons vd., 2006; Fischer ve Karl, 2019; Jafari, Allahyari, Salarzadeh ve Bagheri, 2015).

Lordif tekniği için bir boş model ve iç içe yuvalanmış (nested) üç farklı model kurulmaktadır. Bu teknikte DMF türünü belirlemek amacıyla Olabilirlik Oranı testi kullanılarak, iç içe modeller karşılaştırılmaktadır. Üç model test edildikten sonra ilgili maddede DMF olup olmadığı ve DMF türü (TB ve TBO) belirlenebilmektedir (Choi vd., 2011; Feuerherd vd., 2014). Testteki maddelerin her biri için bağımsız değişkenlerin eklendiği üç aşamalı iç içe yuvalanmış modeller Eşitlik 18'de verilmiştir.

$$\text{Model 0: } \text{logit } P(u_i \geq k) = \alpha_k$$

$$\text{Model 1: } \text{logit } P(u_i \geq k) = \alpha_k + \beta_1 * \text{yetenek} \quad (18)$$

$$\text{Model 2: } \text{logit } P(u_i \geq k) = \alpha_k + \beta_1 * \text{yetenek} + \beta_2 * \text{grup}$$

$$\text{Model 3: } \text{logit } P(u_i \geq k) = \alpha_k + \beta_1 * \text{yetenek} + \beta_2 * \text{grup} + \beta_3 * \text{yetenek} * \text{grup}$$

Eşitlik 18'de verilen her bir modeldeki u_i , i . madde için verilen yanıtları; α_k , k kategorisine bağlı olan kesişim terimini; $P(u_i \geq k)$, k kategorisi ve üzerindeki yanıt olasılığını göstermektedir. Yetenek terimi gizil bir örtük özelliği temsil etmektedir. Eşitlik

18’de görüldüğü üzere boş modele (Model 0) yetenek dahil edilerek Model 1 oluşturulmaktadır. Benzer şekilde Model 2’de grup değişkeni eklenerek madde performansını belirlemek amacıyla hem yetenek hem de grup değişkenleri birlikte kullanılmaktadır. Model 3’te ise yetenek ve grup etkileşimi Model 2’ye dahil edilmektedir. Böylece her açıklayıcı değişken yeni bir regresyon modeline aşamalı olarak eklenmekte ve üç aşamalı iç içe yuvalanmış model oluşturulmaktadır. Analiz sürecinde öncelikle ilgili maddenin DMF’li olup olmadığı Model 1 ve Model 3 karşılaştırılarak değerlendirilmektedir. DMF’li olarak belirlenen bir maddede DMF türünü tespit etmek amacıyla model karşılaştırmalarına devam edilmesi gerekmekte ve Model 1 ile Model 2 karşılaştırıldığında TB DMF, Model 2 ve Model 3 karşılaştırıldığında ise TBO DMF incelenebilmektedir (Choi vd., 2011; Crane, Gibbons vd., 2006; Feuerherd vd., 2014).

Bu teknikte gerçekleştirilen model karşılaştırmalarında izlenen adımlar ve yapılan değerlendirmeler LR tekniği ile aynıdır. Bu bağlamda Swaminathan ve Rogers (1990) yaptıkları araştırmalarında LR tekniği için öncelikle Model 1 ve Model 3’ün karşılaştırılıp ilgili maddede DMF tespit edilmesinden sonra, diğer modellerin birbirleriyle karşılaştırılarak DMF türünün belirlenmesi gerektiğini açıklamışlardır. Bu süreçte Model 1 ile Model 2’nin log olabilirlik oran değerleri arasındaki ki-kare farkı 1 serbestlik derecesindeki ki-kare dağılımındaki değerle karşılaştırılır ve elde edilen anlamlı sonuç ($\beta_2 \neq 0$) farklı yetenek düzeylerindeki gruplar arasında gözlemlenen farkın sabit bir şekilde gruplardan birinin lehine olduğunu gösterir. Benzer şekilde Model 2 ile Model 3’ün log olabilirlik oran değerleri arasındaki ki-kare farkı 1 serbestlik derecesindeki ki-kare dağılımındaki değerle karşılaştırılır ve elde edilen anlamlı sonuç ($\beta_3 \neq 0$) farklı yetenek düzeylerindeki gruplar arasında gözlemlenen farkın yönünün değiştiği şeklinde değerlendirilir. Alanyazında belirtildiği üzere Lordif’in çoklu gruplarda kullanılması durumunda hem TB hem de TBO DMF için kurulan modellerin log olabilirlik oran değerleri arasındaki ki-kare farkı, serbestlik derecesi odak grup sayısına eşit olan ki-kare dağılımındaki değer ile karşılaştırılmaktadır (Choi vd., 2011; Feuerherd vd., 2014; Jafari vd., 2015; Mair, 2018; Vaughn-Coaxum vd., 2016).

Lordif tekniğinde DMF’li maddeler Olabilirlik Oran testine dayalı olarak belirlendikten sonra etki büyüklüğü için sahte R^2 katsayısının yorumlanması gerektiği açıklanmıştır (Choi vd., 2011; Kristjansson, Aylesworth, McDowell ve Zumbo, 2005). Çünkü LR tekniğinde olduğu gibi Lordif tekniğinde de grupların örneklemeleri ne kadar büyükse, tekniğin manidar sonuç ortaya koyma olasılığının yükseleceği ifade edilmiştir (Fischer ve Karl, 2019). Başka bir ifadeyle örneklem büyüklüğü arttıkça iki model için

karşılaştırılan log olabilirlik oran değerlerindeki küçük farkların bile manidar sonuçlara neden olduğu belirtilmiştir. Dolayısıyla büyük örneklemelerin manidarlık testleri üzerindeki etkisi nedeniyle, istatistiksel olarak manidar bulunan olabilirlik oran testi sonucunda DMF büyüklüğünü tespit etmek için McFadden's sahte R^2 , Nagelkerke ile Cox ve Snell etki büyüklüklerinden birinin yorumlanması gerektiği önerilmiştir (Bowe, 2017; Choi vd., 2011). Lordif tekniğinin kullanıldığı araştırmalarda genellikle McFadden's sahte R^2 etki büyüklüğü katsayısının yorumlandığı görülmüştür (Hahn, vd., 2014; Jafari vd., 2015; Vaughn-Coaxum vd., 2016). Bu bilgiler doğrultusunda Olabilirlik Oran testinin istatistiksel gücünün örneklem büyüklüğüne bağlı olduğu göz önüne alındığında, kurulan modeller arasındaki farklar (ΔR^2) belirlenirken McFadden's sahte R^2 veya diğer etki büyüklüğü katsayılarından birine başvurmak gerektiği ifade edilmiştir (Choi vd., 2011; Kristjansson vd., 2005; Vaughn-Coaxum vd., 2016; Zumbo, 1999).

Hem TB hem de TBO DMF'li maddelerde kullanılan etki büyüklüğü katsayıları için öncelikle Zumbo ve Thomas (1996) tarafından üç farklı değerlendirme kriteri ortaya konulmuştur [ihmal edilebilir ($\Delta R^2 < .13$), orta ($.13 \leq \Delta R^2 \leq .26$) ve büyük ($\Delta R^2 > .26$)]. Sonrasında ise daha önce önerilen kriterlerin çok yüksek olması nedeniyle Jodoin ve Gierl (2001) tarafından yapılan çalışmada farklı bir sınıflama önerilmiştir [ihmal edilebilir ($\Delta R^2 < .035$), orta ($.035 \leq \Delta R^2 \leq .070$) ve büyük ($\Delta R^2 > .070$)]. Bu kriterler incelendiğinde Jodoin ve Gierl (2001) tarafından yapılan sınıflamanın Zumbo ve Thomas (1996) tarafından yapılan sınıflamadan daha hassas olduğu görülmüştür.

Mevcut araştırmalardan elde edilen sonuçların aksine Choi ve diğerleri (2011) tarafından yapılan araştırmada, etki büyüklüğü katsayılarının çok büyük olması nedeniyle gerçekte DMF'li olan maddelerin DMF'li olarak belirlenemeyebileceği ileri sürülmüştür. Bu nedenle Crane, Belle ve Larson, (2004) ile Crane, Gibbons ve diğerleri (2006) tarafından yapılan çalışmalarda DMF sonuçlarını değerlendirmek amacıyla alanyazında kabul gören kriterlerden daha katı olan farklı kriterlerin kullanılması gerektiği önerilmiştir. Bu araştırmacılar tarafından TBO DMF için değerlendirme kriterinde bir farklılaşmaya gerek duyulmadığı, ancak TB DMF için β_1 katsayıları arasındaki %10'luk bir farklılaşmanın TB DMF'nin var olduğu şeklinde yorumlanması gerektiği belirtilmiştir. Benzer şekilde bazı araştırmacıların yaptıkları çalışmalarda hesaplanan β katsayıları arasındaki farkın %5 veya %1 olması durumunda da maddenin DMF'li olarak kabul edilebildiği ifade edilmiştir (Cook, Bamer, Amtmann, Molton ve Jensen, 2012; Crane, Hart, Gibbons ve Cook, 2006; Croom ve Svetina, 2021).

Choi, Gibbons ve Crane (2016) tarafından 0.3-3 versiyonu geliştirilen R yazılımındaki Lordif paketinde, eşleştirme kriteri olarak hem MTK'ya dayalı kestirilen yetenek parametreleri, hem de KTK'ya dayalı olarak elde edilen toplam puanlar kullanılarak analizler gerçekleştirilebilmektedir. Ayrıca bu paket kullanılarak McFadden's sahte R^2 , Nagelkerke ile Cox ve Snell etki büyüklüğü katsayıları da hesaplanabilmektedir (Choi vd., 2011).

LR ve SLR tekniklerinden farklı olarak Lordif tekniği için hesaplanan istatistik sonucunu değerlendirmek amacıyla Lordif paketinde yer alan Monte Carlo simülasyon yaklaşımından yararlanılarak ampirik bir eşik değeri belirlenmektedir. Monte Carlo yaklaşımına dayalı olarak simüle edilen veri setleri DMF içermeyen koşullar altında üretilmektedir. Satır ve sütun olarak ampirik veri seti ile aynı boyutlara sahip olacak şekilde simüle edilen bu veri setleri hem örneklem büyüklüğü hem de yetenek kestirimi açısından ampirik veri seti ile benzer özellik göstermektedir. Monte Carlo simülasyonu ile yinelenmeli bir süreç gerçekleştirilerek modeller arasında karşılaştırma yapılırken farklı kritik değerler kullanılabilir ve elde edilen kesme noktasını aşan test istatistikleri DMF'nin varlığına işaret etmektedir. Bu özelliği nedeniyle Monte Carlo simülasyon yaklaşımı kullanılarak Tip I hata oranı kontrol edilebilmektedir. Sahte R^2 değerlerinin büyüklüğünü değerlendirmek amacıyla da yine Monte Carlo simülasyonu yaklaşımının kullanılması tavsiye edilmektedir. Ayrıca Lordif tekniğinde istatistik sonuçlarının doğruluğunu önemli ölçüde artırdığı belirtilen, yinelenmeli bir saflaştırma işlemi paket içerisinde otomatik olarak yapılmaktadır. Bu süreçte DMF'li olmayan maddeler çapa madde olarak ele alınmakta ve analiz sürecinde gerçekleştirilen iki ardışık yinelemede aynı madde kümesi DMF'li olarak tespit edilinceye kadar yineleme işlemi devam etmektedir. Böylelikle testte DMF'li maddeler bulunmasına rağmen, eşleştirme kriterinin DMF'li maddelerin etkisinden arındırılması sonucunda yetenek parametresinin yansız kestirilebildiği ifade edilmektedir. Lordif tekniğinde yinelenmeli bir saflaştırma sürecinin ek bir işlem gerektirmeden otomatik gerçekleştirilebilmesi, gerçek test uygulamalarında bu tekniğin son derece kullanışlı olduğunu göstermektedir. Belirtilen avantajlarının yanı sıra Lordif tekniğindeki hesaplamalarda Monte Carlo simülasyon yaklaşımından yararlanılması nedeniyle, analizin tamamlanması için uzun zamana ihtiyaç duyulduğu belirtilmektedir. Ayrıca diğer tekniklere kıyasla daha yakın zamanda geliştirilmiş olan Lordif tekniğinin sınırlı sayıda çalışmada kullanıldığı ifade edilmektedir (Allahyari, 2020; Choi vd., 2011, 2016; Crane, Gibbons vd., 2006; Fischer ve Karl, 2019; Hahn vd., 2014; Jafari vd., 2015; Opariuc-Dan vd., 2017; Vaughn-Coaxum vd., 2016).

Lordif tekniğindeki parametre kestirimi sürecinde ya Samejima (1969) tarafından ortaya konulmuş olan Graded Response Model (GRM) ya da Muraki (1992) tarafından ortaya konulmuş olan Generalized Partial Credit Model (GPCM) kullanılabilmektedir (Chalmers, 2012; Mair, 2018; Schauburger ve Mair, 2020). Yapılan araştırmalar incelendiğinde ise genellikle GRM'nin kullanıldığı görülmektedir (Crane, Hart vd., 2006; Crins vd., 2018; Jafari vd., 2015; Paz vd., 2017; Vaughn-Coaxum vd., 2016). Bu bilgiler ışığında yapılan araştırma kapsamında da MTK'ya dayalı parametre kestirimlerinde GRM'nin kullanılması tercih edilmiştir.

Bowe (2017) gerçek veri üzerinde yapmış olduğu araştırmasında, yedi grubun eşzamanlı olarak karşılaştırılması durumunda analizin hata uyarısı verdiğini belirterek, Lordif tekniğinin bir sınırlılığını tespit etmiştir. Böylece grup sayısı arttığında hem sonuçların yorumlanmasının zorlaşması hem de analiz sürecinde hata oluşabilmesi nedeniyle, bu tekniğin daha az sayıda grup için uygun olduğu ifade edilmiştir (Fischer ve Karl, 2019). Fischer ve Karl (2019) tarafından yapılan çalışmada Lordif'in gelişmiş bir teknik olmasının yanı sıra kullanıcı dostu da olması nedeniyle özellikle 2010 yılından sonraki DMF belirleme çalışmalarında sıklıkla kullanıldığı açıklanmıştır.

Alanyazında Lordif tekniğinin kullanıldığı araştırmalarda genellikle çoklu puanlanan maddelerden oluşan gerçek veri setleri üzerinde analiz yapılmış ve DMF'nin yalnızca iki grup arasında incelendiği görülmüştür (Bowe, 2017; Crins vd., 2018; Croom ve Svetina, 2021; Demir, 2018; Fallon, Cathcart ve Johnson, 2021; Feuerherd vd., 2014; Geramipour ve Shahmirzadi, 2019; Hahn, vd., 2014; Hong vd., 2018; Jafari vd., 2015; Li, She, Tu ve Cai, 2021; Paz vd., 2017; Tan vd., 2018; Vaughn-Coaxum vd., 2016; Zhong vd., 2014). İki grup dikkate alınarak Lordif tekniğinin kullanıldığı çok sayıdaki çalışmanın haricinde Gregory, Werth, Reilly ve Jevsevar (2021) ile Webster ve Bowe (2021) tarafından yapılan çalışmalar kapsamında yine çoklu puanlanan maddelerden oluşan gerçek veri seti kullanmasına rağmen, ikiden fazla grup arasında DMF incelenmiştir. Tunç'un (2016) yaptığı simülasyon çalışmasında ise çeşitli koşullar dikkate alınarak, iki grupta hem ikili hem de çoklu puanlanan maddeler için veri üretilmiş ve Lordif, MH, ANOVA tekniklerinin performansları karşılaştırılmıştır. Benzer şekilde Schauburger ve Mair (2020) tarafından yapılan simülasyon çalışmasında iki grup için hem ikili hem de çoklu puanlanan maddelerden oluşan veri setleri üretilerek, LRL, MH ve Lordif tekniklerinin performansları incelenmiştir. Ayrıca bu araştırmacılar tarafından çoklu puanlanan maddelerden oluşan gerçek bir veri seti de kullanılmış ve bu veri setindeki maddelerin çoklu gruplar arasında DMF gösterip göstermediği LRL ve Lordif

teknikleri ile belirlenmiştir. Bu araştırmaların tamamı dikkate alındığında, çoklu gruplar için ikili veya çoklu puanlanan gerçek veri setlerinin analizinde Lordif'in kullanıldığı görülmesine rağmen, çoklu gruplar için ikili puanlanan simülatif veri setleri üzerinde bu tekniğin performansının incelendiği herhangi bir araştırmaya rastlanılmamıştır. Bu nedenle çoklu gruplar için ikili puanlanan maddelerde kullanılan Lordif tekniğinin farklı koşullar altındaki performansının incelenerek, başka tekniklerle de karşılaştırılması gerektiği sonucuna ulaşılmıştır.

Yapılan bu araştırma kapsamındaki alanyazın incelemeleri sonucunda çoklu gruplarda, ancak çoklu puanlanan maddelerden oluşan veri setleri kullanılarak çeşitli DMF belirleme tekniklerinin dikkate alındığı çalışmalara da rastlanılmış olmasına rağmen (Allahyari, 2020; Gregory vd., 2021; Langer, 2008; Schauburger ve Mair, 2020; Scott, Fayers, Aaronson, vd., 2009; Tay vd., 2015; Teymoori vd., 2020; Vaughn-Coaxum vd., 2016; Woods vd., 2013) bu araştırmanın amacı doğrultusunda yalnızca ikili puanlanan maddeler için geliştirilen tekniklerin kullanıldığı çalışmalar dikkate alınmıştır. Bu bağlamda ikili puanlanan maddeler için çoklu gruplarda DMF belirlemek amacıyla geliştirilen çeşitli tekniklerin performanslarının simülatif veya gerçek veri kullanılarak araştırıldığı sınırlı sayıda çalışma detaylı bir şekilde incelenmiştir.

Bu araştırmalardan ilki Kim ve diğerleri (1995) tarafından gerçekleştirilmiştir. Alanyazındaki ilk çoklu grup DMF belirleme tekniği olarak belirtilen GLK'nın, Kim ve diğerleri (1995) tarafından geliştirilerek gerçek veri seti üzerinde kullanıldığı araştırmada; analiz edilen veriler Cohen ve Kim'in (1992) çalışmasında kullanılan testten elde edilmiştir. Bu testten elde edilen veri seti, sınav anında hesap makinesi kullanma durumuna göre her biri 200 öğrenciden oluşan üç gruptan oluşmuştur. Gruplardan her birine beş seçenekli çoktan seçmeli olarak geliştirilen ve ikili puanlanan 14 maddelik matematik testi uygulanmıştır. Gruplardan biri referans olarak seçilmiş ve bu grubun sınavda hesap makinesi kullanmasına izin verilmemiştir. Diğer grupların ise farklı özelliklere sahip bilimsel hesap makinesi kullanmaları sağlanmış ve birinci odak grubun hesap makinesi-1'i, ikinci odak grubun ise hesap makinesi-2'yi kullandığı belirtilmiştir. Araştırma kapsamında çeşitli gruplardaki öğrencilerin hesap makinesi kullanıp kullanmamaları ve farklı özellikleri olan hesap makinelerine sahip olma durumları dikkate alınarak, testteki maddelerin DMF açısından incelenmesi amaçlanmıştır. Böylece hangi maddelerin hesap makinesi kullanımına duyarlı olduğu belirlenmiştir. Parametrelerin marginal maximum likelihood (MML) ile kestirildiği araştırmada, veri setinin tek boyutluluk varsayımını karşıladığı ve 2PLM'ye uyum sağladığı ifade

edilmiştir. Analiz sürecinde öncelikle üç grup arasında eş zamanlı olarak DMF incelemesi yapılmış ve sonucun manidar çıkmasının ardından gruplar arasında ikili karşılaştırma yapılarak hangi gruplar arasında DMF olduğu belirlenmiştir. Bu bağlamda GLK, Lord ki-kare ve ikili karşılaştırmalara dayalı Q_j istatistiği kullanılarak elde edilen sonuçlar karşılaştırılmıştır. Araştırmanın bulgularına göre GLK tekniği ile iki maddenin (10 ve 14) DMF gösterdiği sonucuna ulaşılmıştır. Yapılan ikili karşılaştırmalarda yalnızca 14. maddenin hesap makinesi kullanmayan grup ile hesap makinesi-2'yi kullanan grup arasında DMF'li olarak işaretlendiği tespit edilmiştir. Q_j istatistiğinde ise 10 ve 14. maddede hesap makinesi kullanmayan grup ile hesap makinesi-2'yi kullanan grup ve hesap makinesi kullanmayan grup ile hesap makinesi-1'i kullanan grup arasında DMF belirlendiği sonucuna ulaşılmıştır. Lord ki-kare ve Q_j istatistiğinde, farklı parametre dönüştürme yaklaşımı kullanıldığından analiz sonuçlarının farklılaşması beklenen bir sonuç olarak görülmüştür. Bununla birlikte hem Lord ki-kare tekniği hem de Q_j istatistiğine göre hesap makinesi-1 ve hesap makinesi-2 arasında DMF gözlenmemiştir. Bu bilgiler doğrultusunda hesap makinesi kullanmanın matematik testindeki iki maddeye doğru yanıt vermek için avantaj sağladığı ifade edilmiştir.

Penfield (2001) tarafından yapılan araştırmada, grup sayısının ikiden fazla olduğu durumda ikili karşılaştırma yaklaşımının dikkate alınmasından kaynaklı olan sınırlılıkların üstesinden gelmek amacıyla çoklu gruplar için GMH tekniği önerilmiştir. Araştırma çerçevesinde referans grup sayısının bir ve odak grup sayısının bir, iki, üç ve dört olduğu durumlar için MH ki-kare, Bonferroni düzeltmeli MH ki-kare (BMH) ve GMH olmak üzere üç farklı Mantel-Haenszel tekniği kullanılmış ve çeşitli koşullar altında belirtilen üç farklı tekniğin Tip I hata ve istatistiksel güç oranları karşılaştırılmıştır. Araştırmada 3PLM'ye dayalı olarak 60 maddelik simülasyon veri üretilmiş ve grup sayısı (2, 3, 4, 5), DMF içeren odak grup sayısı (1, 2, 3, 4), her bir gruptaki örneklem büyüklüğü (250, 500, 1000), referans ve odak grupların yetenek dağılımları $[N(0, 1), N(-1, 1)]$ ve DMF miktarı ($b_R - b_O = 0.0$; $b_R - b_O = 0.4$) ile ilgili çeşitli koşullar belirlenmiştir. Odak gruplardaki grup sayısına göre bütün olası durumlar dikkate alınmıştır. Örneğin odak grup sayısının üç olduğu durum için yalnızca bir, yalnızca iki ve her üç grupta da DMF'li maddelerin yer alması sağlanmıştır. Örneklem büyüklüğü (250, 500, 1000) için ise referans ve odak grupların örneklem büyüklüklerinin eşit ve eşit olmama durumları ele alınmıştır. Ayrıca incelenen madde dışındaki maddelerin (2, 5 ve 10) de odak gruplar için DMF'li olması sağlanmıştır. Madde parametreleri belirlenirken gerçek veri kullanılarak yapılan çalışmalardaki dağılımlar

dikkate alınmıştır. Güçlük parametresi için $N(0, 1)$ olan normal dağılım; ayırt edicilik parametresi için ise $\log N(0, 0.1125)$ olan log-normal dağılım dikkate alınarak madde parametreleri üretilmiştir. Şans parametresi ise tüm maddeler için .20 olarak ele alınmıştır. Araştırmanın sonuçlarına göre grupların yetenek dağılımlarının eşit ve bütün odak gruplarda DMF olduğu durumda; MH için odak grup sayısı ve örneklem büyüklüğü arttıkça Tip I hata oranı kabul edilemeyecek düzeyde artmıştır. GMH tekniğinde Tip I hata oranının .05 seviyesine yakın olduğu ve odak grup sayısı ile örneklem büyüklüğü arttıkça Tip I hata oranında önemli bir yükselme olmadığı bulunmuştur. Ayrıca grup sayısı arttığında GMH'nin istatistiksel güç oranında da önemli bir artışın söz konusu olmadığı belirlenmiş ve bu teknik için istatistiksel güç oranlarının diğer tekniklere yakın olduğu görülmüştür. Tüm odak gruplar DMF içerdiğinde GMH tekniği için hesaplanan istatistiksel güç oranının bir miktar düşme eğiliminde olduğu belirlenmiştir. Küçük örneklemelerde üç tekniğin istatistiksel güç oranları arasında daha belirgin farklar söz konusu iken büyük örneklemelerde farkın azaldığı ifade edilmiştir. Her üç teknik için de referans ve odak grupların yetenek dağılımlarının eşit olduğu durumlardaki istatistiksel güç oranları yetenek dağılımlarının eşit olmadığı duruma göre daha yüksek bulunmuştur. Buradan hareketle GMH tekniğinin çoklu gruplarda etkili olduğu sonucuna ulaşılmıştır.

Fidalgo ve Scalon (2010) tarafından yapılan araştırmada ikiden fazla grupta DMF'nin nasıl tespit edilebileceği hem ikili hem de çoklu puanlanan maddeler için açıklanarak ayrıntılı bir şekilde gösterilmiştir. Araştırma kapsamında yapılan simülasyon çalışması ile ikili puanlanan maddeler için MH, BMH, GMH, çoklu puanlanan maddeler için ise Mantel Test (MT), Bonferroni düzeltmeli MT ve GMH tekniklerinin DMF belirlemedeki performansları incelenmiştir. Testteki 27 madde arasından ikili puanlanan 20 madde ve çoklu puanlanan yedi madde için önceki araştırmalara dayalı olarak madde parametreleri üretilmiş ve ikili ile çoklu puanlanan üçer madde DMF'li olarak belirlenmiştir. Grup sayısı ise üç ve dört olarak dikkate alınmıştır. İkili puanlanan maddeler için 3PLM'ye göre, çoklu puanlanan maddeler için ise Kısmi Puan Modeli'ne dayalı olarak veri üretilmiştir. İkili puanlanan maddelerde DMF'li üç maddede referans grup için üretilen ayırt edicilik ($a_1 = 1.162$; $a_2 = 0.638$; $a_3 = 1.000$) ve güçlük ($b_1 = -1.75$; $b_2 = -0.50$; $b_3 = 0.00$) parametreleri maddeler arasında farklılaşmış, şans parametresi (.15) ise sabit olarak ele alınmıştır. Çoklu puanlanan maddeler için ise üç maddenin adım güçlükleri ($b_{11} = -0.91$; $b_{12} = 0.98$; $b_{13} = 0.21$; $b_{21} = -2.25$; $b_{22} = -1.8$; $b_{23} = 1.66$; $b_{31} = -0.54$; $b_{32} = -2.11$; $b_{33} = 0.74$) olarak belirlenmiştir. Araştırmada iki örneklem büyüklüğü ($N_R=1000$, $N_{O1}=N_{O2}=N_{O3}=250$) belirlenmiştir. Yalnızca TB DMF'nin incelendiği

çalışmada odak gruplara DMF miktarı olarak 0.4 eklenmiştir. Referans ve odak grupların yetenek dağılımları ise $N(0, 1)$ olarak belirlenmiş ve maddelerin DMF'li olmadığı (%0) veya DMF'li olduğu (%22) durumlar dikkate alınmıştır. Diğer taraftan DMF'nin örüntüsü açısından maddelerin DMF'li olmadığı, DMF'li maddelerin bir grupta olduğu veya DMF'li maddelerin gruplara dengeli dağıldığı durumlar ele alınmıştır. Araştırma sonuçları, DMF verilmeyen durumda hem ikili ve çoklu puanlanan maddelerde hem de iki ve üç odak grup olması durumunda GMH tekniği için Tip I hata oranının iyi kontrol edildiğini göstermiştir. Ancak ikili karşılaştırmalar yapılması nedeniyle MH ve MT teknikleri için hesaplanan Tip I hata oranları yüksek bulunmuştur. Grup sayısı arttıkça MH ve MT teknikleri için hesaplanan Tip I hata oranlarının bir miktar yükseldiği belirlenmesine rağmen, Bonferroni düzeltmesi yapıldığında bu iki teknik için Tip I hata oranlarının iyi kontrol edildiği görülmüştür. Ayrıca maddelerin %22'sinin DMF'li olduğu durumda bütün teknikler için Tip I hata oranının yükseldiği belirtilmiştir. Bununla birlikte grup sayısı arttıkça BMH, BMT ve GMH tekniklerinin istatistiksel güç oranlarında ciddi bir değişiklik olmamıştır. Son olarak her koşulda BMH ve BMT tekniklerinin istatistiksel güç oranlarının GMH tekniğine eşit veya biraz düşük olması nedeniyle hem ikili hem de çoklu puanlanan maddelerde GMH tekniğinin kullanılması gerektiği önerilmiştir.

Magis ve diğerleri (2011) araştırmalarında ikili puanlanan maddeler için GLR tekniğini geliştirmişlerdir. Araştırmada Kanada'da bir eyalet olan Quebec'teki kolejlerde Fransızca konuşan Kanadalı öğrencilerin üniversite öğrenimlerinin başındaki İngilizce performanslarına göre seviyelere ayrılabilmesi için TCALS-II testindeki kısa İngilizce metinlerden oluşan 15 madde kullanılmış ve toplanan gerçek veri üzerinde GLR, GMH ve GLK tekniklerinin performansları karşılaştırılmıştır. Farklı yıllarda yapılan uygulamaların (1998, 2000, 2002 ve 2004) dikkate alındığı araştırmada ilk uygulama yılı referans (referans grup), sonraki uygulamalar ise odaklanılan yıllar (odak gruplar) olarak ele alınmıştır. Referans ve üç odak grubun örneklem büyüklükleri 1277 ile 1547 arasında farklılık göstermiştir. Bütün uygulamalar için hesaplanan negatif çarpıklık katsayıları yüksek puanlı öğrencilerin daha büyük oranda olduğunu ortaya koymuştur. Böylece araştırma kapsamındaki dört uygulamada da aynı testin kullanılması nedeniyle, testteki maddelerin yıllara göre DMF gösterip göstermediği incelenmiştir. Hem TB hem de TBO DMF'nin belirlendiği araştırmada GLK testi için 1PLM ve 2PLM'ye göre kestirim yapılmıştır. GLR tekniğinin analiz sürecinde Wald ki-kare ve Olabilirlik Oran testleri birlikte kullanılmıştır. Araştırmada elde edilen bulgulara göre Wald ki-kare testi ile dört maddenin, Olabilirlik Oran testi ile de beş maddenin DMF gösterdiği sonucuna

ulaşılmıştır. Böylece GLR için elde edilen sonuçlara göre her iki testte yalnızca bir maddenin DMF’li olarak belirlenmesinde farklılaşma olduğu bulunmuştur. İki madde TB DMF, diğer iki madde ise TBO DMF göstermiştir. TB DMF’li olarak belirlenen maddelerin ikisi üç teknikte de ortaktır. Ayrıca TB DMF için GMH ve GLK (1PLM) tekniklerinin aynı maddeleri DMF’li olarak belirlediği sonucuna ulaşılmıştır. TBO DMF için ise GLK ile GLR’nin iki maddede ortak olduğu, iki maddede ise yalnızca GLK’nın DMF gösterdiği tespit edilmiştir. GLK tekniği özelinde 8. ve 10. madde hem 1PLM hem de 2PLM ile yapılan analizler sonucunda DMF’li olarak işaretlenmiş ve bu durum GLK’nın aynı maddeleri hem TB hem de TBO DMF’li olarak belirlediğini göstermiştir. Böylelikle hem TB hem de TBO DMF için yapılan analiz sonuçlarına göre çoklu grup teknikleri tarafından DMF’li olarak tespit edilen maddelerin farklılaştığı belirtilmiştir. Bu anlamda GLR, GMH ve GLK tekniklerinin Tip I hata ve istatistiksel güç oranları açısından farklı koşullardaki performanslarının karşılaştırılabilmesi amacıyla çeşitli simülasyon çalışmalarının yürütülmesi gerektiği önerilmiştir.

Svetina ve Rutkowski (2014) tarafından yapılan bir simülasyon çalışmasında uluslararası geniş ölçekli uygulamaların kapsamı dikkate alınarak, büyük örneklemelerden oluşan çok sayıdaki grup için GLR tekniğinin performansı incelenmiştir. Veriler TIMSS 2007 uygulamasının dördüncü sınıf düzeyindeki madde ve yetenek parametreleri dikkate alınarak 2PLM’ye göre üretilmiş ve TIMSS uygulamasındaki gibi her biri 15 maddeden oluşan 7 blok oluşturulmuştur. Araştırmada grup sayısı (10 veya 20), DMF miktarı [$b(0.50 - 1.00)$, $a(0.40 - 0.80)$], DMF’li maddelerin yüzdesi (%20, %40, %60), DMF’nin türü [TB ve TBO(a)] ve DMF’den etkilenen grupların yüzdesi (%40, %70) olmak üzere çeşitli koşullar dikkate alınarak GLR tekniği kullanılmıştır. DMF’den etkilenen grupların yüzdesi farklılaştığından grup sayısı 10 olduğunda 4 veya 7 grubun, grup sayısı 20 olduğunda ise 8 veya 14 grubun DMF’den etkilendiği belirtilmiştir. Gruplar arasından rastgele seçilmiş olan bir grup (8. Grup) referans olarak dikkate alınmış, geri kalan gruplar ise odak gruplar olarak tanımlanmıştır. Grup sayısının 10 ve 20 olduğu durumlarda referans grubun örneklem büyüklüğünün sırasıyla 6370 ve 5341 olduğu ifade edilmiştir. Odak grupların örneklem büyüklükleri ise TIMSS uygulamasındaki örneklem büyüklüklerine benzer şekilde $\sim N(5000, 1000)$ aralığından rastgele belirlenmiştir. Küçük [$b(0.50)$, $a(0.40)$] ve büyük [$b(1.00)$, $a(0.80)$] DMF miktarları referans grubun madde parametrelerine eklenmiş ve çıkarılmıştır. Hesaplanan istatistiksel güç oranlarının genel olarak orta veya yüksek düzeyde olduğu ifade edilmiş ve grup sayısı arttığında hem güçlük hem de ayırt edicilik parametreleri için istatistiksel güç oranlarının önemli bir

şekilde farklılaşmadığı gözlemlenmiştir. Elde edilen sonuçlar DMF’li madde yüzdesi arttığında GLR tekniğinin TB DMF’li maddeleri doğru tespit etmede iyi sonuç vermediğini göstermiştir. Ayrıca gruptaki örneklemelerin büyük ve DMF’nin güçlük parametresinden kaynaklı olduğu durumda DMF miktarındaki artışın istatistiksel güç oranında ciddi bir değişikliğe yol açmadığı ortaya konulmuştur. Ayırt edicilik parametresi için ise diğer koşullar sabit tutulduğunda DMF miktarı 0.40’tan 0.80’e yükseldiğinde istatistiksel güç oranı da artarken, DMF’li madde yüzdesi %20’den %60’a yükseldiğinde istatistiksel güç oranında genel olarak önemli bir farklılaşma olmamıştır. Bununla birlikte ayırt edicilik parametresi için DMF miktarının 0.40 olduğu durumda DMF’den etkilenen grup yüzdesi arttığında istatistiksel güç oranlarında azalma görülmüştür. Ayrıca grup sayısına bakılmaksızın DMF’nin olmadığı durumda Tip I hata oranlarının yüksek olduğu ifade edilmiştir. DMF’nin olduğu durumda ise Tip I hata oranlarının nispeten düşük olduğu gözlenmiştir. Tip I hata açısından ilginç olduğu belirtilen bu sonucun grup ortalamaları arasındaki farklılıklardan veya referans grubun madde parametreleri arasındaki farktan kaynaklanabileceği düşünülmüştür.

Carroll (2015) tarafından çoklu gruplar için yeni geliştirilmiş olan LGÇGW tekniğinin Wald-1 yaklaşımı dikkate alınarak, bu tekniğin farklı simülasyon koşullarındaki performansı Tip I hata ve istatistiksel güç oranları açısından incelenmiştir. Ayrıca bu çalışmada çapa maddelerinin yanlış tanımlanması durumunda ilgili tekniğin performansı araştırılmıştır. Veriler TIMSS 1999 uygulamasının matematik alt testindeki madde parametreleri dikkate alınarak 3PLM’ye göre üretilmiştir. Üretilen veri setindeki ikili puanlanan 50 maddenin maximum 10 tanesi çapa olarak belirlenmiştir. Uygulamaya katılan ülkeler arasından ortalama yetenekleri heterojen olacak şekilde altı ülke tespit edilmiş ve referans grup için yetenek parametresi standart normal dağılıma göre üretilmiştir. TIMSS uygulamasındaki ülkelerin örneklem büyüklükleri dikkate alınarak, toplam altı grubun da örneklem büyüklüklerinin dengesiz ve 2.962 ile 9.072 arasında değişiklik gösterdiği ifade edilmiştir. TB ve TBO(a) DMF miktarının 0.3 ile 0.7 aralığından bir kere rastgele belirlenerek sabit tutulduğu araştırmada çok yönlü ANOVA ile bağımsız değişkenlerin etkileri incelenmiştir. Araştırma kapsamında; grup sayısı (3 ve 6), çapa maddeler haricindeki DMF’li maddelerin sayısı (1, 2 ve 6) ve doğru belirlenen çapa maddelerin sayısı (5, 8 ve 10) olmak üzere yalnızca 18 koşul dikkate alınarak her koşul için 500 tekrar yapılmıştır. Böylece 0, 2 ve 5 yanlış belirlenmiş, yani DMF içeren çapa maddesi dikkate alınmıştır. Çapa olmayan maddelerde ise 1, 2 ve 6 maddenin DMF’li olması sağlanmıştır. Grup sayısı üç olduğunda yalnızca bir odak grubun, altı

olduğunda ise bir veya iki odak grubun DMF'li maddeler içerdiği belirtilmiştir. Elde edilen sonuçlar grup sayısına bakılmaksızın çapa maddelerinde DMF'li madde olmaması durumunda Tip I hata oranının iyi kontrol edildiği belirlenmiştir. Hem üç hem de altı grup için TB DMF'li çapa maddelerin %20 olduğu durumda Tip I hata oranı .10'un altında, %50 olduğu durumda ise .20'nin üstünde bulunmuştur. Öte yandan çapa maddelerin TB DMF içerdiği durumda grup sayısının artması, Tip I hata oranlarının azalmasına neden olmuştur. Ayrıca TB DMF'li maddelerden etkilenen gruplardaki toplam örneklem büyüklüğü ile çapalardaki DMF'li maddelerin yüzdesi birlikte arttığında hesaplanan Tip I hata oranlarında artış gözlemlenmiştir. Araştırmadaki 10 çapa maddesi doğru bir şekilde tanımlandığı sürece LGÇGW tekniğinin Wald-1 yaklaşımı için yapılan analizlerden elde edilen sonuçlar Tip I hata oranının iyi kontrol edildiğini (.05'in altında), istatistiksel gücün ise mükemmel (0.90 ve üzerinde) olduğunu göstermiştir. Aksi taktirde çapa maddelerin yanlış tanımlanması Tip I hata oranlarının yükselmesine, istatistiksel güç oranlarının ise azalmasına neden olmuştur.

Jin ve diğerleri (2015) yapmış oldukları araştırmada faktöriyel DMF analizi yaklaşımını dikkate alarak, analiz sürecinde GLR tekniğini kullanmışlardır. Araştırma kapsamında her biri iki düzeyli olan iki değişkenin (faktör A ve faktör B) ele alındığı bir simülasyon çalışmasında dört farklı grup oluşturulmuştur. TB ve TBO(a) DMF için iki ayrı simülasyon çalışmasının yapıldığı bu araştırmada grupların örneklem büyüklüğü 125 kişi olarak belirlenmiş ve yetenek parametresi standart normal dağılıma göre üretilmiştir. Madde yanıtları ise sırasıyla Rasch Modele ve 3PLM'ye göre üretilmiştir. Testteki 21 madde içerisinde DMF'li olarak ilgilenilen madde sayısının bir olduğu bu araştırmada, incelenen DMF'li madde dışında kalan diğer maddeler çapa olarak dikkate alınmıştır. Araştırmadaki DMF miktarları güçlük parametresi (0, 0.2, 0.4, 0.6) ve ayırt edicilik parametresi (0, 0.13, 0.26, 0.39) için farklılaştırılmakla birlikte ikisinde de dört koşul belirlenmiştir. Ayrıca DMF'li maddelerin yüzdesi (%0, %10, %20) ve DMF kaynağı (Faktör A, Faktör B, AB etkileşim) ele alınmıştır. Araştırma sonuçları hem TB hem de TBO(a) DMF miktarları arttığında istatistiksel güç oranlarının yükseldiğini, testteki DMF'li çapa madde yüzdesi arttığında ise her iki DMF türünde de istatistiksel güç oranlarının önemli ölçüde farklılaşmadığını göstermiştir. Ayrıca bu araştırmada DMF verilmeyen durumda GLR tekniği için her iki DMF türüne ilişkin hesaplanan Tip I hatanın iyi kontrol edildiği belirtilmiştir. DMF miktarı ve testteki DMF'li çapa madde yüzdesi birbirinden bağımsız veya birlikte arttığında ise TB DMF için Tip I hata artarken, TBO(a) DMF için Tip I hatanın ciddi bir şekilde farklılaşmadığı ifade edilmiştir. Böylece testteki

apa maddelerin DMF’li olmasının, yalnızca TB DMF t r  iin Tip I hata oranının y kselmesine neden olduėu ortaya konulmuştur.

Oshima ve diėerleri (2015) tarafından gerekleştiren araştırmada MG-NCDIF tekniğinin oklu gruplar iin farklı sim lasyon koştullarındaki performansı Tip I hata ve istatistiksel g  oranları aısından incelenmiştir. 3PLM’ye dayalı  retilen veri setindeki 30 madde iin ayırt edicilik parametresinin 0.49 ile 1.74, g l k parametresinin -1.16 ile 1.96, şans parametresinin ise .05 ile .22 aralıėında olduėu belirtilmiştir.  retilen verideki maddelerin %20’si (6 madde) DMF’li olarak belirlenmiştir. Yalnızca ayırt edicilik, yalnızca g l k ve hem ayırt edicilik hem g l k parametrelerinden kaynaklı DMF’nin incelendiėi araştırmada; grup sayısı (3 ve 5), grup sayısına g re DMF’li maddelerden etkilenen odak grup sayısı [3 Grupta (1 ve 2), 5 Grupta (1, 2 ve 4)], DMF t r  [TB, TBO(a), TBO(ab)], ayırt edicilik parametresi iin DMF miktarı (0.0 ve 0.4), g l k parametresi iin DMF miktarı (0.0 ve 0.7),  rneklem b y kl ė  ($N_R = N_O = 1000$; $N_R = 2N_O = 1000$) ve yetenek daėılımları [$N(0, 1)$, $N(-0.5, 1)$, $N(0.5, 1)$] olacak şekilde farklı koştullar sim le edilmiştir. Belirlenen DMF miktarları gruplara eklenip ıkarılarak farklı grupların avantajlı/dezavantajlı olması saėlanmıştır. Grup sayılarına g re grupların  rneklem b y kl klerinin eşt olmadığı durumlar iin [$(N_R = 1000, N_{O1, O2} = 500)$ ve $(N_R = 1000, N_{O1, O2} = 500, N_{O3, O4} = 1000)$] farklı  rneklem b y kl kleri belirlenmiştir. SAS programı kullanılarak gerekleştiren sim lasyon alıřmasında 100 tekrar yapılmıştır. Elde edilen sonular hem g l k (0.7) hem de ayırt edicilik (0.4) parametresinden kaynaklı DMF olduėunda istatistiksel g  oranının y kselerek 1.00’a yaklaştıėını g stermiştir. Farklı DMF t rlerinin dikkate alındıėı b t n koştullar iin toplam  rneklem b y kl ė  arttıka istatistiksel g  oranları da artmıştır. Maddelerin DMF iermediėi koştul iin Tip I hata oranının .07 deėerinin altında, DMF ierdiėi b t n koştullar iin ise Tip I hata oranının .13 deėerinin altında olduėu sonucuna ulaşılmıştır. Bununla birlikte ayırt edicilik parametresinden kaynaklı DMF’nin olduėu koştullarda MG-NCDIF tekniğinin performansının grup sayısından etkilendiėi g r l rken, diėer koştullarda bu tekniğın performansının grup sayısından ciddi bir şekilde etkilenmediėi belirtilmiştir. Elde edilen sonular MG-NCDIF tekniğinin, DMF’nin oklu gruplarda eşt zamanlı olarak belirlenmesi amacıyla kullanılabileceėini g stermiştir.

Tutz ve Schauburger (2015) yaptıkları sim lasyon alıřmasında oklu gruplar iin Lojistik Regresyon Lasso (LRL) tekniğini geliştirmiş ve LRL, GLK, GLR ve GMH tekniklerinin performanslarını Tip I hata ve istatistiksel g  oranları aısından karşılaştırmışlardır. Grupların  rneklem b y kl klerinin 500, testteki madde sayısının ise

20 olduğu bu araştırmada; grup sayısı (5 ve 6), DMF'li maddelerin sayısı (4 ve 8), DMF'li maddelerin yüzdesi (%20 ve %40) ve TB DMF miktarı (düşük, orta, yüksek) olmak üzere çeşitli koşullar dikkate alınarak her bir koşul için 100 tekrar yapılmıştır. Elde edilen sonuçlar DMF miktarı arttığında LRL, GLK, GLR ve GMH tekniklerinin tamamı için hem istatistiksel güç oranlarının hem de Tip I hata oranlarının yükseldiğini göstermiştir. Ayrıca diğer koşulların sabit kalması şartıyla DMF'li madde yüzdesi arttığında bütün tekniklerin istatistiksel güç ve Tip I hata oranları yükselmiştir. Yalnızca LRL tekniği için DMF miktarının düşük olduğu durumda grup sayısı arttıkça istatistiksel güç oranının da yükseldiği sonucuna ulaşılmıştır. DMF miktarı ve grup sayısı etkileşimi en fazla GLK, en az ise LRL tekniği için hesaplanan Tip I hata oranlarında artışa sebep olmuştur. Diğer taraftan DMF miktarı ve DMF'li madde yüzdesi etkileşimi ise en az GLK, en fazla GLR ve GMH tekniklerinin Tip I hata oranlarının artmasına neden olmuştur. DMF miktarının orta ve yüksek olduğu durumlarda LRL tekniğinin diğer tekniklerle benzer istatistiksel güç oranı ve daha düşük Tip I hata oranına sahip olduğu ifade edilmiştir. Buna rağmen DMF miktarının düşük olduğu durum için LRL tekniğinin istatistiksel güç ve Tip I hata oranlarının diğer tekniklerden daha düşük olduğu belirtilmiştir. Sonuç olarak özellikle grup sayısının az ve DMF miktarının düşük olduğu durumlarda GLK, GLR ve GMH tekniklerinin tercih edilmesi gerektiği vurgulanmıştır.

Finch (2016) tarafından yapılan simülasyon çalışmasında yalnızca TB DMF'li maddeleri belirlemek amacıyla çoklu gruplar için GLR, GMH, GLK ve Bayes alignment teknikleri dikkate alınmış ve tekniklerin performansları Tip I hata ve istatistiksel güç oranları açısından karşılaştırılmıştır. Değişimlenen faktörlerin ana etki ve etkileşim etkilerinin Tip I hata ve istatistiksel güç oranları açısından belirlenmesi amacıyla ANOVA yapılmıştır. Araştırma kapsamında 2PLM'ye dayalı üretilen veri setindeki 20 madde için ayırt edicilik parametresinin 0.84 ile 1.53, güçlük parametresinin ise -1.06 ile 1.39 aralığında olduğu dağılımlar dikkate alınmıştır. GLR, GMH ve GLK teknikleri için analiz sürecinde difR paketi kullanılarak 1000 tekrar yapılırken, Bayes alignment tekniği için Mplus 7.11 programı kullanılarak 10.000 tekrarda analizler gerçekleştirilmiştir. TB DMF belirleme tekniklerinin performanslarının karşılaştırıldığı araştırmada; grup sayısı (2, 3 ve 6), referans grubun örneklem büyüklüğü (500, 1000 ve 2000), odak grupların örneklem büyüklüklerinin referans gruba oranı ($N_R = N_O$, $N_R = 2N_O$), TB DMF miktarı (0, 0.4, 0.6 ve 0.8) ve yetenek düzeyi ($\theta_R = \theta_O$; $\theta_R > \theta_O$) olacak şekilde farklı koşullar simüle edilmiştir. Ayrıca odak grupların örneklem büyüklüğünün her durumda birbirine eşit olması sağlanmış ve yalnızca bir maddenin DMF'li olduğu kabul edilmiştir. Elde

edilen sonuçlar grup sayısının altı ve referans ile odak grupların örneklem büyüklüklerinin farklı olduğu koşullarda GLR tekniğinin Tip I hata oranlarının çok yüksek olduğunu göstermiştir. GLK tekniği için ise genel olarak bütün koşullarda Tip I hata oranının .05 değerinin altında olduğu ifade edilmiştir. Araştırmada referans ile odak grupların örneklem büyüklüklerinin eşit olmadığı durumda grup sayısı arttıkça GLR, GMH, GLK ve Bayes alignment teknikleri için Tip I hata oranlarının arttığı bulunmuştur. Bununla birlikte tekniklerin tümü için örneklem büyüklüğü ve DMF miktarı arttıkça istatistiksel güç oranlarının da yükseldiği görülmüştür. Referans grubun örneklem büyüklüğünün 2000, DMF miktarının ise 0.80 olduğu durumun haricinde, GLK için hesaplanan istatistiksel güç oranının diğer tekniklere göre daha düşük olduğu ifade edilmiştir. Ayrıca GMH ve Bayes alignment'ın çoklu gruplarda TB DMF'yi belirlemede etkili teknikler oldukları belirlenmesine rağmen grup sayısı iki veya üç olduğunda GLR tekniğinin de belirtilen tekniklere alternatif olabileceği sonucuna ulaşılmıştır.

Schauberger ve Tutz (2016) yaptıkları simülasyon çalışmasında çoklu gruplar için boosting (yükseltme) yaklaşımını ilk defa kullanmış ve DMFboost, GLK, GLR ve GMH tekniklerinin performanslarını Tip I hata ve istatistiksel güç oranları açısından karşılaştırmıştır. TB DMF'yi dikkate alan bu çalışmada yetenek ve madde güçlük parametreleri standart normal dağılımdan üretilmiştir. Grupların örneklem büyüklüklerinin 500, testteki madde sayısının ise 20 olduğu araştırma kapsamında; grup sayısı (2 ve 5), DMF'li maddelerin sayısı (0, 4 ve 8), DMF'li maddelerin yüzdesi (%0, %20 ve %40) ve TB DMF miktarı (düşük, orta ve yüksek) olmak üzere çeşitli koşullar dikkate alınarak her bir koşul için 100 tekrar yapılmıştır. Çalışmada Tip I hata DMF'li olmayan maddeler üzerinden hesaplanmıştır. Elde edilen sonuçlar genel olarak GLR ve GMH tekniklerinin performanslarının birbirlerine oldukça benzer olduğunu ortaya koymuştur. Grup sayısı arttığında Tip I hata oranlarının nispeten yükseldiğini ve hem iki hem de beş grup için DMF olmayan durumda DMFboost, GLK, GLR ve GMH tekniklerinin Tip I hata oranlarının .05 kabul edilebilir değerin altında olduğunu göstermiştir. Ayrıca grup sayısı ve TB DMF miktarı arttığında bütün teknikler için hesaplanan istatistiksel güç oranlarının yükseldiği açıklanmıştır. Diğer koşulların sabit kalması şartıyla DMF'li madde yüzdesi arttığında istatistiksel güç ve Tip I hata oranları; GLR ve GMH teknikleri için yükselmiş, GLK tekniği için genel olarak sabit kalmış ve DMFboost yaklaşımı için ise düşmüştür. Böylece çoklu gruplarda DMFboost yaklaşımının diğer tekniklerle benzer istatistiksel güç oranına ve nispeten daha düşük Tip I hata oranına sahip olduğu belirtilmiştir. Sonuç olarak çoklu gruplar için bütün

tekniklerin oldukça benzer performans göstermeleri nedeniyle sonraki çalışmalarda DMFboost yaklaşımının da alternatif olarak tercih edilebileceği ifade edilmiştir.

Tutz ve Berger (2016) tarafından yapılan simülasyon çalışmasında çoklu gruplar için madde odaklı Rasch Ağacı (item-focussed trees) tekniği ilk defa kullanmış ve madde odaklı Rasch Ağacı ile GLK, GLR ve GMH tekniklerinin performansları Tip I hata ve istatistiksel güç oranları açısından karşılaştırılmıştır. Çalışmada grupların örneklem büyüklükleri 500, testteki madde sayısı ise 20 olarak belirlenmiş ve TB DMF dikkate alınmıştır. Yetenek ve madde güçlük parametrelerinin standart normal dağılımdan üretildiği araştırma kapsamında; grup sayısı (2 ve 5), DMF'li maddelerin sayısı (0 ve 4), DMF'li maddelerin yüzdesi (%0 ve %20) ve TB DMF miktarı (düşük, orta ve yüksek) olmak üzere çeşitli koşullar dikkate alınarak her bir koşul için 100 tekrar yapılmıştır. Madde odaklı Rasch Ağacı tekniğinde analizler her bir permütasyon testi 100 permütasyona dayalı olarak yapılmıştır. Elde edilen sonuçlara göre TB DMF miktarı arttıkça madde odaklı Rasch Ağacı, GLK, GLR ve GMH teknikleri için elde edilen istatistiksel güç oranlarının yükseldiği görülmüştür. Hem iki hem de beş grup için DMF olmayan durumda madde odaklı Rasch Ağacı, GLK, GLR ve GMH tekniklerinin Tip I hata oranlarının .06 değerinin altında olduğu sonucuna ulaşılmıştır. Ayrıca GLR ve GMH tekniklerinin genel olarak birbirlerine yakın performans ortaya koydukları ifade edilmiştir. Bütün koşullarda GLR ve GMH tekniklerinin GLK tekniğine göre daha yüksek Tip I hata ve istatistiksel güce sahip olduğu belirtilmiştir. Beş grupta madde odaklı Rasch Ağacı tekniği için hesaplanan istatistiksel güç oranının, GLK, GLR ve GMH tekniklerine göre nispeten daha yüksek ve Tip I hata oranının da GLR ve GMH tekniklerinden daha düşük olması, sonraki çalışmalarda madde odaklı Rasch Ağacı tekniğinin de kullanılabileceğini göstermiştir.

Uyar ve Kaya Uyanık (2016) tarafından yapılan bir araştırmada PISA 2012 matematik ve fen okuryazarlığı alt testlerinin ikili puanlanan maddeleri, OECD ortalamasının üstünde (Finlandiya, Hong Kong) ve altında (Türkiye, Rusya) olan dört ülke arasında DMF gösteren madde olup olmadığı incelenmiştir. Öğrenci ve madde sayılarının fazla olması nedeniyle yedinci kitapçığın seçildiği araştırmada, her bir ülkeden matematik okuryazarlığı alt testinde 326 öğrenci, fen okuryazarlığı alt testinde ise 372 öğrenci verisi rastgele belirlenmiştir. Araştırma kapsamında Finlandiya referans grup diğer ülkeler ise odak gruplar olarak dikkate alınmıştır. Matematik okuryazarlığı alt testinde ikili puanlanan 28 maddede, fen okuryazarlığı alt testinde ise 34 maddede TB DMF belirlemek amacıyla GMH ve GLR teknikleri kullanılmış ve bu teknikler tarafından

DMF gösteren maddeler için ülkeler arasında altı farklı ikişerli karşılaştırma yapılmıştır. Elde edilen bulgulara göre matematik ve fen okuryazarlığı alt testleri için maddelerin çoğunluğunun GMH ve GLR teknikleri tarafından DMF’li olarak tespit edildiği sonucuna ulaşılmıştır. Matematik ve fen okuryazarlığı alt testlerinde GMH ve GLR teknikleri ile DMF’li olarak belirlenen maddeler arasında sırasıyla %89.3 ve %100 uyum belirlenmiştir. İkili karşılaştırma sonuçlarına göre; GMH ve GLR teknikleri tarafından DMF’li olarak belirlenemeyen maddeler MH ve LR teknikleri ile DMF’li olarak tespit edilmiş ve bu bulgu ikili karşılaştırma yaklaşımında Tip I hata yapma olasılığının arttığına ilişkin alanyazında var olan bilgiyi doğrulamıştır. MH ve LR teknikleri karşılaştırıldığında MH tekniğinin daha fazla sayıda DMF’li madde tespit ettiği ve analiz sonuçlarına göre orta ile yüksek düzeyde DMF içeren madde sayısının oldukça fazla olduğu belirtilmiştir. Sonuç olarak dikkate alınan ülkeler arasında göreceli olarak kültürel olarak farklılıklar arttıkça, DMF’li olarak işaretlenen madde sayısının da arttığı ifade edilmiştir.

Rollins III’in (2018) yaptığı simülasyon çalışmasında grup sayısının dört olduğu durum için gözlenen puan yaklaşımlarından GMH, GLR ve LRL tekniklerinin performanslarını karşılaştırmıştır. 3PLM’ye dayalı üretilen veri setindeki 40 madde için ayırt edicilik parametresi ortalaması 1 ve standart sapması 0.1225 olan lognormal dağılımdan, şans parametresi ise .20 olacak şekilde üretilmiştir. Güçlük parametresi ise Test Bilgi Fonksiyonunun (TBF) dağılımına göre ortalama yetenek, 0.00 ve -1.15 olacak şekilde değişimlenmiştir. Toplam örneklem büyüklüğünün 2000 olduğu ve yalnızca güçlük parametresinden kaynaklı DMF’nin incelendiği araştırmada; örneklem büyüklüğü [“dengeli” (500, 500, 500, 500), “dengesiz” (800, 600, 400, 200)], DMF’li maddelerin yüzdesi (%0, %5, %10), TBF’nin şekli [“dar” (s.s = 0.5), “yaygın” (s.s = 1.5)], TBF’nin özelliği ($TBF_{max-\theta} \approx 0$; $TBF_{max-\theta} \approx -1.15$), grupların yetenek dağılımları [“eşit” (0.00, 0.00, 0.00, 0.00), “yarım logit” (0.00, -0.17, -0.33, -0.50)] ve DMF’nin konumu (TBF’den uzak, TBF’den yakın) olacak şekilde farklı koşullar simüle edilmiştir. R yazılımı kullanılarak gerçekleştirilen simülasyon çalışmasında 250 tekrar yapılmıştır. Elde edilen bulgulara göre DMF’li maddelerin %0 olduğu koşul için Tip I hata oranlarının .052 değerinin altında, %5 (2 madde) ve %10 (4 madde) olduğu koşullar için ise Tip I hata oranlarının .07 değerinin altında olduğu sonucuna ulaşılmıştır. Grupların yetenek parametrelerinin farklılaşması, LRL tekniği için hesaplanan Tip I hata oranının önemli miktarda artmasına neden olurken GMH ve GLR teknikleri için hesaplanan Tip I hata oranlarında ciddi bir farklılaşma gözlemlenmemiştir. TBF’nin sağa çarpık olduğu

durumda grupların yetenek parametrelerinin farklılaşması hem GLR hem de LRL teknikleri için hesaplanan Tip I hata oranlarının artmasına neden olmuştur. Grupların yetenek parametrelerinin ve örneklem büyüklüğünün eşit olduğu durumda GMH ve GLR teknikleri için hesaplanan istatistiksel güç oranının LRL tekniğinden yüksek olduğu ve testteki DMF'li madde sayısı arttıkça tekniklerin istatistiksel gücünün azaldığı görülmüştür. Grupların yetenek parametreleri farklılaştığında ise GMH ve GLR tekniklerinin istatistiksel güç oranlarında azalma söz konusu olmazken, LRL tekniği için hesaplanan istatistiksel güç oranlarında azalma olduğu belirtilmiştir. Genel olarak sonuçlar grupların örneklem büyüklüklerinin farklılaşmasından ziyade yetenek parametrelerinin farklılaşmasının DMF belirleme tekniklerinin performanslarını daha fazla etkilediğini göstermiştir. Ayrıca GMH ve GLR tekniklerinin istatistiksel güç ve Tip I hata açısından benzer performans gösterdiği belirtilmesine rağmen, LRL tekniğinin istatistiksel gücünün düşük, Tip I hatasının ise yüksek hesaplanması sonucunda performansının diğer tekniklerden farklılaştığı ifade edilmiştir.

Battauz (2019) tarafından yapılan çalışmada Lord ki-kare Wald tekniğinin kuramsal özelliklerinde iki farklı iyileştirmeye gidilerek BGÇGW tekniği geliştirilmiş ve GLK ile BGÇGW tekniklerinin performansları Tip I hata ve istatistiksel güç oranları açısından karşılaştırılmıştır. Hem TB hem de TBO(α) DMF belirleme tekniklerinin performanslarının karşılaştırıldığı araştırmada; grup sayısı (2 ve 3), parametre kestirim modeli (1PLM, 2PLM ve 3PLM), her grup için örneklem büyüklüğü (500, 1000, 2000, 4000 ve 8000), test uzunluğu (20 madde ve 40 madde), grupların yetenek dağılımları [$N(0, 1)$, $N(0.5, 1.2)$, $N(-0.5, 1.2)$], DMF'li maddelerin yüzdesi (%0, %5, %20), TB DMF miktarı (0.4 ve 0.6), TBO(α) DMF miktarı (0.3 ve 0.5) olacak şekilde farklı koşullar simüle edilmiştir. Veriler ayırt edicilik parametresi 0.7 ile 1.3 aralığında tekdüze dağılımdan, güçlük parametresi standart normal dağılımdan üretilmiştir. Şans parametresi ise bütün maddeler için .20 olarak belirlenmiştir. DMF'li maddelerin %0 olduğu durumda üretilen veriler kullanılarak Tip I hata oranları hesaplanmıştır. Elde edilen sonuçlar genel olarak 1PLM ve 2PLM'de hem GLK hem de BGÇGW teknikleri için hesaplanan Tip I hata oranlarının nominal manidarlık düzeyine (.05) eşit veya düşük olduğunu göstermiştir. Her iki MTK modeline göre de bütün koşullarda GLK tekniğinin Tip I hata oranları BGÇGW tekniğine göre düşük bulunmuştur. 1PLM'de test uzunluğu ve örneklem büyüklüğü arttığında GLK ile BGÇGW teknikleri için hesaplanan Tip I hata oranlarında ciddi bir farklılaşma olmadığı, grup sayısı arttığında ise tekniklerin Tip I hata oranlarında bir miktar azalma olduğu gözlemlenmiştir. 2PLM için ise test uzunluğu ve

örneklem büyüklüğünün yanı sıra grup sayısı arttığında da hesaplanan Tip I hata oranlarında önemli bir farklılaşma olmadığı belirtilmiştir. 3PLM dikkate alındığında test uzunluğu, örneklem büyüklüğü ve grup sayısı arttığında tekniklerin Tip I hata oranları yükselmiştir. Diğer koşullar sabit kaldığında üç grup için her iki tekniğin Tip I hata oranları .06 ile .50 arasında değişiklik göstermiştir. 3PLM’de genel olarak örneklem büyüklüğü arttıkça GLK tekniği için Tip I hata oranı azalırken, BGÇGW tekniği için Tip I hata oranında artış gözlemlenmiştir. Diğer koşullar sabit kalmak şartıyla test uzunluğu arttıkça GLK tekniği için Tip I hata oranı azalmıştır. Ayrıca genel olarak bütün koşullarda 3PLM için BGÇGW tekniğinin Tip I hata oranları GLK tekniğine göre düşük bulunmuştur. İstatistiksel güç oranlarına göre tekniklerin 2PLM’de daha yüksek performans gösterdikleri sonucuna ulaşılmıştır. Bütün MTK modellerinde hem GLK hem de BGÇGW teknikleri için gruptaki örneklem büyüklüğü arttıkça istatistiksel güç oranları yükselerek 1.00’a yaklaşmıştır. Benzer şekilde her üç model için küçük örneklemelerde grup sayısı arttıkça istatistiksel güç oranının arttığı gözlemlenmiştir. 1PLM ve 2PLM’de tekniklerin istatistiksel güç oranları birbirlerine oldukça yakın iken, 3PLM için GLK tekniğinin performansının önemli bir şekilde yüksek olduğu bulunmuştur. Özellikle 3PLM’de BGÇGW tekniği için hesaplanan Tip I hata oranlarının GLK’ya göre nispeten daha düşük olmasına rağmen, istatistiksel güç oranlarının genel olarak kabul edilebilir sınırın altında olması ilginç bir sonuç olarak değerlendirilmiştir.

Morales-Ortiz ve Peña-Cardenas (2019) tarafından yapılan araştırmada Amerika Birleşik Devletleri’nde yaşayan İspanyol, Meksikalı ve Hispanik (Amerika Birleşik Devletleri’nde ikamet eden Meksikalılar) bireyler olmak üzere üç farklı grup için Sözel Seçici Hatırlatma Testinin (“VSRT”, Verbal Selective Reminding Test) İspanyolca versiyonunun birinci formunda yer alan maddelerin DMF açısından incelenmesi amaçlanmıştır. Çalışmada, farklı gruplar arasında hiçbir DMF’li madde olmadığına ilişkin hipotez test edilmiştir. Grupların örneklem büyüklükleri hem alt gruplar ($N_{İspanyol} = 211$, $N_{Meksikalı} = 201$, $N_{Hispanik} = 205$) açısından hem de yaş, eğitim düzeyi ve cinsiyet açısından dengelidir. Uygun örnekleme yaklaşımının tercih edildiği bu araştırma kapsamında, İspanyol bireyler referans grubu Meksikalı ve Hispanik bireyler ise odak grupları oluşturmuştur. On iki maddeden oluşan VSRT’nin altı deneme uygulamasından elde edilen veriler kullanılarak, GMH ve GLR (TB ve TBO) teknikleri ile DMF belirleme çalışması gerçekleştirilmiştir. Elde edilen bulgulara göre; beş maddenin GMH, altı maddenin ise GLR tarafından TB DMF’li olarak belirlendiği sonucuna ulaşılmıştır. TBO DMF için ise GLR tekniği kullanılarak yapılan analizler sonucunda hiçbir maddenin

DMF’li olmadığı ifade edilmiştir. GMH ve GLR teknikleri için elde edilen sonuçların yalnızca bir maddede farklılaşması, bu tekniklerin genel olarak benzer sonuçlar gösterdiğini ortaya koymuştur. Yalnızca MH tekniği ile gruplar arasında ikili karşılaştırma yapılmış ve İspanyol ile Meksikalı bireyler arasında bir ve İspanyol ile Hispanik bireyler arasında iki maddenin DMF’li olarak tespit edildiği belirtilmiştir. ‘Pollo’ isimli maddenin hem İspanyol ile Meksikalı hem de İspanyol ile Hispanik bireyler için DMF’li olarak belirlenmesi önemli bulunmuştur. Sonuç olarak hem çoklu grupların eş zamanlı olarak hem de ikili olarak karşılaştırılması durumunda, GMH ve GLR tekniklerinin genellikle tutarlı sonuçlar ortaya koydukları ifade edilmiştir.

Dell-Ross’un (2021) çalışmasında GMH, MG-NCDIF ile MIMIC tekniklerinin performansları Tip I hata ve istatistiksel güç oranları açısından karşılaştırılmıştır. Farklı özelliklerdeki DMF belirleme tekniklerinin performanslarının karşılaştırıldığı araştırmada; grup sayısı (3 ve 5), grup sayısına göre DMF’li maddelerden etkilenen odak grup sayısı [3 Grupta (1 ve 2), 5 Grupta (1, 2 ve 4)], örneklem büyüklüğü ($N_R = N_O = 1000$; $N_R = 2N_O = 1000$), DMF türü [TB, TBO(*a*), TBO(*ab*)], DMF örüntüsü [(0/0/0), (0/-0.4/0), (0/-0.4/0.4), (0/0.7/0), (0/0.7/-0.7), (0/-0.4/0), (0/0.7/0), (0/-0.4/0.4) ve (0/0.7/-0.7)], DMF miktarı (0.4 ve 0.7), DMF’li maddelerin yüzdesi (%0, %20) ve grupların yetenek dağılımları [$N(0, 1)$, $N(0.5, 1)$, $N(-0.5, 1)$] simüle edilmiştir. DMF miktarları gruplara eklenip çıkarılarak farklı grupların avantajlı/dezavantajlı olması sağlanmıştır. Grup sayılarına göre grupların örneklem büyüklüklerinin eşit olmadığı durumlar için [$(N_R = 1000, N_{O1, O2} = 500)$ ve $(N_R = 1000, N_{O1, O2} = 500, N_{O3, O4} = 1000)$] farklı örneklem büyüklükleri belirlenmiştir. Madde sayısının 30 olduğu araştırmada, ayırt edicilik parametresi 0.49 ile 1.74 arasında, güçlük parametresi -1.16 ile 1.96 arasında ve şans parametresi ise 0.05 ile 0.22 arasında olmak üzere veriler 3PLM’ye dayalı olarak üretilmiş ve 100 tekrar yapılmıştır. Elde edilen sonuçlar Tip I hata oranlarının GMH ile MIMIC teknikleri için nominal manidarlık düzeyinden (.05) düşük olduğunu, MG-NCDIF tekniği için ise liberal kritere göre kabul edilebilir düzeyde olduğunu göstermiştir. Ayrıca grup sayısı ve toplam örneklem büyüklüğü arttığında GMH ile MIMIC teknikleri için hesaplanan Tip I hata oranlarında ciddi bir farklılaşma olmadığı, MG-NCDIF tekniğinde ise Tip I hata oranlarının yükseldiği gözlemlenmiştir. Maddelerin DMF içerdiği koşullarda örneklem büyüklüğü ve DMF miktarı arttıkça GMH ile MIMIC teknikleri için hesaplanan Tip I hata oranlarının genel olarak yükseldiği, MG-NCDIF tekniği için ise Tip I hata oranlarının örneklem büyüklüğünden etkilenerek yükselmesine rağmen elde edilen değerlerin DMF türü ve DMF miktarına göre ciddi bir şekilde

farklılaşmadığı saptanmıştır. Öte yandan diğer koşullar sabit tutulduğunda örneklem büyüklüğü ve/veya DMF miktarı arttıkça GMH, MG-NCDIF ve MIMIC teknikleri için hesaplanan istatistiksel güç oranlarının yükselme eğiliminde olduğu anlaşılmıştır. TB DMF türünde GMH tekniğine ilişkin performansın diğerlerinden yüksek olduğu, TBO(a) ve TBO(ab) DMF türlerinde ise en yüksek istatistiksel güce sahip olan tekniğin MIMIC olduğu ortaya konulmuş ve DMF miktarı ile örneklem büyüklüğü arttıkça istatistiksel güç oranlarının yükselerek 1.00'a yaklaştığı tespit edilmiştir. GMH ve MG-NCDIF teknikleri için istatistiksel güç oranlarının ise TBO(a) DMF türünde düşük iken TBO(ab) DMF türünde değerlerin yükseldiği belirlenmiştir. Elde edilen bulgular sonucunda; GMH, MG-NCDIF ve MIMIC tekniklerine ilişkin performansların çeşitli koşullar altında farklılaşabildiği saptanmıştır.

Çoklu grup DMF belirleme tekniklerinin performanslarının araştırıldığı simülasyon çalışmaları detaylı bir şekilde incelenerek, alanyazında en yaygın dikkate alınan faktör ve koşullar tespit edilmiştir. Bu anlamda çoklu grup tekniklerinin kullanıldığı simülasyon çalışmalarında dikkate alınan koşullar aşağıda özetlenmiştir.

- Penfield (2001) tarafından yapılmış olan ilk simülasyon çalışmasında, grup sayısı (2, 3, 4, 5), DMF'li maddelerin etkilediği odak grup sayısı (1, 2, 3, 4), örneklem büyüklüğü (250, 500, 1000), DMF miktarı ($b_R - b_O = 0.0$, $b_R - b_O = 0.4$), incelenen madde haricinde DMF içeren madde sayısı (2, 5, 10) ve yetenek dağılımı [$N(0, 1)$, $N(-1, 1)$] olmak üzere altı faktör için MH, BMH ve GMH tekniklerinin performansları Tip I hata ve istatistiksel güç oranları açısından karşılaştırılmıştır.
- Fidalgo ve Scalón'un (2010) çalışmalarında grup sayısı (3, 4), DMF'li madde yüzdesi (%0, %22) ve DMF'nin örüntüsü (DMF yalnızca bir grupta, DMF gruplara dengeli dağılmış) koşullarında MH, BMH ve GMH teknikleri için Tip I hata ve istatistiksel güç oranları hesaplanmıştır.
- Svetina ve Rutkowski (2014) araştırmalarında grup sayısı (10, 20), DMF'den etkilenen grupların yüzdesi (%40, %70), DMF'li maddelerin yüzdesi (%20, %40, %60), DMF türü [TB, TBO(a)] ve DMF miktarı [$(b_R - b_O = 0.5, a_R - a_O = 0.4)$, $(b_R - b_O = 1.0, a_R - a_O = 0.8)$] koşulları için GLR tekniğinin performansını Tip I hata ve istatistiksel güç oranları açısından incelemiştir.
- Carroll (2015) çalışmasında grup sayısı (3 ve 6), çapa maddeler haricindeki DMF'li maddelerin sayısı (1, 2 ve 6) ve doğru belirlenen çapa maddelerin sayısı (5, 8 ve 10) koşulları için LGÇGW tekniğinin performansını Tip I hata ve istatistiksel güç oranları açısından araştırmıştır.

- Jin ve diğerleri (2015) grup sayısının dört olduğu durum için yaptıkları simülasyon çalışmalarında DMF miktarı $[(b_R - b_O = 0.0, a_R - a_O = 0.0), (b_R - b_O = 0.2, a_R - a_O = 0.13), (b_R - b_O = 0.4, a_R - a_O = 0.26), (b_R - b_O = 0.6, a_R - a_O = 0.39)]$, DMF'li maddelerin yüzdesi (%0, %10, %20) ve DMF kaynağı (Faktör A, Faktör B, AB etkileşim) koşullarında GLR tekniğinin performansını Tip I hata ve istatistiksel güç oranları açısından incelemiştir.
- Oshima ve diğerleri (2015) araştırmalarında grup sayısı (3 ve 5), grup sayısına göre DMF'li maddelerden etkilenen odak grup sayısı (1 ve 2; 1, 2 ve 4), DMF türü [TB, TBO(a), TBO(ab)], DMF miktarı $[(b_R - b_O = 0.0, a_R - a_O = 0.0), (b_R - b_O = 0.7, a_R - a_O = 0.4)]$, örneklem büyüklüğü ($N_R = N_O = 1000$; $N_R = 2N_O = 1000$) ve yetenek dağılımı $[N(0, 1), N(-0.5, 1), N(0.5, 1)]$ koşulları için MG-NCDIF tekniğinin performansını Tip I hata ve istatistiksel güç oranları açısından araştırmıştır.
- Tutz ve Schauburger (2015) çalışmalarında, grup sayısı (5 ve 6), DMF'li maddelerin sayısı (4 ve 8), DMF'li maddelerin yüzdesi (%20 ve %40) ve TB DMF miktarı (düşük, orta, yüksek) koşullarında LRL, GLK, GLR ve GMH tekniklerinin performanslarını Tip I hata ve istatistiksel güç oranları açısından karşılaştırmıştır.
- Finch (2016) yapmış olduğu araştırmada, grup sayısı (2, 3, 6), referans grubun örneklem büyüklüğü (500, 1000, 2000), referans ve odak grupların örneklem büyüklüğü oranı ($N_R = N_O, N_R = 2N_O$), TB DMF miktarı $[(b_R - b_O = 0.0, b_R - b_O = 0.4, b_R - b_O = 0.6, b_R - b_O = 0.8)]$ ve yetenek düzeyi ($\theta_R = \theta_O; \theta_R > \theta_O$) olmak üzere beş faktör için GLR, GLK, GMH ve Bayes tekniklerinin performanslarını karşılaştırmıştır.
- Schauburger ve Tutz (2016) araştırmalarında, grup sayısı (2 ve 5), DMF'li maddelerin sayısı (0, 4 ve 8), DMF'li maddelerin yüzdesi (%0, %20 ve %40) ve TB DMF miktarı (düşük, orta, yüksek, çok yüksek) koşullarında DMFboost yaklaşımı ile GLK, GLR ve GMH tekniklerinin performanslarını Tip I hata ve istatistiksel güç oranları açısından karşılaştırmıştır.
- Tutz ve Berger (2016) tarafından yapılan çalışmada, grup sayısı (2 ve 5), DMF'li maddelerin sayısı (0 ve 4), DMF'li maddelerin yüzdesi (%0 ve %20) ve TB DMF miktarı (düşük, orta ve yüksek) olmak üzere çeşitli faktörler için madde odaklı Rasch Ağacı, GLK, GLR ve GMH tekniklerinin performansları Tip I hata ve istatistiksel güç oranları açısından karşılaştırılmıştır.

- Rollins III (2018) grup sayısının dört olduğu durum için yapmış olduğu çalışmada, örneklem büyüklüğü [“dengeli” (500, 500, 500, 500), “dengesiz” (800, 600, 400, 200)], DMF’li maddelerin yüzdesi (%0, %5, %10), Test Bilgi Fonksiyonunun (TBF) şekli [“dar” (s.s = 0.5), “yaygın” (s.s = 1.5)], Test Bilgi Fonksiyonunun özelliği (TBFmax- $\theta \approx 0$; TBFmax- $\theta \approx -1.15$), grupların yetenek dağılımları [“eşit” (0.00, 0.00, 0.00, 0.00), “yarım logit” (0.00, -0.17, -0.33, -0.50)] ve DMF’nin konumu (TBF’den uzak, TBF’den yakın) olmak üzere toplam altı faktör için GLR, GMH ve LRL tekniklerinin performanslarını karşılaştırmıştır.
- Battauz (2019) tarafından yapılan çalışmada grup sayısı (2 ve 3), parametre kestirim modeli (1PLM, 2PLM ve 3PLM), her grup için örneklem büyüklüğü (500, 1000, 2000, 4000 ve 8000), test uzunluğu (20 madde ve 40 madde), grupların yetenek dağılımları ($N(0, 1)$, $N(0.5, 1.2)$, $N(-0.5, 1.2)$), DMF’li maddelerin yüzdesi (%0, %5, %20), TB DMF miktarı (0.4 ve 0.6) ve TBO(a) DMF miktarı (0.3 ve 0.5) olmak üzere çeşitli faktörler için GLK ile BGÇGW tekniklerinin performansları incelenmiştir.
- Dell-Ross’un (2021) çalışmasında ise grup sayısı (3 ve 5), DMF’li maddelerden etkilenen odak grup sayısı [3 Grupta (1 ve 2), 5 Grupta (1, 2 ve 4)], DMF türü [TB, TBO(a), TBO(ab)], DMF örüntüsü [(0/0/0), (0/-0.4/0), (0/-0.4/0.4), (0/0.7/0), (0/0.7/-0.7), (0/-0.4/0), (0/0.7/0), (0/-0.4/0.4) ve (0/0.7/-0.7)], örneklem büyüklüğü ($N_R = N_O = 1000$; $N_R = 2N_O = 1000$), DMF miktarı (0.4 ve 0.7), DMF yüzdesi (%0, %20) ve yetenek dağılımları [$N(0, 1)$, $N(0.5, 1)$, $N(-0.5, 1)$] olmak üzere çeşitli faktörler için GMH, MG-NCDIF ile MIMIC tekniklerinin performansları Tip I hata ve istatistiksel güç oranları açısından karşılaştırılmıştır.

Çoklu gruplarda ikili puanlanan maddeler için geliştirilen tekniklerin dikkate alındığı bu simülasyon çalışmalarında; farklı çoklu grup DMF belirleme tekniklerinin çeşitli koşullar altındaki performansları incelenmesine rağmen genellikle benzer faktörlerin dikkate alındığı, GLK tekniğinin performansının sınırlı sayıda araştırmada incelendiği, Lordif tekniğinin performansının ise bu çalışmaların hiçbirinde araştırılmadığı görülmüştür. Üstelik belirtilen çalışmalarda genellikle TB ve TBO(a) DMF türlerinin ele alındığı, TBO(ab) DMF türünün sınırlı sayıda araştırmada incelendiği, Tip III hata oranlarının ise hiçbir araştırmada hesaplanmadığı tespit edilmiştir.

BÖLÜM 3

YÖNTEM

Bu bölümde araştırmanın modeli, simülasyon deseni, verilerin üretilmesi, verilerin analizi ve simülasyonun geçerliği başlıkları ele alınmıştır.

Araştırmanın Modeli

Araştırma kapsamında çoklu grup DMF belirleme tekniklerinin hangi koşullar altında nasıl sonuçlar ürettiğinin belirlenmesi amacıyla simülasyon çalışması gerçekleştirilmiştir. Simülasyon çalışmaları, rastgele veriler oluşturmak için belirli parametrelerin kullanıldığı bilgisayar destekli araştırmalardır (Mooney, 1997; Morris vd., 2019; Rubinstein ve Kroese, 2017). Gerçek veriye ulaşmanın zor veya fazla zaman gerektirdiği, maliyetli olduğu, madde ve birey düzeyindeki potansiyel karıştırıcıların etkilerinin ortadan kaldırılamadığı, gerçek uygulamalarda dışsal etkilerden arınık bir ortam sağlanamadığı ve gerçek verilerin genellikle kayıp veri içerdiği gibi durumlar (Bulut ve Sünbül, 2017; Hallgren, 2013; Harwell, Kohli ve Peralta, 2017) nedeniyle farklı koşulların eş zamanlı olarak değişimlenebildiği simülasyon çalışmaları giderek önem kazanmıştır (Feinberg ve Rubright, 2016; Harwell, Stone, Hsu ve Kirisci, 1996).

Alanyazında kuramsal problemlerin çözümü için yararlanılan simülasyon çalışmalarının güçlü yöntemsel araştırmalar oldukları belirtilmiş (Morris vd., 2019) ve bu çalışmalar ile kuramların gelişimine önemli katkılar yapılabileceği vurgulanmıştır (Harwell, vd., 2017). Bu özelliği nedeniyle simülasyon çalışmalarının genellikle bir metodolojik yaklaşım olarak kabul edildiği ifade edilmiştir (Davis, Eisenhardt ve Bingham, 2007; Happach ve Tilebein, 2015). Bu bilgiler doğrultusunda bu araştırma kapsamında gerçekleştirilen simülasyon çalışması ile GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin belirlenen koşullar altında birbirlerine olan üstünlük ve sınırlılıkları Tip I hata, istatistiksel güç ve Tip III hata oranları ile belirlendiğinden, yapılan bu çalışmanın alanyazındaki kuramsal bilgi birikimine katkı sağladığı ve bu nedenle temel araştırma niteliğindeki bir simülasyon çalışması olduğu söylenebilir.

Simülasyon Deseni

Simülasyon çalışmalarında ilgi çekici sonuçlara ulaşmak için simülasyon desenine önem verilmeli ve kuramsal olarak ilgili alana katkı sağlaması beklenen faktörlerin dikkate alınmasına özen gösterilmelidir. Simülasyona dahil edilmek istenilen faktörlerin belirlenmesi sürecinde daha önce yapılmış olan çalışmalar incelenerek kuramsal bilgi eksikliğinin olduğu faktörler tespit edilip hangi faktörlerin dikkate alınmasının daha etkili olacağına karar verilmesi gerekmektedir (Davis vd., 2007; Harwell, Kohli ve Peralta-Torres, 2018; Tureson ve Odland, 2018). Kim (2010) tarafından belirtildiği üzere mevcut çalışmalarda örneklem büyüklüğü, test uzunluğu, grupların yetenek düzeyi, MTK modeli, DMF türü veya henüz keşfedilmemiş çok sayıda faktör istatistiksel tekniklerin DMF belirleme performanslarını etkileyebilmektedir. Bu bağlamda simülasyon çalışmalarının kapsamını yönetilebilir kılmak amacıyla değişimlenen ve sabit tutulan faktörlerin ve bu faktörlerin düzeylerine karşılık gelen koşul sayısının makul olması son derece önemli görülmektedir (Harwell vd., 2018; Sigal ve Chalmers, 2016). Öte yandan dikkate alınan faktörler ile koşulların gerçek uygulamaları yansıtmasının gerekliliği de çok sayıda araştırmacı tarafından vurgulanmaktadır (Bulut ve Sünbül, 2017; Harwell vd., 1996; Scott, 2014).

Bu araştırmacılar tarafından yapılan açıklamalara ilaveten Berends ve Romme (1999) ile Sigal ve Chalmers (2016) tarafından da belirtildiği gibi alanyazındaki diğer araştırmalarda ele alınmamasına rağmen gerçek uygulamalardaki durumları yansıtan faktörlere odaklanılması, konu alanına ilişkin derin bir bakış açısı oluşması açısından anahtar bir rol üstlenmektedir. Ancak her ne kadar simülasyon çalışmalarında ele alınan koşulların gerçek durumu yansıtmasının önemi belirtilmiş olsa da çeşitli faktörlerin alt ve üst sınırlarında istatistiksel tekniklere ilişkin performansların incelendiği çalışmalara da rastlanılmaktadır. Yapılan bu araştırmalarda her bir grubun örneklem büyüklüğünün 30 veya 40 (Rockoff, 2018; Wood, 2011), DMF'li maddelerin %50 veya %60 (Rockoff, 2018; Svetina ve Rutkowski, 2014; Tay vd., 2016) ve DMF miktarının 1.5 (Muñiz vd., 2001) olduğu gerçek uygulamalarda örneğine zor rastlanan uç veya aykırı koşulların dikkate alındığı da görülmektedir. Yapılan bu araştırma kapsamında ise değişimlenen faktörlerin gerçek test koşullarını yansıtması amaçlanmış (Feinberg ve Rubright, 2016; Guilera vd., 2013) ve ilgili araştırmalarda dikkate alınan bazı faktörlerin düzeyleri farklılaştırılarak simülasyon deseni oluşturulmuştur. Araştırma kapsamında değişimlenen ve sabit tutulan koşullar aşağıda sırasıyla açıklanmıştır.

Değişimlenen Koşullar

Yapılan bu simülasyon çalışmasında gerçek uygulama verilerine olabildiğince yakın olan bazı faktörler belirlenmiş ve bu faktörlere ilişkin çeşitli koşullar sistematik olarak değişimlenmiştir. Bu kısımda simülasyon kapsamında değişimlenen faktörler ve bu faktörlere ilişkin çeşitli koşullar hakkında bilgi verilmiş ve bu koşulların dikkate alınma gerekçeleri detaylı bir şekilde açıklanmıştır.

DMF Teknikleri. Yapılan bu araştırma kapsamında hem MTK hem de KTK'ya dayalı olan GLK, GLR-LRT, GLR-Wald, GMH ve Lordif çoklu grup teknikleri birlikte kullanılmış ve tekniklerin çeşitli koşullar altındaki performansları karşılaştırılmıştır. Magis ve diğerlerinin (2011) çalışmalarında GLR tekniği için Wald ki-kare ve Olabilirlik Oran testlerinden her ikisinin de dikkate alınması önerildiğinden, bu çalışmada toplam beş farklı tekniğin birbirlerine olan üstünlük ve sınırlılıkları incelenmiştir. Çalışmada ele alınan beş farklı çoklu grup tekniği TB DMF'yi belirlemek için, GMH haricindeki dört teknik ise TBO DMF'yi belirlemek için kullanılmıştır. Bu çalışmada DMF'nin belirlenmesi amaçlandığından, ikili gruplar düzeyinde çoklu karşılaştırmalar yapılmamasına karar verilmiştir (Rollins III, 2018).

Çoklu grup DMF belirleme tekniklerinin farklı varsayımlara da sahip olmaları nedeniyle bir testte DMF'li olarak tespit edilen maddeler, kullanılan DMF belirleme tekniklerine bağlı olarak değişebilmektedir (Rollins III, 2018). Magis ve diğerlerinin (2011) çalışmalarında belirttiği gibi GLK tekniği için parametrelerin 1PLM'ye dayalı kestirilmesi durumunda TB DMF, 2PLM'ye dayalı kestirilmesi durumunda ise TBO DMF belirlenebilmektedir. Gerçek veri üzerinde yapılan uygulamalarda bir maddenin DMF gösterip göstermediğinin ve DMF türünün önceden bilinmemesi nedeniyle analiz sürecinde farklı DMF türünü belirlemek amacıyla geliştirilen birden fazla istatistiksel tekniğin birlikte kullanılması önerilmektedir (Karami ve Nodoushan, 2011). Böylece DMF belirlemek amacıyla farklı tekniklerin birlikte kullanılması durumunda hem Tip I hata (DMF içermeyen maddeleri belirleme) hem de Tip II hata (DMF içeren maddeleri belirleyememe) yapma olasılığının azaldığı belirtilmektedir (Fidalgo vd., 2004a).

Genel olarak DMF belirleme tekniklerinin metodolojik yaklaşım (MTK, KTK), örneklem büyüklüğü, DMF miktarı, DMF türü (TB, TBO), grup sayısı (iki veya daha fazla) gibi faktörlerden etkilenebildiği çok sayıda çalışmada belirtilmiştir (Carroll, 2015; Finch, 2016; Finch vd., 2016; Schauburger ve Tutz, 2016; Svetina ve Rutkowski, 2014;

Tutz ve Schauburger, 2015). Bu nedenle araştırma kapsamında grup sayısı, referans grubun örneklem büyüklüğü, odak grupların örneklem büyüklüğü oranı, DMF miktarı ve DMF türü değişkenlerinin değişimlenmesine karar verilmiştir.

Grup Sayısı. Çoklu grupların dikkate alındığı simülasyon çalışmalarında değişimlenen koşullardan biri de grup sayısı olmuştur. Bu nedenle yapılan çalışmada bir referans grup ve iki veya üç odak grup olmak üzere, grup sayısının üç ve dört olduğu iki koşul dikkate alınmış ve bütün odak grupların DMF’li madde içermesi sağlanmıştır.

Çoklu grupların dikkate alındığı araştırmalardaki grup sayıları incelendiğinde genel olarak grup sayılarının üç ile altı arasında değiştiği görülmüştür (Battauz, 2019; Carroll, 2015; Finch, 2016; Penfield, 2001; Schauburger ve Tutz, 2016; Tutz ve Schauburger, 2015). Bu araştırmalardan birinde Finch (2016) sosyo-ekonomik düzey veya etnik köken gibi değişkenlerin araştırılması durumunda karşılaşılabilecek grup sayısının üç ile altı arasında olabileceğini ifade etmiştir. Öte yandan geniş ölçekli uluslararası uygulamaların dikkate alındığı bir araştırmada ise grup sayısının 10 ve 20 olduğu belirtilmiştir (Svetina ve Rutkowski, 2014). Penfield (2001) ise pratikteki DMF belirleme çalışmalarında örneklemde üç veya dört alt grup oluşmasının gerçekçi olduğunu açıklamıştır. Bu bağlamda DMF ile ilgili simülasyon çalışmalarında (Carroll, 2015; Li ve Zumbo, 2009) dikkate alınan bazı koşulların eşleştirilmesi, bir yandan sonuçların yorumlamasını kolaylaştırırken diğer yandan simülasyonun geçerli sonuçlar ortaya koyup koymadığının belirlenmesine imkân sağlamıştır.

Referans Grubun Örneklem Büyüklüğü ve Odak Grupların Örneklem Büyüklüklerinin Referans Gruba Oranı. DMF belirleme çalışmalarında grupların örneklem büyüklüğünün istatistiksel tekniklerin performanslarını etkileyen önemli ve temel bir faktör olduğu çok sayıda araştırmada ifade edilmiştir (Kim, 2010; Kristjansson vd., 2005; Rollins III, 2018). Bu nedenle çalışma kapsamında MTK’ya dayalı parametre kestirimlerinin yapılması nedeniyle referans grubun örneklem büyüklüğü 600 ve 1800 olarak belirlenmiştir. Ayrıca referans ve odak gruplar arasındaki örneklem büyüklüğü oranlarının 1:1/1, 1:1/2 ve 1:2/3 olması sağlanarak, odak grup örneklem büyüklüklerinin en az 300, en fazla 1800 arasında değişmesi sağlanmıştır. Böylece toplam örneklem büyüklüğü; üç grubun olduğu durum için 1200 ile 5400 arasında, dört grubun olduğu

durum için ise 1500 ile 7200 arasında deęişiklik göstermiştir. Grup sayısının üç ve dört olduęu durumlar için grupların örneklem büyüklükleri Ek 4’te verilmiştir.

Örneklemin küçük olması gerçekten DMF’li olan maddelerin doğru tespit edilememesine neden olabilirken, büyük örneklerde gerçekten DMF’li olan maddelerin doğru tespit edilmesi kesinleşmekte ve DMF’li olmayan veya DMF miktarı düşük olan maddelerin dahi DMF’li olarak belirlenmesi söz konusu olabilmektedir (Kim, 2010). Bu nedenle genel olarak DMF belirleme tekniklerinin çıkarımsal istatistiklere dayanması nedeniyle, örneklem büyüklüğünün istatistiksel tekniğin gücünü etkileyebildięi belirtilmektedir (Wood, 2011).

Araştırmacılar testteki DMF’li maddelerin doğru belirlenmesi için gerekli olan minimum örneklem büyüklüğü için farklı önerilerde bulunmuşlardır. İki gruplu ve ikili puanlanan maddeler için yapılmış olan araştırmalardan birinde (Candell ve Drasgow, 1988) 300 örneklem büyüklüğünün yetersiz olduęu belirtilmişken, başka araştırmalarda 200 veya 250 (Narayanan ve Swaminathan, 1996; Swaminathan ve Rogers, 1990) örneklem büyüklüğünün alt sınır olabileceęi iddia edilmiştir. Bununla birlikte DMF belirleme tekniklerinin küçük örneklerdeki performanslarını belirlemek amacıyla yapılmış olan bazı simülasyon çalışmalarında 30 odak, 30 referans (Rockoff, 2018), 40 odak 40 referans (Wood, 2011) veya 50 odak 50 referans (Fidalgo vd., 2004b; Muñiz vd., 2001) gibi oldukça küçük örneklerin dikkate alındıęı da görülmüştür.

Alanyazında gruplar için kabul edilen örneklem büyüklüklerinin farklılaşmasının en önemli sebebi olarak, istatistiksel tekniklerin farklı varsayımlara sahip olması gösterilmiştir. Bu çerçevede Camilli’nin (2006) çalışmasında, genel olarak MTK’ya dayalı DMF belirleme teknikleri için büyük örneklerin gerekli olduğuna işaret edilmiştir. Magis ve diğerleri (2015) tarafından yapılan araştırmada ise gözlenen puana dayalı teknikler için 100 örneklem büyüklüğünün yeterli olduęu ifade edilmiştir. Tay ve diğerleri (2015) DMF ile ilgili yaptıkları doküman incelemesi çalışmasında, çoğunlukla grupların en az birinin 500’den (%51.2) daha küçük örnekleme sahip olduğunu ve yine grupların örneklem büyüklüklerinin de genellikle eşit (%37.2) olduğunu belirtmişlerdir. Yapılmış olan açıklamalar çerçevesinde MTK’ya dayalı DMF belirleme tekniklerinin yeterli performans gösterebilmeleri için nispeten orta veya büyük örneklerin gerektięi, KTK’ya dayalı teknikler içinse göreceli olarak küçük örneklerin de yeterli olduęu açıklanmıştır (Fidalgo ve Scalón, 2010). Yapılan bu araştırma kapsamında da hem MTK hem de KTK’ya dayalı tekniklerin birlikte kullanılması nedeniyle kararlı sonuçlar elde etmek için odak grupların çok küçük olmamasına özen gösterilmiştir.

Çoklu gruplarda yapılan simülasyon çalışmalarında referans grubun örneklem büyüklüğü ve her bir odak grubun örneklem büyüklüğünün referans grubuna oranı, istatistiksel tekniklerin performanslarını etkileyen önemli faktörler olarak kabul edilmiş ve çalışmalarda bu faktörlere ilişkin koşulların değişimlendiği görülmüştür (Finch, 2016; Oshima vd., 2015; Penfield, 2001). DMF belirlemek amacıyla MTK'ya dayalı tekniklerin kullanıldığı simülasyon çalışmalarında (Battauz, 2019; Finch, 2016; Oshima vd., 2015), grupların örneklem büyüklüğünün 500 ve 8000 arasında, KTK'ya dayalı tekniklerin kullanıldığı çalışmalarda ise (Finch, 2016; Penfield, 2001; Rollins III, 2018) grupların örneklem büyüklüğünün 200 ve 2000 arasında farklılaştığı tespit edilmiştir. Geniş ölçekli test uygulamalarının dikkate alındığı başka bir çalışmada ise grup sayısı 10 ve 20 olarak ele alınmış ve referans grubun örneklem büyüklüğü sırasıyla 6370 ve 5341 olarak belirlenmiştir. Odak grupların örneklem büyüklükleri ise, grup sayısının 10 olduğu durum için 4494 ile 6545 arasında, grup sayısının 20 olduğu durum için ise 3297 ile 6545 arasında belirlenmiştir (Svetina ve Rutkowski, 2014).

Örneklem büyüklüğünün gruplar arasında eşit olması ideal durum olarak görülse de gerçek uygulamalardaki odak grupların örneklem büyüklüklerinin farklılaştığı durumlara da rastlamak mümkün olmaktadır. Nitekim Kim ve diğerleri (1995) tarafından yapılan çalışmada genellikle referans ve odak grupların örneklem büyüklüklerinin farklı olduğu belirtilmiştir. Odak grupların azınlık oldukları göz önüne alındığında gerçek veriye dayalı araştırmalarda odak gruplardaki örneklem büyüklüğünün referans gruptan küçük olabileceği ifade edilmiştir (Fidalgo ve Scalón, 2010; Yu, Moses, Puhán ve Dorans, 2008). Çünkü gerçek uygulamalarda özellikle odak gruplara ilişkin örneklem büyüklüğü, zor veya çok maliyetli olabilmektedir (Wood, 2011). Finch'in (2016) çalışmasında, odak gruplardaki örneklem büyüklüklerinin her zaman aynı büyüklükte olamayabileceğine dikkat çekilmiş ve farklı örneklem büyüklüklerine sahip odak grupların dikkate alınması gerektiği açıklanmıştır. Ong (2012) tarafından gerçek veriye dayalı yapılan bir araştırmada, referans grubun örneklem büyüklüğünün, odak gruptan yaklaşık 13 kat daha büyük olduğu ortaya konulmuştur. Böylece simülasyon çalışmalarında çok sayıda bireyden oluşan bir referans grup ile az sayıda bireyden oluşan birkaç odak grubun dikkate alınması gerektiğine ilişkin görüş birliğinin oluşması nedeniyle (Battauz, 2019; Fidalgo ve Scalón, 2010; Finch, 2016; Oshima vd., 2015; Rollins III, 2018) odak grupların örneklem büyüklüğünün referans gruba oranının tekniklerin performanslarını etkileyebileceği kabul edilerek (Kristjansson vd., 2005; Penfield, 2001) referans ve odak grupların örneklem büyüklükleri karşılaştırılmıştır.

DMF Miktarı. DMF ile ilgili yapılan simülasyon çalışmalarında genellikle odak grupların madde parametreleri referans gruba göre azaltılıp/arttırılarak farklılaştırılmaktadır (Harwell vd., 1996). Bu nedenle bu araştırma kapsamında ilgili alanyazındaki çalışmalarda (Battaaz, 2019; Finch, 2016; Jin vd., 2015; Oshima vd., 2015; Penfield, 2001; Svetina ve Rutkowski, 2014) yaygın olarak dikkate alınan DMF miktarlarının temsil edilmesi amacıyla odak grupların madde parametrelerine 0.40 ve 0.70 DMF miktarı eklenerek, referans ve odak grupların madde parametreleri arasındaki farkın 0.00, 0.40 ve 0.70 olması sağlanmıştır. Grup sayısının üç ve dört olduğu durumlar için odak grupların madde parametrelerine eklenen DMF miktarları Ek 5'te verilmiştir. Maddelerin DMF içermediği koşulda güçlük ve ayırt edicilik parametreleri referans ve odak gruplar için aynı kaldığından (Fidalgo ve Scalón, 2010; Svetina ve Rutkowski, 2014) DMF içermeyen maddeler ($b_R = b_{O1} = b_{O2}$; $a_R = a_{O1} = a_{O2}$) kullanılarak Tip I hata, DMF içeren maddeler ($b_R < b_{O1} = b_{O2}$; $a_R < a_{O1} = a_{O2}$) kullanılarak da istatistiksel güç (Clark, 2010; Jin vd., 2015; Langer, 2008; Penfield, 2001) ve Tip III hata hesaplanmıştır.

Çoklu gruplarda Penfield (2001), Carroll (2015) ve Finch (2016) tarafından yapılan çalışmalarda bütün odak gruplar için maddelerin güçlük parametrelerinin eşit miktarda arttırıldığı görülmüştür. Böylece referans grubun her zaman avantajlı, odak grupların ise eşit düzeyde dezavantajlı oldukları kabul edilmiştir. Bu çalışmalarda olduğu gibi yapılan simülasyon çalışmasında da referans grup belirlenerek, bu referans grubun her zaman avantajlı olması sağlanmıştır. Ayrıca Finch'in (2016) çalışmasında önerildiği gibi yapılan bu çalışmada DMF'li olarak belirlenen maddelerin parametrelerine farklı DMF miktarları eklenerek odak grupların hem güçlük hem de ayırt edicilik parametreleri değişimlenmiş ve böylelikle odak grupların da kendi aralarında avantajlı/dezavantajlı olma durumları simüle edilmiştir. Öte yandan Oshima ve diğerlerinin (2015) çalışmasında olduğu gibi yapılan bu araştırma kapsamında da yalnızca güçlük, yalnızca ayırt edicilik ve hem güçlük hem de ayırt edicilik parametreleri için odak grupların madde parametreleri farklılaştırılmıştır.

Alanyazında grup sayısı önemli olmaksızın farklı DMF miktarlarını dikkate alan simülasyon çalışmalarına rastlamak mümkündür. TB DMF'de ayırt edicilik parametresinin gruplar arasında eşit olması sağlanırken, güçlük parametresi değişimlenerek (Hidalgo, Gómez-Benito ve Zumbo, 2014) ilgili maddeler odak gruplar için zorlaştırılmaktadır (Finch, 2016; Stark vd., 2006; Woods, 2008). Farklı araştırmacılar tarafından TB DMF için iki grupta yapılmış olan çalışmalarda, referans ve odak grubun güçlük parametreleri arasındaki DMF miktarları; 0.20 (zayıf) ve 0.35 (orta) (Shepard,

Camilli ve Williams, 1985); 0.48 (orta) ve 0.64 (yüksek) (Swaminathan ve Rogers, 1990); 0.50 (orta) ve 1.00 (büyük) (Kim, 2010); 0.30 (orta) ve 0.60 (yüksek) (Tay vd., 2016); 0.40 (orta) ve 0.80 (yüksek) (Tay vd., 2013); 0.45 (düşük) ve 0.90 (yüksek) (Socha vd., 2015); 0.40 (düşük) ve 0.80 (yüksek) (Magis ve De Boeck, 2012) olmak üzere çeşitli şekillerde dikkate alınmıştır. Benzer şekilde çoklu gruplarda yapılmış olan çalışmalarda referans ve odak grupların güçlük parametreleri arasındaki DMF miktarları; 0.50 (düşük) ve 1.0 (yüksek) (Svetina ve Rutkowski, 2014); 0.20 (düşük), 0.40 (orta) ve 0.60 (yüksek) (Jin vd., 2015); 0.40 (düşük), 0.60 (orta) ve 0.80 (yüksek) (Finch, 2016); 0.40 (orta) ve 0.60 (yüksek) (Battauz, 2019) olmak üzere farklı şekillerde tanımlanmıştır.

TBO(a) DMF türünde güçlük parametresinin gruplar arasında eşit olması sağlanırken, ayırt edicilik parametresi değişimlenerek ilgili maddeler odak gruplar için daha ayırt edici hale getirilmektedir (Bulut ve Sünbül, 2017; Desjardins ve Bulut, 2018; Lopez Rivas, 2012; Oshima vd., 2015; Woods, 2008). Ayırt edicilik parametresi ne kadar yüksekse madde karakteristik eğrileri de o kadar dikleşmektedir (Muñiz vd., 2001). Farklı araştırmacılar tarafından TBO(a) DMF türü için iki grupta yapılmış olan çalışmalarda referans ve odak grubun ayırt edicilik parametreleri arasındaki DMF miktarları; 0.40 (orta) ve 0.80 (yüksek) (Lopez Rivas, 2012); 0.30 (düşük) ve 0.60 (yüksek) (Güler ve Penfield, 2009) olmak üzere çeşitli şekillerde tanımlanmıştır. Benzer şekilde çoklu gruplarda yapılmış olan çalışmalarda referans ve odak grupların ayırt edicilik parametreleri arasındaki DMF miktarları; 0.40 (düşük) ve 0.80 (yüksek) (Svetina ve Rutkowski, 2014); 0.13 (düşük), 0.26 (orta) ve 0.39 (yüksek) (Jin vd., 2015); 0.30 (orta) ve 0.50 (yüksek) (Battauz, 2019) olmak üzere çeşitli şekillerde dikkate alınmıştır. TB DMF’de olduğu gibi TBO(a) DMF için de ilgili araştırmalardaki DMF miktarlarına göre yapılan tanımlamalar son derece tartışmalıdır. Lee ve diğerleri (2016) tarafından da belirtildiği gibi DMF miktarı etki büyüklüğü gibi mutlak bir değeri temsil etmediğinden düşük, orta ve yüksek DMF miktarlarının göreceli olarak yorumlanması gerekmektedir.

DMF Türü (Kaynağı). Simülasyon çalışmalarında odak grupların madde parametrelerine eklenen DMF’nin miktarı ile DMF’nin türü arasında ilişki olması nedeniyle odak gruplara yalnızca güçlük parametresi eklenerek TB DMF’li, yalnızca ayırt edicilik parametresi eklenerek TBO(a) DMF’li ve hem güçlük hem de ayırt edicilik parametreleri birlikte eklenerek TBO(ab) DMF’li maddeler simüle edilmiştir. Yapılan çalışmada mevcut araştırmalardan farklı olarak TBO(ab) DMF türüne de odaklanılmış ve

GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin TB DMF, TBO(a) DMF ve TBO(ab) DMF içeren maddelerdeki performansları ayrı ayrı ortaya konulmuştur.

Madde yazma ve test geliştirme aşamasındayken testteki maddelerin farklı gruplar için DMF gösterip göstermeyeceği test geliştiren araştırmacılar tarafından kesin olarak bilinmediğinden (Wood, 2011), geliştirilmiş testlerin gerçek uygulamalarda kullanılması durumunda, testteki maddelerin DMF açısından incelenmesi (Finch ve French, 2008) ve DMF’li olarak belirlenen maddelerin de DMF türünün tespit edilmesi gerekmektedir. Bir madde TB DMF göstermişse odak ve referans grubun arasındaki farkın tek bir yönü olduğu ifade edilebilmektedir. TBO DMF’de ise referans grup odak gruptan daha iyi bir performans gösterebilse de bu durumun belirli bir yetenek düzeyinin üzerindeki bireyler için tersine dönmesi nedeniyle gruplar arasındaki farkın tek bir yönünün olmadığı belirtilebilmektedir (Carroll, 2015; Desjardins ve Bulut, 2018). İlgili araştırmalarda olduğu gibi (Carroll, 2015; Li ve Zumbo, 2009) bu çalışma kapsamında da TB DMF’de referans grubun odak gruplardan daha avantajlı olduğu kabul edilmektedir.

DMF ile ilgili çalışmalarda genellikle güçlük parametresinden kaynaklı DMF’nin ortaya çıktığı ifade edilmiştir (Smith ve Reise, 1998; Steinberg, 2001). Bununla birlikte yapılan birkaç araştırmada maddelerin %15 ila %20’sinde ayırt edicilik parametresinden kaynaklı DMF’ye de rastlanıldığı belirtilmiştir (Maller, 2001; Morales, Reise ve Hays, 2000; Rodebaugh, Woods, Heimberg, Liebowitz ve Schneier, 2006). TB ve TBO DMF’nin farklı nedenlere dayalı olarak oluştuğunun öne sürülmesi nedeniyle (Penfield, 2010), maddeleri uygun şekilde düzenlemek için DMF türünün doğru tanımlanmasının kritik bir öneme sahip olduğu açıklanmıştır (Finch ve French, 2008).

Alanyazında TB ve TBO olmak üzere yalnızca iki DMF türünün var olduğu belirtilmesine rağmen bazı çalışmalarda TBO DMF’nin de iki farklı türünün olduğu vurgulanmıştır (Hope vd., 2018; Penfield, 2003; Smith, 2002; Walker vd., 2001). Bazı DMF belirleme tekniklerinin farklı DMF türüne sahip maddeleri belirlemede daha iyi performans gösterebilmeleri nedeniyle, TB DMF’nin yanı sıra TBO(a) ve TBO(ab) DMF türlerinin de dikkate alınması gerektiği tavsiye edilmiştir (Rollins III, 2018).

Çoklu gruplarda DMF belirlemek amacıyla yapılan simülasyon çalışmalarında TB DMF (Battauz, 2019; Carroll, 2015; Finch, 2016; Jin vd., 2015; Rollins III, 2018; Schauburger ve Tutz, 2016; Svetina ve Rutkowski, 2014; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) ve TBO(a) DMF (Battauz, 2019; Carroll, 2015; Jin vd., 2015; Svetina ve Rutkowski, 2014) sıklıkla incelenmiş ve yapılacak yeni çalışmalarda hem güçlük hem de ayırt edicilik parametrelerinin birlikte ele alındığı [TBO(ab)] DMF türünün de

araştırılması gerektiği belirtilmiştir. Oshima ve diğerleri (2015) tarafından yapılan çalışmada üç DMF türü dikkate alınmasına rağmen yalnızca MG-NCDIF tekniğinin performansı incelendiğinden yapılan çalışma kapsamında GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin TB DMF, TBO(*a*) DMF ve TBO(*ab*) DMF içeren maddelerdeki performanslarının belirlenmesi önemli görülmüştür.

Sonuç olarak gerçek uygulamalara dayalı çalışmalarda çoklu grup teknikleri kullanılmadan önce, bu tekniklerin hangi özelliklere sahip veri setlerini analiz etmek amacıyla tercih edilmesi gerektiği belirlenmelidir. Bu anlamda yapılan bu simülasyon çalışması ile grup sayısı, referans grubun örneklem büyüklüğü, odak grupların örneklem büyüklüğü oranları, DMF miktarı, DMF türü (Kaynağı) ve DMF tekniği faktörlerine dayalı olarak çeşitli koşullar dikkate alınmıştır. Araştırma kapsamında değişimlenen koşulların özeti Tablo 1’de verilmiştir.

Tablo 1

Simülasyon Deseni

Grup Sayısı	Referans Grubun Örneklem Büyüklüğü	Odak Grubun Örneklem Büyüklüğünün Referans Gruba Oranı	DMF Miktarı	DMF Türü	DMF Teknikleri
3	600 ve 1800	1:1/1, 1:1/2, 1:2/3	$a(0.0,0.4,0.7)$, $b(0.0,0.4,0.7)$	DMF Yok, TB(<i>b</i>), TBO(<i>a</i>), TBO(<i>ab</i>)	GLK, GLR-LRT, GLR-Wald, GMH, Lordif
4	600 ve 1800	1:1/1, 1:1/2, 1:2/3	$a(0.0,0.4,0.7)$, $b(0.0,0.4,0.7)$	DMF Yok, TB(<i>b</i>), TBO(<i>a</i>), TBO(<i>ab</i>)	GLK, GLR-LRT, GLR-Wald, GMH, Lordif
3 ve 4	600 ve 1800	1:1/1, 1:1/2, 1:2/3	DMF Yok, Düşük, Orta Düzey, Yüksek	DMF Yok, TB(<i>b</i>), TBO(<i>a</i>), TBO(<i>ab</i>)	GLK, GLR-LRT, GLR-Wald, GMH, Lordif

Tablo 1’den görüldüğü üzere genel amacı çoklu gruplar için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif DMF belirleme tekniklerinin performanslarını karşılaştırmak olan bu araştırma kapsamında; grup sayısı (3, 4), referans grubun örneklem büyüklüğü (600, 1800), odak grubun örneklem büyüklüğünün referans gruba oranı (1:1/1, 1:1/2, 1:2/3), DMF miktarı [$a(0.00, 0.40, 0.70)$, $b(0.00, 0.40, 0.70)$] ve DMF türü [DMF Yok, TB(*b*), TBO(*a*), TBO(*ab*)] faktörlerinin değişimlendiği bir simülasyon çalışması

yürütülmüştür. Testteki maddelerin DMF içermediği (DMF Yok) durum için bir testte aynı puana sahip tüm bireylerin, hangi grupta olduklarına bakılmaksızın maddeleri doğru yanıtlama olasılıklarının aynı olduğu belirtilmiş ve bu koşul için üretilen veriler kullanılarak temel Tip I hata hesaplanmıştır. Çalışmanın simülasyon deseninde değişimlenen koşullar araştırma sorularına dayalı olarak aşağıda özetlenmiştir.

Birinci araştırma sorusunda öncelikle grup sayısının üç olduğu durum ele alınmış ve iki (referans grubun örneklem büyüklüğü) x dokuz (odak grupların örneklem büyüklüklerinin, referans grubun örneklem büyüklüğüne oranı) x yirmi beş (gruplara eklenen DMF miktarı) x beş (DMF tekniği) olmak üzere toplam 2250 koşul simüle edilmiştir. Simüle edilmiş koşulların tamamı Ek 6'da verilmiş ve Seybert ve Stark (2012) tarafından da belirtildiği üzere TBO (*ab*) DMF türü için koşul sayısının arttığı görülmüştür. Çalışma kapsamında referans grubun örneklem büyüklüğü 600 veya 1800 olarak belirlenmiştir. Odak grupların örneklem büyüklüklerinin referans grubun büyüklüğüne oranlanması ile en az 300, en fazla 1800 arasında değişen toplam 9 farklı odak grup örneklem büyüklüğü oluşmuştur. Böylelikle grup sayısının üç olduğu durumda referans ve odak grupların örneklem büyüklüklerinin eşit veya referans grup örneklem büyüklüğünün daha büyük olduğu koşullar dikkate alınmış ve toplam örneklem büyüklüğünün 1200 ile 5400 arasında değişmesi sağlanmıştır. Ayrıca odak gruplara eklenen DMF miktarları ise 25 farklı koşuldan [1(DMF Yok) + 4(TB) + 4(TBO(*a*)) + 16(TBO(*ab*))] oluşmuştur. Bu anlamda her koşulda 100 tekrar gerçekleştirilerek toplam $2 \times 9 \times 25 \times 5 \times 100 = 225,000$ satırlık simülatif veri seti üretilmiştir.

Grup sayısının dört olduğu durum ele alındığında ise iki (referans grubun örneklem büyüklüğü) x yirmi yedi (odak grupların örneklem büyüklüklerinin, referans grubun örneklem büyüklüğüne oranı) x seksen bir (gruplara eklenen DMF miktarı) x beş (DMF tekniği) olmak üzere toplam 21870 koşul simüle edilmiştir. Çalışma kapsamında referans grubun örneklem büyüklüğü 600 veya 1800 olarak belirlenmiştir. Odak grupların örneklem büyüklüklerinin referans grubun büyüklüğüne oranlanması ile en az 300, en fazla 1800 arasında değişen toplam 27 farklı örneklem büyüklüğü oluşmuştur. Böylece grup sayısının dört olduğu durumda referans ve odak grupların örneklem büyüklüklerinin eşit veya referans grup örneklem büyüklüğünün daha büyük olduğu koşullar dikkate alınmış ve toplam örneklem büyüklüğünün 1500 ile 7200 arasında değişmesi sağlanmıştır. Ayrıca odak gruplara eklenen DMF miktarları ise 81 farklı koşuldan [1(DMF Yok) + 8(TB) + 8(TBO(*a*)) + 64(TBO(*ab*))] oluşmuştur. Böylece her koşulda 100 tekrar yapılarak toplam $2 \times 27 \times 81 \times 5 \times 100 = 2,187,000$ satırlık simülatif veri seti üretilmiştir. Bu anlamda dört

grupta odak grup sayısının artmasına bağılı olarak hem örneklem büyüklüğü hem de DMF miktarı faktörlerindeki toplam koşul sayısının arttığı gözlemlenmiştir.

Yapılan bu simülasyon çalışması kapsamında öncelikle grup sayısının yalnızca üç ve yalnızca dört olduğu birinci araştırma sorusu için Tip I hata, istatistiksel güç ve Tip III hata oranları hesaplanmıştır. Ardından ikinci araştırma sorusuna yanıt aramak amacıyla üç ve dört grup için üretilen veri setleri birleştirilmiş ve grup sayısı da bir faktör olarak dikkate alınmıştır. Bu bağlamda grup sayısının değişimlenmesine bağılı olarak toplam DMF miktarının da tanımlanması gerekmiş ve odak gruplardaki DMF miktarlarının 0.40 olması durumunda düşük düzey DMF'nin, 0.70 yüksek olması durumunda ise yüksek düzey DMF'nin var olduğu kabul edilmiştir. Ayrıca odak gruplardan birine 0.40, diğerine 0.70 DMF miktarı eklendiğinde ise toplam DMF miktarının orta düzey olduğu dikkate alınarak DMF miktarı için “DMF Yok”, “Düşük DMF”, “Orta Düzey DMF”, “Yüksek DMF” olmak üzere dört farklı düzey tanımlanmıştır.

Grup sayısının hem üç hem de dört olduğu durumda “odak grupların örneklem büyüklüklerinin referans gruba oranı” faktörlerinin ana etkileri ile çoklu grup DMF belirleme teknikleri ile ikili etkileşimlerine ilişkin etki büyüklüklerinin çok düşük olması nedeniyle yapılan Faktöriyel ANOVA'larda bu faktörün dikkate alınmasına gerek olmadığına karar verilmiştir. Bu durumda çalışma kapsamındaki ikinci araştırma sorusu için grup sayısı (3 ve 4), referans grubun örneklem büyüklüğü (600 ve 1800), DMF miktarı (DMF Yok, Düşük DMF, Orta Düzey DMF ve Yüksek DMF), DMF türü ise [DMF Yok, TB, TBO(a) ve TBO(ab)] olmak üzere çeşitli koşullar dikkate alınmıştır. Böylece birleştirilmiş yeni veri setleri kullanılarak Tip I hata, istatistiksel güç ve Tip III hata oranları için üç farklı Faktöriyel ANOVA gerçekleştirilmiş ve GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin değişimlenen koşullar altındaki performansları karşılaştırılmıştır.

Sabit Tutulan Koşullar

Yapılan bu simülasyon çalışmasında bazı faktörler sistematik olarak değişimlenirken, diğerleri sabit tutularak kontrol edilmiştir. Çalışmada değişimlenen koşulların yanı sıra sabit tutulan koşulların da gerçek uygulama verilerine olabildiğince yakın olması amaçlanmıştır. Bu kısımda simülasyon kapsamında kontrol edilen faktörler ile bu faktörlere ilişkin sabit tutulan koşullar hakkında bilgi verilmiş ve seçilme gerekçeleri detaylı bir şekilde açıklanmıştır.

Veri Üretmede Kullanılan Model. İkili puanlanan maddeler için çoklu gruplarda DMF incelemek amacıyla yapılan simülasyon çalışmalarında hem TB hem de TBO DMF için genellikle 2PLM'ye dayalı veri üretildiğinin belirtilmesi (Finch, 2016; Rollins III, 2018; Svetina ve Rutkowski, 2014), 3PLM ile yapılan kestirimlerde her bir grubun örneklem büyüklüğü için en az 500 kişinin gerekmesi (Kim ve Oshima, 2013), gerçek test uygulamalarından elde edilen verilerin daha çok 2PLM'ye uyum sağlaması (Bergan, 2010) ve şans parametresi sebebiyle 3PLM ile gerçekleştirilen kestirimlerin standart hatalarının büyük olması (Hambleton vd., 1991) gibi nedenlerle, yapılan bu araştırma kapsamında veri setlerinin 2PLM'ye dayalı olarak üretilmesine karar verilmiştir.

DMF, ölçülen özellik açısından eşit düzeye sahip farklı gruplardaki bireylerin, belirli bir maddeyi doğru yanıtlama olasılıklarının farklı olduğunu ortaya koymaktadır. Bu farkın altında yatan özelliği belirlemek için ise olasılıkların nasıl modellendiğini anlamakta yarar vardır (Finch ve French, 2008). Hem tek boyutlu hem de çok boyutlu veriler için çeşitli MTK modelleri geliştirilmesine rağmen, yapılan bu çalışma kapsamında tek boyutlu modeller dikkate alınmıştır.

Tek boyutlu testlerde ikili puanlanan maddelere verilen yanıtlar, genellikle ölçülen gizil özelliği belirli bir maddeye özgü özelliklerle ilişkilendiren birkaç lojistik modelden biri ile karakterize edilmektedir. Bu modellerden 1PLM'de yalnızca güçlük, 2PLM'de güçlük ve ayırt edicilik, 3PLM'de ise güçlük, ayırt edicilik ve şans parametreleri yer almaktadır. Böylelikle MTK modelleri, maddeye özgü tahmin edilen bilgi miktarı açısından farklılık göstermektedir. MTK'da model-veri uyumu sağlandığı takdirde, madde parametreleri örneklemden bağımsız, yetenek parametreleri ise maddelerden bağımsız olarak kestirilebilmektedir (de Ayala, 2009; DeMars, 2010).

Araştırmalarda evet/hayır, katılıyorum/katılmıyorum veya doğru/yanlış gibi ikili puanlanan maddelerin MTK'ya dayalı olarak analiz edilmesi sürecinde sırasıyla 2PLM (%66.7), 3PLM (%29.2) ve 1PLM (%4.1) için model veri uyumunun sağlandığı ifade edilmiştir (Tay vd., 2015). Tay ve diğerleri (2015) tarafından yapılan doküman analizi çalışmasında ortaya konulduğu gibi az sayıda araştırmada verilerin 1PLM'ye uyum sağladığı belirlenmiştir. Bununla birlikte ikili puanlanan maddelere odaklanılan DMF çalışmalarında veri 1PLM'ye dayalı üretildiğinde, testteki maddelerin ayırt edicilik parametrelerinin farklılaşmaması nedeniyle TBO DMF incelenemediğinden (Lai vd., 2005) DMF belirleme tekniklerinin hem TB hem de TBO DMF türündeki performanslarının belirlenebilmesi için simülasyon çalışmalarında 2PLM veya 3PLM'ye dayalı veri üretilmesi gerektiği açıklanmıştır (Battauz, 2019). Battauz (2019) tarafından

yapılan arařtırmada  zellikle LK ve GLK tekniklerinin veri  retilirken dikkate alınan MTK modelinden etkilendiđi ve 2PLM'ya dayalı olarak veri  retilendiđinde bu tekniklerin Tip I hata a ısından daha iyi performans g sterdiđi belirtilmiřtir.

Test Uzunluđu (Madde Sayısı). Yapılan bu sim lasyon  alıřmasında test uzunluđuğunun sabit bir fakt r olarak dikkate alınması uygun g r lm ř ve elde edilen sonu ların ilgili arařtırmalarla karřılařtırılabilmesi amacıyla ikili puanlanan 20 maddenin  retilmesine karar verilmiřtir. Test uzunluđuğunun sim lasyon  alıřmalarında sıklıkla dikkate alınan fakt rlerden biri olması nedeniyle belirlenen madde sayısının ger ek durumu yansıtmasının son derece  nemli olduđu vurgulanmıřtır (Bulut ve S n  l, 2017). Eđitim ve psikometri alanlarındaki arařtırmalarda test uzunluđuğunun  oklu puanlanan maddeler i in genellikle 20 ila 40 madde arasında (Hidalgo vd., 2014), ikili puanlanan maddeler i in ise 10 ile 80 madde arasında deđiřtiđi ifade edilmiřtir (Wood, 2011).

Alanyazında test uzunluđuğunun belirlenmesi ile ilgili farklı g r řler bulunmaktadır. Yapılmıř olan  ok sayıdaki  oklu grup DMF belirleme  alıřmasında ger ek uygulamalardaki madde sayıları dikkate alınmıř ve genellikle 20 maddeden oluřan test uzunluđu sabit fakt r olarak sim lasyona dahil edilmiřtir (Finch, 2016; Jin vd., 2015; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015).  oklu gruplar arasında DMF belirleme ile ilgilenen diđer arařtırmacılar ise 15 madde (Svetina ve Rutkowski, 2014), 27 madde (Fidalgo ve Scalón, 2010), 30 madde (Oshima vd., 2015), 40 madde (Rollins III, 2018), 50 madde (Carroll, 2015) ve 60 madde (Penfield, 2001) olmak  zere farklı test uzunluklarını sabit fakt r olarak ele almıřlardır.  oklu grupların incelendiđi bařka bir sim lasyon  alıřmasında ise test uzunluđu deđiřimlenerek 20 ve 40 madde dikkate alınmıřtır (Batta , 2019).

 oklu gruplara benzer řekilde iki grup i in yapılan bir ok sim lasyon  alıřmasında da madde sayısının genellikle 20, 40 veya 60 olarak belirlendiđi g r lm řtir (DeMars, 2009; Dodeen, 2004; Jodoin ve Gierl, 2001; Kim, 2010; Magis vd., 2015; Price, 2014; Rogers ve Swaminathan, 1993). Bu bađlamda yapılan bu arařtırmanın amacı dođrultusunda ger ek madde parametrelerine dayalı olarak sim le edilen 20 maddenin, hem sınıf i i  l me uygulamalarındaki  oktan se meli testlerdeki hem de geniř  l ekli test uygulamalarının alt testlerindeki madde sayıları ile uyumlu olması ama lanmıřtır.

DMF’li Madde Yüzdesi (DMF İçeren Madde Sayısı). Araştırma kapsamında DMF’li madde yüzdesi sabit koşul olarak ele alınmış ve testteki maddelerin %20’sinin DMF’li olması sağlanacak şekilde veri üretilmiştir. Araştırmada üretilen veri setlerindeki madde sayısının 20 olması nedeniyle, Finch’in (2016) çalışmasına benzer şekilde bütün tekrarlar da aynı dört maddenin DMF içermesi, geri kalan 16 maddenin ise DMF göstermemesi amaçlanmıştır. Ayrıca temel Tip I hata oranının hesaplanabilmesi amacıyla 20 maddelik testte hiçbir DMF’li maddenin bulunmadığı durum da dikkate alınmıştır. Clark (2010) ve Scott (2014) tarafından yapılan çalışmalarda temel Tip I hata oranının hesaplanabilmesi amacıyla testteki hiçbir maddenin DMF göstermemesi gerektiği belirtilmiştir. Diğer taraftan testin DMF’li madde içermesi durumunda, odak gruplar için DMF içermeyen maddeler kullanılarak da Tip I hata hesaplanabildiği açıklanmıştır.

Gerçek uygulamalarda kullanılan testlerin birden daha fazla sayıda (Wang ve Yeh, 2003) ve yaklaşık %20 veya daha yüksek yüzdeye sahip DMF’li madde içermeye eğiliminde olduğu ifade edilmektedir (Gierl, Gotzmann ve Boughton, 2004). Ayrıca yapılan çalışmalarda genellikle kısa testlerin kullanılmasına bağlı olarak DMF’li madde sayısının bir ile beş arasında değişiklik gösterdiği belirtilmektedir (Kim vd., 1995; Magis vd., 2011). Bu anlamda ideal olan testteki maddelerin DMF göstermemesi olsa da istatistiksel tekniklerin performanslarını karşılaştırmak amacıyla yapılan simülasyon çalışmalarında DMF’li madde yüzdesinin sabit olarak dikkate alındığı veya değişimlendiği görülmektedir (Wood, 2011).

Rockoff (2018) tarafından yapılan araştırmada test uzunluğu (10 madde, 20 madde, 40 madde) ile DMF’li madde yüzdesinin (%10, %20, %50) eş zamanlı olarak dikkate alınması gerektiği vurgulanmaktadır. Bu nedenle sınıf içi veya geniş ölçekli uygulamalarda kullanılan testlerdeki madde sayısına göre DMF’li madde yüzdesinde de artış olabileceği ifade edildiğinden araştırmalarda testin uzunluğu ile DMF içeren madde sayısının birlikte değerlendirilmesinin son derece önemli olduğu belirtilebilmektedir.

İki grubun dikkate alındığı simülasyon çalışmaları incelendiğinde; yalnızca %10 (Kim, 2010); %0, %10 ve %20 (Scott, 2014); %20 ve %40 (Cao vd., 2017); %0, %10, %20, %30 ve %40 (Jin vd., 2018) olmak üzere farklı DMF’li madde yüzdelерinin araştırıldığı görülmüştür. Çoklu gruplarda yapılan simülasyon çalışmaları incelendiğinde ise; %0 ve %22 (Fidalgo ve Scalón, 2010); %20, %40 ve %60 (Svetina ve Rutkowski, 2014); %0, %10 ve %20 (Jin vd., 2015); %20 ve %40 (Tutz ve Schauburger, 2015); %0, %20 ve %40 (Schauburger ve Tutz, 2016); %0 ve %20 (Tutz ve Berger, 2016); %0, %5 ve %10 (Rollins III, 2018); %0, %5 ve %20 (Battaüz, 2019) olmak üzere DMF’li

maddelerin yüzdesinin sıklıkla değişimlendiği belirlenmiştir. Böylelikle değişimlenen koşul olarak dikkate alınan DMF'li madde yüzdesinin %5 ile %60 arasında farklılaştığı ve genellikle testteki maddelerin %10 veya %20'sinin DMF'li kabul edildiği sonucuna ulaşılmıştır (Cohen ve Kim, 1993; Finch ve French, 2007; French ve Maller, 2007).

DMF çalışmalarında aynı yetenek düzeyinde, ancak farklı gruplardaki bireylerin eşleştirilmesi gerektiğinden dikkate alınması gereken önemli konulardan biri de eşleştirme kriterinin belirlenmesidir. KTK ve MTK'ya dayalı tekniklerde eşleştirme kriterlerinin farklılaştığı ifade edilmektedir (Sireci ve Rios, 2013). Testteki DMF'li maddelerin KTK'ya dayalı tekniklerde toplam test puanı üzerinde, MTK'ya dayalı tekniklerde ise bireyler için kestirilen yetenek parametreleri üzerinde istenmeyen bir etkisinin olduğu belirtilmektedir (Magis vd., 2010).

DMF çalışmalarında incelenen madde dışında kalan diğer maddeler çapa veya DMF içermeyen maddeler olarak tanımlanmakta (Choi vd., 2011; Frederickx, Tuerlinckx, De Boeck ve Magis, 2010) ve bu maddelerin DMF içermediği varsayılmaktadır (Millsap ve Everson, 1993). Ancak çapa maddelerin DMF içermediği kabul edilmesine rağmen, bu maddelerde de DMF gözlenebildiği vurgulanmaktadır (Narayanan ve Swaminathan, 1996; Rogers ve Swaminathan, 1993). Başka bir anlatımla gerçek uygulamalarda tüm çapa maddelerin DMF'den arınık (saf) olması garanti edilememekte (Carroll, 2015) ve çapa olarak dikkate alınan maddelerin DMF'li olması durumunda, eşleştirme kriteri için hesaplanan puanlarda kirlenme (contamination) olduğu belirtilmektedir (Kopf, Zeileis ve Strobl, 2015a; Woods, 2009a).

Rollins III'in (2018) çalışmasında, testte yer alan DMF'li madde sayısındaki artışın eşleştirme puanında kirlenmeye yol açması nedeniyle genellikle gözlenen puana dayalı istatistiksel teknikler için Tip I hata ve istatistiksel gücün etkilenebildiği ifade edilmiştir. Guilera ve diğerleri (2013) tarafından yapılan bir meta analiz çalışmasında ise, maddelerin %20'den fazlasının DMF içermesinin genellikle Tip I hata oranlarının artmasına, istatistiksel güç oranlarının ise azalmasına neden olduğu belirtilmiştir.

Araştırma sonuçlarından görüldüğü üzere çalışmalarda saflaştırma işleminin yapılmasına gerek olup olmadığına, testteki DMF'li madde yüzdesi dikkate alınarak karar verilmelidir. Genel olarak testteki DMF'li madde sayısının birden fazla olması durumunda, eşleştirme puanında saflaştırma yapılması önerilmesine rağmen saflaştırmanın hangi koşullar altında DMF belirleme tekniklerinin performansını etkilediği konusunda alanyazında net bir sonuç ortaya konulmadığı belirtilmektedir (Magis ve Facon, 2013).

DMF ile ilgilenen araştırmacılar çalışmalarında hangi koşullarda saflaştırma yapılmasının gerektiğini veya saflaştırmanın hangi yaklaşıma dayalı olarak yapılmasının daha etkili olduğunu belirlemek amacıyla çalışmalarına devam etmektedirler (Craig, 2017; Fikis ve Oshima, 2017; French ve Maller, 2007; Kopf vd., 2015a; Kopf, Zeileis ve Strobl, 2015b; Woods, 2009a). Bu konudaki ilk araştırmalardan biri Zwick (1990) tarafından yapılmış ve DMF'li maddeleri testten çıkarmanın çeşitli sorunlara neden olabileceği ortaya konulmuştur. Benzer şekilde yine Zwick ve diğerleri (1993) tarafından yapılan çalışmada, eşleştirme puanı hesaplanırken gerçek uygulamalardaki testlerde de DMF'li maddelerin yer aldığı durumlar dikkate alınarak saflaştırma işleminin gerçekleştirilmediği açıklanmıştır. Clauser, Mazor ve Hambleton (1993) ile French ve Maller, (2007) yaptıkları araştırmalarında saflaştırmanın tüm koşullar için istatistiksel güç ve Tip I hata oranlarında önemli bir iyileşmeye yol açmadığını belirtmişlerdir.

Yapılan bu araştırma kapsamında da saflaştırma yapılıp yapılmamasına karar verilirken DMF ile ilgili yapılan önceki çalışmalar incelenmiş ve hem iki grupta (Finch, 2005; Lan, 2014; Tunç, 2016) hem de çoklu gruplarda (Jin vd., 2015; Rollins III, 2018; Svetina ve Rutkowski, 2014) yapılan bazı çalışmalarda saflaştırma yöntemlerine başvurulmadan DMF analizlerinin gerçekleştirildiği görülmüştür. Bu araştırmada dikkate alınan Lordif için yinelemeli bir saflaştırma işleminin otomatik olarak gerçekleştirilmesi nedeniyle beş farklı teknikte standart bir saflaştırma yönteminin kullanılması mümkün olmamıştır. Bu anlamda beş teknik için saflaştırma yöntemlerinin standart olmamasına bağlı olarak madde saflaştırmanın çalışmaya ek bir karışıklık getireceği düşünüldüğünden, saflaştırma işleminin yapılmamasına karar verilmiştir.

Yetenek ve Madde Parametrelerinin Dağılımı. Yapılan bu araştırmada ikili puanlanan madde yanıtlarını 2PLM'ye dayalı simüle etmek amacıyla Finch'in (2016) çalışmasındaki gerçek test uygulamasından elde edilen madde parametrelerinin dağılımları dikkate alınarak, güçlük parametresi için $b \sim U(-1.06, 1.39)$, ayırt edicilik parametresi için ise $a \sim U(0.84, 1.53)$ olacak şekilde tekdüze (uniform) dağılımdan 20 madde için rastgele olarak veri üretilmiştir. Aşırı zor maddelerin diğer maddelere kıyasla daha küçük varyans oluşturmaları nedeniyle, 20 maddenin güçlük parametrelerinin -1.06 ile 1.39 aralığında olması tamamen homojen (tümü sıfırlardan veya birlerden oluşan) tepki örüntülerinin önlenmesine olanak sağlamış ve maddelerin zorluk düzeylerinin sonuçları etkileme şansı en aza indirilmiştir. Ayrıca yapılan bu çalışmada referans ve odak

grupların yetenek parametrelerine ilişkin dağılımlar, -3 ile +3 aralığında ortalaması 0, standart sapması 1 olan normal dağılımdan [$\theta \sim N(0, 1)$] üretilmiştir.

Alanyazında referans ve odak gruplar için yetenek dağılımlarının, istatistiksel tekniklerin DMF’li maddeleri tespit etme gücü üzerinde önemli bir etkiye sahip olduğu belirtilmiştir (Wood, 2011). Çoklu gruplar için yapılan simülasyonlarda referans ve odak gruplar için farklı yetenek dağılımlarının dikkate alındığı (Battauz, 2019; Finch, 2016; Oshima vd., 2015; Penfield, 2001; Rollins III, 2018) ifade edilmesine rağmen, Wood (2011) tarafından yapılan çalışmada grupların yeteneklerinin standart normal dağılıma sahip olmasının önemli olduğu belirtilmiştir. Çoklu grupların dikkate alındığı çok sayıda araştırmada referans ve odak grupların yetenek dağılımlarının sabit bir koşul olarak ele alınıp normal dağılımdan üretilmesi (Carroll, 2015; Fidalgo ve Scalón, 2010; Jin vd., 2015; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016) nedeniyle bu araştırmada da grupların yetenek dağılımları standart normal dağılımdan üretilmiştir.

Madde güçlük ve ayırt edicilik parametreleri için ise ilgili alanyazında yapılan simülasyon çalışmalarında referans grubu için farklı dağılımların ele alındığı tespit edilmiştir. Genellikle MTK’ya dayalı veri üretilen simülasyon çalışmalarında madde parametrelerinin dağılımları belirlenirken ulusal düzeydeki gerçek testlerden (Fidalgo ve Scalón, 2010; Finch, 2016) veya TIMSS gibi geniş ölçekli uygulamalardan (Carroll, 2015; Svetina ve Rutkowski, 2014) elde edilen güçlük ve ayırt edicilik parametre değerlerinin dikkate alındığı görülmüştür. Svetina ve Rutkowski (2014) yaptıkları araştırmalarında TIMSS 2007 uygulamasının dördüncü sınıf düzeyindeki madde parametrelerini kullanmış ve referans grup için ayırt edicilik parametresinin 0.46 ile 1.76 aralığında, güçlük parametresinin ise -2.00 ile 1.09 aralığında olmasını sağlayacak şekilde 15 madde üretmiştir. Finch (2016) ise çalışmasında eyalet çapında uygulanan ulusal ölçekli bir başarı testinin madde parametre değerlerini dikkate almış ve referans grup için ayırt edicilik parametresinin 0.84 ile 1.53 aralığında, güçlük parametresinin ise -1.06 ile 1.39 aralığında olacak şekilde 20 madde üretmiştir.

DMF belirleme çalışmalarında incelenen maddelerin referans gruptaki madde parametreleri ile kullanılan istatistiksel tekniklerin DMF belirleme gücü arasında ilişki olduğu pek çok araştırmacı (Ankenmann, Witt ve Dunbar, 1999; Clauser, Mazor ve Hambleton, 1994; Jiang ve Stout, 1998; Monahan ve Ankenmann, 2005) tarafından belirtilmiştir. Başka bir ifadeyle simülasyon çalışmalarında DMF’li olduğu kabul edilen maddelerin referans gruptaki madde parametrelerinin, ilgilenilen tekniklerin DMF belirleme performanslarını etkileyebildiği açıklanmıştır (Clauser vd., 1994). Böylelikle

DMF’li olarak dikkate alınacak maddelere karar verilmeden önce, referans grubun madde parametrelerinin incelenmesinin ve farklı güçlük ile ayırt edicilik parametrelerine sahip maddelerin DMF’li olarak belirlenmesinin önemli olduğu ifade edilmiştir (Muñiz vd., 2001). Bu bilgiler dikkate alınarak yapılan araştırma kapsamında hangi maddelerin DMF’li olacağına karar verirken, 20 maddenin referans gruptaki parametreleri incelenmiş ve DMF’li olan dört maddenin madde parametrelerinin birbirlerine yakın olması amaçlanmıştır. Böylelikle referans grubun madde parametrelerinin çoklu grup tekniklerinin performansları üzerindeki etkileri kontrol altına alınmaya çalışılmış ve dört maddenin farklı güçlük ve ayırt ediciliğe (kolay-yüksek ayırt edici; kolay-düşük ayırt edici; zor-yüksek ayırt edici; zor-düşük ayırt edici) sahip olmaları sağlanmıştır. Referans grubun madde parametreleri ile odak gruplar için DMF’li olarak belirlenen dört maddenin güçlük ve ayırt edicilik parametrelerine eklenen DMF miktarları Ek 7’de verilmiştir.

Ek 7 incelendiğinde DMF içermeyen 16 madde için referans ve odak grupların madde parametrelerinin aynı olduğu görülmüştür. Güçlük parametreleri -0.79 ile 1.00 aralığında, ayırt edicilik parametreleri ise 0.86 ile 1.31 aralığında olan DMF’li dört madde için ise odak grupların madde parametreleri değişimlenerek (Scott, 2014) belirlenen DMF miktarları (0.40, 0.70) ayırt edicilik ve güçlük parametrelerine eklenmiştir. Böylece dört maddede madde parametreleri farklılaştırılmış ve ilgili maddelerin referans ve odak gruplar arasında TB DMF, TBO(a) DMF ve TBO(ab) DMF göstermeleri sağlanmıştır.

Verilerin Üretilmesi

Araştırma kapsamında veri setleri üretilirken DMF ile ilgili alanyazındaki çalışmalar incelenmiş (Clark, 2010; Kim, 2010; Rollins III, 2018; Sandilands, 2014; Tay vd., 2016) ve bu çalışmalarda izlenen adımlar dikkate alınmıştır. Bu simülasyon çalışmasında ikili puanlanan ve tek boyutlu olarak üretilen verilerde çoklu gruplar için geliştirilen tekniklerin DMF’li maddeleri doğru belirleme performanslarının karşılaştırılmasına odaklanıldığından, veri üretilmesi sürecinde Birnbaum (1968) tarafından kuramsal temelleri ortaya konulan MTK’ya dayalı tek boyutlu 2PLM kullanılmıştır. Bu model yardımıyla öncelikle bireylerin yetenekleri ve madde parametrelerini içeren bir fonksiyon oluşturulmuştur (Lord, 1980).

Bireylerin yetenek (θ) parametreleri standart normal dağılıma $[N(0, 1)]$ ve ayırt edicilik $[a \sim U(0.84, 1.53)]$ ile güçlük $[b \sim U(-1.06, 1.39)]$ parametreleri ise belirtilen aralıklara göre tekdüze dağılıma dayalı olarak bir kere üretilmiş ve elde edilen değerler

bütün tekrarlar da kullanılmıştır. Böylece 2PLM çerçevesinde yetenek ile madde parametrelerine dayalı olarak her bir bireyin testteki ikili puanlanan maddeleri doğru yanıtlama olasılıkları hesaplanmıştır. Benzer şekilde yine her bir madde ve birey için tekdüze $U(0, 1)$ dağılım gösteren ve 0 ile 1 aralığında değerler alabilen rassal sayılar üretilerek, olasılık matrisi oluşturulmuştur. Sonrasında ise 2PLM'ye dayalı olarak elde edilen olasılıklar ile tekdüze dağılımdan elde edilen rassal sayılar karşılaştırılmıştır. Bu aşamada her bir hücredeki olasılık değeri, karşılaştırma matrisinde aynı hücre konumundaki rastgele değerden düşükse, bireyin bu maddeyi yanlış yanıtladığı kabul edilerek madde yanıtı için "0" değeri atanmıştır. Aksine olasılık değeri tekdüze dağılıma ilişkin matristeki aynı hücre konumuna karşılık gelen rastgele değerden yüksekse, bireyin bu maddeyi doğru yanıtladığı kabul edilerek madde yanıtı için "1" değeri atanmıştır. Böylelikle yapılan karşılaştırma işlemleri sonucunda her bir birey için 1 ve 0'lardan oluşan 20 maddelik veri seti üretilmiştir.

Yapılan simülasyon çalışmasında test uzunluğunun %20'sine karşılık gelen dört madde, DMF'li olacak şekilde belirlenmiş ve bütün tekrarlar da bu dört madde sabit olarak ele alınmıştır. Referans grubun belirlenen dört maddedeki güçlük ve ayırt edicilik parametrelerine, farklı DMF miktarları (0.40, 0.70) eklenerek, odak gruplar için bu dört maddedeki parametre değerleri farklılaştırılmıştır. Farklı odak grupların her biri için, referans grubun madde parametrelerine eklenen DMF miktarına bağlı olarak hem referans grubun odak gruplardan hem de odak grupların kendi aralarında avantajlı olmaları sağlanmıştır. Referans ve odak gruplar için örneklem büyüklükleri, DMF miktarları ve DMF türü değişimlendiğinden, yapılan her bir tekrarda bu faktörlerin düzeylerinin tüm olası kombinasyonlarını yinlemek amacıyla R yazılımında çeşitli fonksiyonlardan yararlanılarak veri üretimi gerçekleştirilmiştir.

Çalışmadaki tüm kodlar R yazılımında hazırlanmış ve dikkate alınan faktörler çerçevesinde simülasyon çalışması yapılandırılmıştır. Alanyazında yapılan çalışmalarda ANOVA (Analysis of Variance) sonuçlarının kolay ve doğru yorumlanabilmesi için simülasyon deseninin tamamen çaprazlanmış faktör desenine (fully crossed factorial design) göre oluşturulmasının önemli olduğu belirtilerek verilerin bu desene uygun üretilmesi önerilmiştir (Feinberg ve Rubright, 2016; Harwell vd., 1996; Skrandal, 2000). Bu bağlamda DMF ile ilgili birçok simülasyon çalışmasında olduğu gibi (Finch, 2005; Finch ve French, 2008; Frederickx vd., 2010; Goodman vd., 2011; Magis ve Facon, 2012; Svetina ve Rutkowski, 2014; Rockoff, 2018; Rollins III, 2018) yapılan bu araştırma kapsamında da tamamen çaprazlanmış faktöriyel desene dayalı olarak veri üretilmiştir.

Tamamen çaprazlanmış desende faktörlerin her bir kombinasyonu dikkate alınarak, bütün etkileşimlerin incelenebilmesi nedeniyle bu desene dayalı veri üretilmesinin birçok avantajının olduğu belirtilmektedir (Bulut ve Sünbül, 2017; Feinberg ve Rubright, 2016; Skrondal, 2000). Tamamen çaprazlanmış desene uygun planlanan simülasyon çalışmalarında faktör düzeylerinin birbirleriyle çaprazlanması gerektiğinden üretilen verilere ilişkin yapılan analiz sonuçlarının özetlenmesi ve sunumu daha kolaydır. Bu kapsamda Feinberg ve Rubright (2016) tarafından da açıklandığı üzere tamamen çaprazlanmış desene dayalı üretilen simülasyon sonuçlarının tablo ve grafiklerden yararlanılarak özetlenmesi uygun olmaktadır.

Simülasyon çalışmalarında veri üretirken önemli konulardan biri de doğru sonuçlar elde etmek için kaç tekrarın yapılması gerektiğidir (Feinberg ve Rubright, 2016; Harwell vd., 1996; Mundform vd., 2011). Simülasyon çalışmalarında birden çok veri seti üretildiğinden (Bulut ve Sünbül, 2017) tekrar sayısı az olan çalışmaların sonuçlarının geçersiz olabileceği, gerekenden daha fazla tekrar yapılmasının ise zaman ve kaynak israfına yol açabileceği belirtilmektedir (Feinberg ve Rubright, 2016; Mundform vd., 2011). Bu bağlamda ilgilenilen parametrelerin daha doğru ve güvenilir bir şekilde kestirilebilmesi için daha fazla tekrar yapmanın her zaman daha iyi olacağı düşünülmese de (Bulut ve Sünbül, 2017; Feinberg ve Rubright, 2016), simülasyondaki koşul sayısının, örneklem büyüklüğünün ve tekrar sayısının artmasına bağlı olarak simülasyonun işlem süresinin de uzayacağını göz ardı edilmemesi gerekmektedir (Bulut ve Sünbül, 2017; Hofert ve Mächler, 2016).

MTK tabanlı Monte Carlo simülasyon çalışmaları için en az 25 tekrarın yapılması önerilmesine rağmen, çalışmanın amacı ve belirlenen koşullar göz önünde bulundurularak tekrar sayısına karar verilmesi gerektiği vurgulanmaktadır (Harwell vd., 1996). İlgili alanyazın incelendiğinde çoklu gruplarla çalışılan, ancak az sayıda koşulun dikkate alındığı çalışmalarda 1000 tekrar (Fidalgo ve Scalón, 2010; Finch, 2016; Jin vd., 2015; Penfield, 2001) ve 500 tekrar (Battauz, 2019; Carroll, 2015; Woods vd., 2013) yapıldığına rastlanılmakla birlikte, 250 tekrarın (Rollins III, 2018) ve 100 tekrarın (Oshima vd., 2015; Schauburger ve Tutz, 2016; Svetina ve Rutkowski, 2014; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) yapıldığı çalışmaların da var olduğu görülmektedir. Sonuç olarak çoklu gruplarda DMF belirlemek amacıyla yapılmış olan simülasyon çalışmalarında farklı tekrar sayıları dikkate alınmasına rağmen genellikle 100 tekrarın yapıldığı belirlenmiştir. Nitekim Kim'in (2010) çalışmasında NCME'in 2009 yılındaki yayınlarında yaygın olarak kabul edilen tekrar sayısının 100 olduğu belirtilmiş

ve Tip I hata oranı açısından kararlı sonuçlar elde etmek amacıyla 100 tekrarın yeterli olduğu açıklanmıştır. Bu nedenle araştırma kapsamında koşul sayılarına bağlı olarak, çok sayıda veri üretilmesi gerektiğinden hem kararlı sonuçlar elde etmek hem de işlem süresinin artmamasını sağlamak amacıyla, her bir koşul için 100 tekrar yapılması uygun görülmüştür.

Verilerin Analizi

Simülasyon çalışmaları veri üretmeyi gerektiren bilgisayar deneyleri olarak kabul edilmektedir (Morris vd., 2019). Bu çalışmalar kapsamında dikkate alınan koşullara ve yapılan tekrar sayısına göre veriler üretilerek, bu veriler analiz edilmektedir (Sigal ve Chalmers, 2016). Veri üretilmesi sürecinde farklı programlardan yararlanılabilse de (Morris vd., 2019) psikometri alanındaki çalışmalarda R yazılım dilinin sıklıkla kullanılması nedeniyle, bu simülasyon çalışmasında da kodların tamamı R'ın 3.6.1 sürümünde (R Development Core Team, 2019) hazırlanmıştır.

Simülasyonun yapılandırılması, veri manipülasyonu, programlama, veri üretme, istatistiksel tekniklerin değerlendirilmesi ve grafiksel gösterim gibi birçok farklı amaç için kullanılabilen R, açık kaynaklı ve aynı zamanda gelişmiş bir yazılım dili olması nedeniyle araştırmacılar tarafından tercih edilmektedir. İstatistiksel paketlerin ücretsiz olarak kullanılmasına olanak sağlayan R yazılımı, özellikle MTK'ya dayalı simülasyon çalışmalarını tasarlamak ve uygulamak için yararlanılan en popüler yazılımlardan biri olarak kabul görmektedir (Bulut ve Sünbül, 2017; Morris vd., 2019; Rusch, Mair ve Hatzinger, 2013; Sigal ve Chalmers, 2016).

Yapılan bu simülasyon çalışmasında verilerin üretilmesi için R yazılımında yer alan çok sayıda paket içerisindeki fonksiyonlardan (Choi ve Asilkalkan, 2019; Li vd., 2012; Morris vd., 2019; Rusch vd., 2013) yararlanılmasına rağmen, veri analizi sürecinde genellikle *ltm 1.1-1* (Rizopoulos, 2018), *lme4 1.1-21* (Bates, Maechler, Bolker ve Walker, 2015), *difR 5.0* (Magis, Béland ve Raiche, 2018) ve *Lordif 0.3-3* (Choi vd., 2016) paketleri kullanılmıştır. Bu paketlerden *difR*'in 5.0 versiyonu içerisinde yer alan GLK, GLR-LRT, GLR-Wald, GMH teknikleri ile *Lordif*'in 0.3-3 versiyonu içerisinde yer alan Lordif tekniği kullanılarak toplam beş farklı çoklu grup tekniği ile DMF analizleri gerçekleştirilmiştir.

Magis ve diğerleri (2010) çalışmalarında ikili puanlanan maddeler için ikili ve çoklu gruplarda hem KTK hem de MTK'ya dayalı tekniklerle TB ve TBO DMF'yi tespit

edebilen çok sayıda tekniğin yer aldığı *difR* paketinin, araştırmacılar tarafından yaygın olarak kullanıldığını belirtmişlerdir. Ancak alanyazında sıklıkla kullanılan *difR* paketinin bazı dezavantajlarının olduğu da ifade edilmiştir. İlk olarak bu pakette yalnızca ikili puanlanan tekniklerin kullanılabilmesi önemli bir sınırlılık olarak görülmüştür. İkinci olarak ise bu paketteki MTK'ya dayalı teknikler için parametre kestirimlerinin *ltm* ve *lme4* paketleri aracılığıyla yapılıyor olmasının, analizlerin gerçekleşmesi için gerekli olan süreyi etkilediği bildirilmiştir. Özellikle madde sayısı fazla ve örneklem büyüklüğü geniş olan araştırmalarda, yani büyük veri setlerinde, MTK'ya dayalı DMF belirleme teknikleri kullanılarak yapılan analizler için gereken hesaplama süresinin arttığı açıklanmıştır (Magis vd., 2010). Svetina ve Rutkowski'nin (2014) çalışmalarında da benzer duruma dikkat çekilmiş ve çoklu grup çalışmalarında grup sayısı ile toplam örneklem büyüklüğünün artması nedeniyle hesaplama süresinin sorun oluşturabildiği vurgulanmıştır. Benzer şekilde *Lordif* paketinde de MTK'ya dayalı parametre kestirimleri için *ltm* paketinin kullanılması (Choi vd., 2011) ve analiz aşamasında istatistik sonucunu değerlendirmek için Monte Carlo simülasyon yaklaşımından yararlanılması gibi nedenlere bağlı olarak analizin tamamlanması için uzun zamana ihtiyaç olduğu belirtilmiştir. Nitekim birçok araştırmada da vurgulandığı üzere (Hallgren, 2013; Harwell vd., 1996; Sigal ve Chalmers, 2016; Skrondal, 2000; Wood, 2011) bu çalışma kapsamında da koşul sayısının fazla olmasına, örneklemelerin büyük olmasına, GLK ile Lordif tekniklerinde MTK'ya dayalı parametre kestirimleri için *ltm* paketinden yararlanılmasına ve 100 tekrar yapılmasına bağlı olarak çalışmanın kapsamı genişlemiş ve simülasyonun standart bir bilgisayar yardımıyla tamamlanabilmesi için hatırı sayılır bir zamana (yaklaşık 150 gün) ihtiyaç duyulduğu deneyimlenmiştir.

R yazılımının bu sınırlılığından yola çıkarak büyük ölçekli simülasyon (large-scale simulations) çalışmalarını gerçekleştirebilmek amacıyla işlevsel paketlerin geliştirildiği belirtilmiştir (Hofert ve Mächler, 2016). Bu bağlamda simülasyonun işlem sürecini hızlandırmak amacıyla *doParallel* 1.0.15 paketindeki (Ooi, Microsoft Corporation, Weston ve Tenenbaum, 2019) fonksiyonlar kullanılarak, paralel hesaplama yapılmasının simülasyonun süresini önemli ölçüde kısaltabileceği ifade edilmiştir (Alfons, Templ ve Filzmoser, 2010). Özellikle çok sayıda tekrarın gerçekleştirildiği simülasyon çalışmalarında paralel hesaplama yapmanın yararlı olduğu açıklanmıştır (Bulut ve Sünbül, 2017). Paralel hesaplama yapılan simülasyon çalışmalarındaki veri üretimi sürecinde yapılan işlemlerin, çok çekirdekli bir işlemci kullanılarak veya birkaç bilgisayarın bağlandığı bir ağdan yararlanılarak gerçekleştirilebileceği vurgulanmıştır

(Bulut ve Sünbül, 2017; Hofert ve Mächler, 2016). Bu bağlamda yapılan bu araştırmada simülasyonun çalışma süresini kısaltmak amacıyla *doParallel* paketindeki fonksiyonlardan yararlanılmış ve simülasyonların zamanında tamamlanabilmesi için TÜBİTAK tarafından kullanıma açılan bir ağ üzerinde paralel hesaplama yaptırılmıştır. Bu süreçte TÜBİTAK ULAKBİM Yüksek Başarım ve Grid Hesaplama Merkezi tarafından desteklenen Türk Ulusal Bilim e-Altyapısı (TRUBA) kullanılarak, Linux işletim sistemi içerisine yüklenen R yazılımının 3.6.1 sürümündeki ilgili paketler yardımıyla veri üretimi gerçekleştirilmiştir.

Değerlendirme Ölçütleri

Bu çalışmada istatistiksel tekniklerin performanslarını belirlemek için Tip I hata, istatistiksel güç ve Tip III hata oranları, parametre kestirimlerinin doğruluğunu değerlendirmek için ise mutlak yanlılık (bias) ve RMSE hesaplanmıştır. Araştırma kapsamında Tip I hata için yürütülen hesaplamalar DMF’li olmayan maddeler üzerinde, istatistiksel güç ve Tip III hata için yapılan hesaplamalar ise DMF’li maddeler üzerinde gerçekleştirilmiş ve çoklu grup tekniklerinin farklı koşullardaki performansları karşılaştırılmıştır. DMF ile ilgili bazı araştırmalarda olduğu gibi (Lei ve Li, 2013; Scott, Fayers, Aaronson, vd., 2009) yapılan bu simülasyon çalışmasında da istatistiksel güç .80’e eşit veya büyükse tekniklerin gücünün yüksek, .60 ile .80 arasında ise tekniklerin gücünün kabul edilebilir ve .60’tan küçükse tekniklerin gücünün yetersiz olduğu üç farklı ölçüt dikkate alınmıştır. İlgili araştırmalara (Clark, 2010; Finch, 2016; Hidalgo vd., 2014; Jin vd., 2015) benzer şekilde yapılan bu çalışmada da öncelikle testteki hiçbir maddenin DMF göstermediği durum için temel Tip I hata oranları hesaplanarak, elde edilen oranlar Bradley (1978) tarafından önerilen liberal kritere göre değerlendirilmiştir. Temel Tip I hata oranı için hesaplanan değerlerin .075’ten büyük olması durumunda istatistiksel gücün yorumlanmasının yanıltıcı olabilmesi nedeniyle (Lopez Rivas, 2012; Rockoff, 2018) istatistiksel güç oranları hesaplanmadan önce temel Tip I hata oranlarına bakılmıştır. Ayrıca DMF’li olarak belirlenen dört madde için madde parametrelerinin değişimlendiği koşullarda da Tip I hata hesaplanarak (Clark, 2010; Tay vd., 2013) iki farklı Tip I hata raporlanmıştır. Bu çalışmada hesaplanan Tip III hata oranları için ise Lopez Rivas’ın (2012) çalışmasına benzer şekilde .05 kritik değeri dikkate alınmış ve elde edilen oranlar Bradley (1978) tarafından önerilen liberal kritere göre değerlendirilmiştir. Öte yandan araştırma kapsamında güçlük ve ayırt edicilik

parametrelerine ilişkin kestirimlerin doğruluğu mutlak yanlışlık ve RMSE kriterlerine göre incelenmiş ve hem mutlak yanlışlık hem de RMSE'nin sıfır olması durumunda madde parametre kestirimlerinin hatasız, sıfıra yakın olması durumunda ise kestirimlerin son derece etkili olduğu şeklinde yorum yapılmıştır.

Fidalgo ve diğerleri (2004b) tarafından da belirtildiği gibi Tip I hata için belirlenen değerlendirme kriteri çalışmanın özelliğine bağlı olabilmektedir. Yapılan bu araştırma kapsamında, testin DMF'li madde içerdiği koşullar için saflaştırma işleminin yapılmaması nedeniyle eşleştirme puanlarının DMF'li maddelerden etkilenebilmesi mümkün olabilmektedir (Rollins III, 2018; Tay vd., 2013). Başka bir anlatımla, testte DMF'li maddelerin yer aldığı koşullarda saflaştırma işleminin yapılmamış olmasının Tip I hata oranlarını olumsuz etkilemesi beklenen durum olarak kabul edilmektedir.

DMF ile ilgili bazı araştırmalarda genel olarak Tip I ve Tip II hata oranlarının hesaplanmasının yeterli olduğu belirtilse de (Fidalgo vd., 2004a; Flowers, Oshima ve Raju, 1999), Socha ve diğerleri (2015) tarafından yapılan çalışmada istatistiksel güç ve Tip I hata oranları ile birlikte, parametre kestirimlerinin doğruluğunu kanıtlamak amacıyla mutlak yanlışlık ve RMSE kriterlerinin de hesaplanması gerektiği ifade edilmiştir. Bu kapsamda simülasyon sonuçlarını değerlendirmek amacıyla çoklu gruplarla ilgili yapılmış olan önceki çalışmalardan farklı olarak mutlak yanlışlık, RMSE, Tip I hata, istatistiksel güç (1 - Tip II hata) ve Tip III hata ayrı ayrı hesaplanarak değerlendirilmiştir. Böylece çoklu grup tekniklerinin performansları üç farklı kritere dayalı olarak karşılaştırılmış, simülasyonun geçerliği için ise mutlak yanlışlık ve RMSE hesaplanarak yapılan parametre kestirimlerinin doğruluğu test edilmiştir.

Yapılan bu simülasyon çalışmasında değişimlenen faktörlere ilişkin ana etki ve etkileşim etkilerinin dikkate alınabilmesi amacıyla tamamen çaprazlanmış faktör desenine göre veri üretilmiştir. Bağımlı değişkenlerin Tip I hata, istatistiksel güç, Tip III hata, mutlak yanlışlık ve RMSE, değişimlenen faktörlerin ise bağımsız değişkenler olduğu çalışmada simülasyon sonuçlarını analiz etmek amacıyla faktöriyel ANOVA tekniği kullanılmış ve ayrı ayrı beş analiz yapılmıştır. Yapılan analizlerde değişimlenen faktörlere ilişkin tüm ana etkilerin yanı sıra etkileşim etkileri de incelenmiş ve sonuçlar tablo halinde sunulmuştur. DMF ile ilgili yapılmış olan birçok çalışmada olduğu gibi (Carroll 2015; Carter ve Zickar, 2011; Clark, 2010; Huggins, 2014; Jin ve Chen, 2020; Socha vd., 2015) bu çalışma kapsamında da Faktöriyel ANOVA sonucundaki manidar farklılıkların yorumlanması için ana etkilerde eta kare, etkileşim etkilerinde ise kısmi eta kare etki büyüklüğü katsayıları hesaplanmış ve elde edilen katsayılar Gray ve Kinnear

(2012) tarafından belirlenen aralıklara göre yorumlanmıştır. Faktöriyel ANOVA sonucunda manidar bulunan etkileşimlerin grafikleri çizilerek sonuçlar özetlenmiştir. Böylelikle Tip I hata, istatistiksel güç, Tip III hata, mutlak yanlışlık ve RMSE değerleri için değişimlenen faktörlerden her birinin diğer faktörlerin düzeylerinde farklılaşım farklılaşmadığı tespit edilmiştir. Grafiklerin çiziminde ise *ggplot2* paketindeki (Wickham, 2009) fonksiyonlardan yararlanılmıştır.

Tip I Hata, İstatistiksel Güç ve Tip III Hata. Yapılan bu araştırmada sırasıyla Tip I hata, istatistiksel güç (1-Tip II hata) ve Tip III hata kavramları öncelikle istatistiksel anlamları, sonrasında ise DMF alanyazınındaki anlamları bağlamında detaylı bir şekilde açıklanmıştır. Alanyazında yapılmış olan çok sayıda çalışmada, genellikle araştırmacıların hipotez testlerinde ortaya çıkabilen Tip I hata (yanlış pozitif) ve Tip II hata (yanlış negatif) ile ilgilenmeleri nedeniyle bu iki hata türüne aşına oldukları vurgulanmaktadır (Mosteller, 1948; Tate, 2018; Yadav ve Korukonda, 1985). Bu nedenle de simülasyon çalışmalarında Tip I hata ve Tip II hata oranlarının hesaplandığı belirtilmektedir (Kleijnen, 1995; de Micheaux ve Tran, 2016). İstatistiksel gösterimi α olan Tip I hata, gerçekte doğru olan bir yokluk hipotezinin yapılan hipotez testi sonucunda yanlışlıkla reddedilmesi olarak ifade edilmektedir (Cohen, Manion ve Morrison, 2018; Glass, Peckham ve Sanders, 1972; Kleijnen, 1995; Mosteller, 1948). DMF açısından tanımlandığında ise; gerçekte DMF’li olmayan maddelerin, analizlerin gerçekleştirildiği istatistiksel teknikler tarafından yanlışlıkla DMF’li olarak belirlenmesi Tip I hata yapıldığını göstermektedir (Chun, 2014; Finch, 2016; Frederickx vd., 2010; Yao ve Li, 2015; Wang, 2016).

Simülasyon çalışmalarında DMF’li maddeler değişimlendiğinden, analizlerin öncesinde bu maddeler bilinmektedir. Bu nedenle DMF belirleme tekniklerinin performanslarını tespit etmek amacıyla hesaplanan Tip I hata oranı, bütün koşullarda yanlışlıkla belirlenen DMF’li maddelerin yapılan tekrar sayısına bölünmesi sonucunda oluşan yüzde değerine karşılık gelmektedir (Lee, Cohen ve Toro, 2009). Örneğin 1000 tekrar yapılan bir çalışmada gerçekte DMF’li olmayan bir maddenin istatistiksel teknik tarafından 48 defa DMF’li olarak belirlenmesi, bu teknik için hesaplanan Tip I hata oranının 0.048 olduğunu göstermektedir (Kim, 2010). Birçok araştırma kapsamında DMF belirleme teknikleri hakkında derin bir bakış açısı kazanmak ve daha ayrıntılı sonuçlar ortaya koyabilmek amacıyla testin tamamen DMF’li olmayan maddelerden oluştuğu ve

testin DMF'li madde içerdği iki farklı durum için Tip I hata oranlarının ayrı ayrı hesaplanmasının gerekli olduğu vurgulanmaktadır (Battauz, 2019; Kopf vd., 2015a; Langer, 2008; Lei ve Li, 2013).

Araştırmalarda dikkate alınan bir diğer hata türü ise Tip II hatadır. İstatistiksel gösterimi β olan Tip II hata gerçekte yanlış olan bir yokluk hipotezinin, yapılan hipotez testi sonucunda yanlışlıkla reddedilememesini ifade etmektedir (Cohen vd., 2018; Glass vd., 1972; Kleijnen, 1995; Mosteller, 1948). Böylece gerçekte DMF'li olan madde veya maddelerin, analizlerin gerçekleştirildiği teknik tarafından DMF'li olarak belirlenememesi Tip II hata olarak tanımlanmaktadır (Camilli, 2013; Frederickx vd., 2010; Meij, Kelderman ve van der Flier, 2010; Yao ve Li, 2015). Çalışmalarda Tip II hata yapma olasılığı ne kadar düşükse, tekniklerin DMF'li maddeleri doğru belirleme olasılığı o kadar yüksek olmaktadır (Yu vd., 2008). Bu bağlamda araştırmalarda Tip I hatanın yanı sıra Tip II hata yapmaktan da kaçınarak, bu hata türlerinin etkisini en aza indirmek amaçlanmaktadır (Ahmadi ve Bazvand, 2016; Cohen vd., 2018).

Tip II hata ve istatistiksel güç ters orantılıdır, yani Tip II hatanın azalması istatistiksel gücün arttığını göstermektedir (Wood, 2011). İstatistiksel gösterimi $1 - \beta$ olan istatistiksel güç (gerçek pozitif), gerçekte yanlış olan bir yokluk hipotezinin, yapılan hipotez testi sonucunda reddedilmesi olarak belirtilmektedir (Descôteaux, 2007; Glass vd., 1972). Böylelikle gerçekte DMF'li olan maddelerin analizin gerçekleştirildiği teknik tarafından doğru bir şekilde DMF'li olarak belirlenmesi istatistiksel güç olarak tanımlanmaktadır (Chun, 2014; Finch, 2016; Frederickx vd., 2010; Tay vd., 2016; Wang, 2016). Bütün koşullarda doğru belirlenen DMF'li maddelerin yapılan tekrar sayısına bölünmesi sonucunda oluşan yüzde değeri istatistiksel güç oranı olarak değerlendirilmekte ve hesaplanan oranların DMF belirleme tekniklerinin performanslarını tespit etmek amacıyla kullanıldığı belirtilmektedir (Cheng vd., 2013; Lee vd., 2009; Lopez Rivas, 2012).

İstatistiksel güç ne kadar yüksekse Tip II hata yapma olasılığının o kadar düşmesi nedeniyle, 0 ile 1 arasında değişebilen istatistiksel gücün mümkün oldukça 1.00'a yakın olması istenmektedir (Cohen vd., 2018, Descôteaux, 2007). Guilera ve diğerlerinin (2013) çalışmalarında Tip II hata ve istatistiksel güç için yürütülen hesaplamaların DMF'li maddeler üzerinde yapılması gerektiği belirtilmektedir. Bu nedenle yapılan bu simülasyon çalışmasında dört maddenin güçlük ve ayırt edicilik parametreleri odak gruplar için değişimlendiğinden, yapılan analizler sonucunda bu maddeler ya doğru bir şekilde DMF'li olarak belirlenmekte (istatistiksel güç) ya da yanlışlıkla DMF'li olarak

işaretlenememektedir (Cheng vd., 2013; Lopez Rivas, 2012). Bu anlamda Tip II hata ve istatistiksel gücün birbirlerinin tersi olması nedeniyle yapılan araştırma kapsamında istatistiksel güç oranlarının raporlanması uygun görülmüştür.

Alanyazındaki çalışmalarda Tip I ve Tip II hatalara ek olarak üçüncü bir hata türünün de olabileceği belirtilmiş (Mosteller, 1948) ve istatistikçilerin Tip III hata olarak isimlendirdikleri ilgi çekici başka bir hata türü daha tanımlanmıştır (Shaffer, 1995; Tate, 2018; Umesh, Peterson, McCann-Nelson ve Vaidyanathan, 1996). İstatistiksel gösterimi γ olan Tip III hata (Leventhal ve Huynh, 1996; Macdonald, 1999) çeşitli araştırmacılar tarafından bilimsel çalışmalarda dikkate alınarak, tarihsel süreçte farklı zamanlarda birkaç şekilde tanımlanmıştır (Leventhal, 1999; Tate, 2018). Ayrıca Tip III hatanın hangi şartlarda incelenmesi gerektiği birçok araştırmada tartışılmıştır (Leventhal, 1999; Leventhal ve Huynh, 1996; Rasch, 2012; Trzpiot, 2015). İlk kez Mosteller'in (1948, s. 61) çalışmasında incelenen Tip III hata “yokluk hipotezini yanlış bir nedenden dolayı doğru bir şekilde reddetmek” tanımlanmıştır. Sonrasında Hopkins (1973, s. 31) çalışmasında Tip III hatayı “hipotez testi bağlamında çıkarımsal bir yön hatası” olarak açıklamıştır. Son olarak Tip III hata “uygun olmayan bir istatistiksel teknik kullanıldığı için yanlış sonuca varılması” şeklinde ele alınmıştır (Carrier ve Wallace, 1989, akt., Umesh vd., 1996).

Yapılan bu tanımlara göre yanlış bir yokluk hipotezinin doğru bir şekilde reddedilmesinin yanı sıra etkinin yönünün yanlış belirlenmesi Tip III hata yapıldığı anlamına gelmektedir (Trzpiot, 2015). Gerçekte DMF'li olan maddelerin analizin gerçekleştirildiği teknik tarafından doğru bir şekilde DMF'li olarak belirlenmesine rağmen, bu maddelerin DMF türünün hatalı tanımlanması Tip III hata yapıldığı şeklinde yorumlanabilmektedir. Böylece testteki maddelerin gerçekte hangi madde parametresinden kaynaklı DMF gösterdiği ile istatistiksel tekniğin belirlemeyi amaçladığı DMF türü arasındaki uyumsuzluk Tip III hata yapılmasının altında yatan temel varsayım olarak dikkate alınmaktadır (Lopez Rivas, 2012).

TB ve TBO DMF'li maddelerin türünün DMF belirleme teknikleri tarafından yanlış tanımlanmasının mümkün olduğu ilk olarak Finch ve French (2008) tarafından yapılan çalışmada açıklanmıştır. Çalışmada TB DMF veya TBO DMF simüle edildiği durumlar için sırasıyla TBO veya TB DMF belirleme teknikleri kullanılmış ve DMF'li maddelerin hatalı olarak belirlenme sıklığı “anormal Tip I hata” (anomalous Type I error) olarak isimlendirilmiştir. Sonrasında ise Lopez Rivas (2012) tarafından yapılan çalışmada DMF türünün hatalı belirlenmesi, Tip III hata olarak yorumlanmıştır. Böylece Tip III hata

hesaplanarak gerçekte TB ve TBO maddelerin DMF türünün istatistiksel teknikler tarafından doğru belirlenip belirlenemediği ortaya konulmuştur (Finch ve French, 2008; Lopez Rivas, 2012). Tip III hatanın ilk kez dikkate alındığı DMF belirleme çalışması olan Finch ve French'in (2008) araştırmasından sonra bu hata türünün dikkate alındığı sınırlı sayıda araştırmaya daha rastlanılmıştır (Chun, 2014; Chun, Stark, Kim ve Chernyshenko, 2016; Lopez Rivas, 2012).

Araştırmacılar tarafından Tip III hatanın var olduğu bilinse de çalışmalarda yokluk hipotezinin kabulü veya reddine odaklanması nedeniyle, bu hata türünün genellikle göz ardı edildiği vurgulanmaktadır (Kimball, 1957; Mosteller, 1948; Yadav ve Korukonda, 1985). Bu anlamda bazen bilerek görmezden gelinen, bazı araştırmacılar tarafından ise varlığı inkâr edilen Tip III hatanın, çalışmalarda en az Tip I ve Tip II hata kadar yaygın olması nedeniyle araştırmacıların bu hata türüne dikkat etmeleri önerilmektedir (Yadav ve Korukonda, 1985). Özellikle DMF belirleme çalışmalarında gerçekte TB olan bir maddenin DMF belirleme tekniği tarafından TBO DMF'li olarak belirlenmesi durumunda, bu maddenin hangi gruba avantaj sağladığı hakkında doğru sonuçlara ulaşılamadığı belirtilmektedir (Finch ve French, 2008; Lopez Rivas, 2012).

Yapılan bu simülasyon çalışması kapsamında yalnızca güçlük parametresinden kaynaklı olan DMF'li maddeler TB DMF belirleme teknikleri tarafından, yalnızca ayırt edicilik parametresinden kaynaklı olan DMF'li maddeler ise TBO DMF belirleme teknikleri tarafından tespit edildiğinde bu durum istatistiksel güç olarak değerlendirilmiştir. Diğer taraftan yalnızca güçlük parametresinden kaynaklı olan DMF'li maddeler TBO DMF belirleme teknikleri tarafından, yalnızca ayırt edicilik parametresinden kaynaklı olan DMF'li maddeler ise TB DMF belirleme teknikleri tarafından tespit edildiğinde ise bu durum Tip III hata olarak ele alınmıştır. Son olarak hem güçlük hem de ayırt edicilik parametrelerinden kaynaklı olan DMF'li maddeler, TBO DMF türünde kabul edilmiş (Gao, 2019; Hidalgo ve López-Pina, 2004) ve bu maddelerin TBO DMF belirleme teknikleri tarafından tespit edilmesi istatistiksel güç, TB DMF belirleme teknikleri tarafından tespit edilmesi ise Tip III hata olarak dikkate alınmıştır.

İstatistiksel güç, Tip I hata ve Tip III hata tanımlarından anlaşıldığı üzere DMF çalışmalarında birincil amaç bir testte gerçekte DMF'li olan tüm maddelerin doğru tespit edilmesi, yine DMF'li olan maddelerin DMF türünün doğru belirlenmesi ve gerçekte DMF içermeyen maddelerin ise DMF'li olarak işaretlenmemesidir (Cohen vd., 1996; Lopez Rivas, 2012). Böylelikle iyi bir DMF belirleme tekniğinin yüksek istatistiksel güç

oranına ve düşük hata oranlarına (Tip I ile Tip III) sahip olması gerekmektedir (Wang, 2004). Yapılan analizler sonucunda Tip I ve Tip III hata oranları 0'a ne kadar yakın olursa hatanın o kadar düşük, istatistiksel güç oranları ise 1.00'a ne kadar yakın olursa tekniğin gücünün o kadar yüksek olduğu belirtilmektedir (Cohen vd., 2018). Detaylandırmak gerekirse istatistiksel güç oranlarının "1" ve Tip I ile Tip III hata oranlarının "0" olması durumunda DMF açısından incelenen maddelerin kullanılan istatistiksel teknik tarafından tamamen doğru veya hatasız bir şekilde belirlendiği, aksi durumda ise incelenen maddelerin DMF açısından hatalı değerlendirildiği sonucuna ulaşılmaktadır (Lee vd., 2009; Svetina ve Rutkowski, 2014; Wyse ve Mapuranga, 2009).

Mevcut araştırmalarda açıklandığı üzere DMF'li maddelerin varlığında bu maddelerin istatistiksel teknik tarafından tespit edilememesi veya bu maddelerin DMF türünün yanlış belirlenmesi çoğu zaman olasıdır. Öte yandan gerçekte DMF'li olmayan maddelerin de yanlışlıkla DMF'li olarak tanımlanması mümkün olmaktadır. Bu nedenle Tip I hata, istatistiksel güç ve Tip III hata oranlarını değerlendirmek için hangi kriterlerin dikkate alınacağı, kullanılan istatistiksel tekniklerin sonuçları üzerinde önemli bir etkiye sahip olabilmektedir (Descôteaux, 2007; Kim, 2010; Lopez Rivas, 2012).

İstatistiksel güç ile hata (Tip I, Tip III) oranlarının değerlendirilmesine ilişkin alanyazında farklı ölçütlerin kullanıldığı araştırmalara rastlanılmıştır. İlgili çalışmalar incelendiğinde Tip I ve Tip III hata oranlarını değerlendirmek için genellikle .05 (Cao vd., 2017; Finch, 2016; Jin vd., 2015; Lee ve Geisinger, 2016; Lopez Rivas, 2012; Tay vd., 2016) ve .075 (Fidalgo ve Scalón, 2010) kritik değerlerinin dikkate alındığı görülmüş ve tipik olarak Tip I hatanın .01, .05 veya .10 olabileceği açıklanmıştır. Zwick'in (2012) araştırmasında ise Tip I hata oranının .09'a kadar kabul edilebileceği belirtilmiştir. Öte yandan Tip I hata oranına benzer şekilde Tip III hata oranı da .05 olarak belirlenen bir kritik değer ile kontrol edilmiştir (Lopez Rivas, 2012). Aslında Tip I hatanın değerlendirilmesi için dikkate alınan kriterlerin farklılaşması Bradley (1978) tarafından yapılmış olan araştırmadan sonra gündeme gelmiştir. Bradley (1978) yapmış olduğu çalışmada hipotez testinin sağlamlılığını değerlendirmek amacıyla Tip I hata oranı için çeşitli tolerans kriterleri önermiştir. Böylelikle eğer α nominal (.05) olarak kabul edilirse liberal kriter ($.025 \leq p \leq .075$; $.05 \pm .025$) ve katı kriter ($.045 \leq p \leq .055$; $.05 \pm .005$) olmak üzere iki farklı simetrik aralığın dikkate alınabileceği belirtilmiştir. Başka bir ifadeyle Bradley'in yaklaşımında α değerine karar verildikten sonra kabul edilebilen Tip I hata miktarı için bir sapma aralığının oluşturulabileceği ifade edilmiştir (Robey ve Barcikowski, 1992).

DMF ile ilgili simülasyon çalışmalarında Tip I hata ile birlikte Tip II hatayı da kontrol etmek amaçlandığından, istatistiksel güç sonucunu değerlendirmek için de bir kritere ihtiyaç duyulmaktadır (Cohen vd., 1996). Alanyazındaki çalışmalarda istatistiksel güç oranı için .80 (Choi vd., 2011; Clark, 2010; Tay vd., 2016; Wang, 2004) veya .70 (Cuevas ve Cervantes, 2012; French ve Maller, 2007; González-Romá, Hernández ve Gómez-Benito, 2006) sınır değerlerinin dikkate alındığı anlaşılmaktadır. Diğer yandan Scott, Fayers, Aaronson ve diğerleri (2009) ile Lei ve Li (2013) tarafından yapılmış olan çalışmalarda .80 ve üzerindeki istatistiksel güç oranlarının yüksek (yeterli), .60 ile .80 arasındaki istatistiksel güç oranlarının kabul edilebilir, .60'ın altındaki istatistiksel güç oranlarının ise yetersiz olarak değerlendirildiği görülmektedir.

Araştırmalarda Tip I hata ve istatistiksel güç oranları yorumlanırken bu iki kriterin birbirleri ile ilişkili olduklarının unutulmaması gerekmekte ve birlikte değerlendirilmeleri önerilmektedir (Cheng vd., 2013; Rockoff, 2018). Tip I hatanın yüksek olduğu durumda DMF'li olmayan maddelerin dahi kolaylıkla istatistiksel teknik tarafından DMF'li olarak belirlenmesi söz konusu olabilmektedir. Bu nedenle istatistiksel gücün sahte (spurious) olabilmesi nedeniyle dikkatli yorumlanması gerektiği çok sayıda araştırmacı tarafından vurgulanmaktadır (Feinberg ve Rubright, 2016; Gao, 2019; Lopez Rivas, 2012; Morris vd., 2019). Finch'in (2005) çalışmasında belirtildiği gibi temel Tip I hata oranının .10'un üzerinde olduğu simülasyon koşullarında istatistiksel gücün yorumlanması anlamlı değildir. Böylece Tip I hata, ve istatistiksel güç ve Tip III hata oranlarının beklendiği gibi kontrol edilmesi, DMF tekniklerinin etkili olduğunu göstermektedir (Chun, 2014; Lan, 2014). Aksine hataların belirlenen seviyeden yüksek, istatistiksel gücün ise kabul edilebilir düzeyden düşük olması durumunda, kullanılan DMF belirleme tekniği ile elde edilen sonuçların geçerliği sorgulanabilmektedir.

Yanlılık ve RMSE. Simülasyon çalışmalarında MTK'ya dayalı veri üretilmesi durumunda, sonuçların kararlılığının önemi nedeniyle kestirilen parametrelerin doğruluğuna (item parameter recovery) ilişkin çalışmaların yapılması önerilmektedir (Bolt, Cohen ve Wollack, 2002; Wen, 2014). Bu anlamda simülasyona dayalı çalışmalarda işlemlerin beklendiği gibi gerçekleştiğinden, dikkate alınan MTK modeline dayalı olarak yapılan kestirimlerin doğru ve sonuçların makul olduğundan emin olmak için, araştırmacıların madde parametrelerine ilişkin kestirim hatalarını kontrol ederek sonuçların doğru olduğunu kanıtlamaları gerekmektedir (Sandilands, 2014; Sargent,

2013; Tay vd., 2015). Arařtırmalarda gerek madde parametreleri ile kestirilen madde parametreleri arasındaki farkın miktarı, kestirimin doęruluęu hakkında bilgi vermektedir (Feinberg ve Rubright, 2016; Sachse, Roppelt ve Haag, 2016). Bařka bir ifadeyle her bir tekrar iin parametre kestirimleri ile gerek parametre deęerleri arasındaki farka gre parametre kestirimlerinin doęruluęu deęerlendirilebilmektedir (Harwell vd., 1996).

Parametre kestirimlerinin doęruluęunu tespit etmek amacıyla, ilgili alanyazında yapılmıř olan alıřmalar incelendięinde yanlılık, mutlak yanlılık (absolute bias), RMSE, kestirilen ve gerek parametreler arasında hesaplanan korelasyon katsayısı, Standart Hata (standart error) ve Oran Hatası (proportion error) gibi farklı deęerlendirme kriterlerinin hesaplandıęı grlmektedir (Atar, 2007; Bilir, 2009; Cho vd., 2016; DeMars ve Lau, 2011; Harwell vd., 1996; Luo ve Liang, 2019; Sandilands, 2014; Socha vd., 2015; Strobl vd., 2015; Wen, 2014, Yuan, Tong ve Zhang, 2015). Alanyazındaki alıřmalarda ilgili kriterlerden birkaının birlikte deęerlendirilmesi gerektięi belirtilmekte (Feinberg ve Rubright, 2016; Morris vd., 2019; Yuan vd., 2015) ve genellikle yanlılık ve RMSE'nin birlikte dikkate alındıęı gzlemlenmektedir (Harwell, 2018; Harwell vd., 2018). Alanyazında yanlılık ve RMSE kriterlerinin yaygın olarak dikkate alınmasına baęlı olarak (DeMars, 2003; Sigal ve Chalmers, 2016), yapılan bu arařtırma kapsamında da parametre kestirimlerinin doęruluęunu (tutarlılıęını/kararlılıęını) incelemek amacıyla yanlılık ve RMSE'nin birlikte deęerlendirilmesine karar verilmiřtir.

Genel olarak hesaplanan RMSE deęerleri parametre kestirimindeki toplam hatayı, yanlılık deęerleri ise sistematik hatayı gstermektedir (Atar, 2007; Feinberg ve Rubright, 2016; Luo ve Liang, 2019). Yanlılık, gerek ve kestirilen parametreler arasındaki farkın ortalama bir lsdr. Bu anlamda kestirilen ve gerek parametreler arasındaki fark, yapılan kestirimlerin bir miktar hata ierdięini ortaya koymaktadır. Kestirilen parametre gerek parametre deęerinden byk olduęunda pozitif yanlılık, kk olduęunda ise negatif yanlılık sz konusu olmaktadır (Feinberg ve Rubright, 2016). Sistematik olarak ok yksek veya dřk kestirim deęerleri yanlılıęın varlıęına iřaretken, kestirilen ve gerek parametreler arasındaki farkın sıfır olması yanlılıęın olmadıęı anlamına gelmektedir (Morris vd., 2019; Sigal ve Chalmers, 2016).

Yanlılık deęerlerinin hem negatif hem de pozitif olabilmesi nedeniyle yanlılık iin ortalama hesaplandıęında negatif deęerlerin ortalamayı etkilemesi sz konusu olabilmektedir. Bu sebeple Atar (2007) ve Bilir (2009) alıřmalarında yanlılık deęerlerinin karesini almıřlardır. Benzer řekilde negatif yanlılık deęerlerinin ortalamayı etkilememesi iin arařtırmalarda mutlak yanlılık kriterinin dikkate alınması gerektięi

önerilmiştir (Yuan vd., 2015). Yanlılık kriterinin aksine mutlak yanlılık hesaplandığında kestirilen ve gerçek parametreler arasındaki fark sıfır veya pozitif olabilmektedir (Wang ve Chen, 2005; Woods, 2009b; Woods ve Grimm, 2011). Bu durumda ise ANOVA tekniği kullanılarak simülasyonda dikkate alınan faktörlerin ana ve etkileşim etkileri test edilebilmektedir (Masyn, 2017; Yuan vd., 2015). Mutlak yanlılığa benzer şekilde RMSE kriteri için de hesaplanan değerlerin yalnızca pozitif olabilmesi nedeniyle, RMSE'nin sıfıra yakın ve küçük olması parametre kestirimlerinin doğru olduğu anlamına gelmektedir (Sigal ve Chalmers, 2016; Strobl vd., 2015).

Yapılan bu simülasyon çalışmasında hem güçlük hem de ayırt edicilik parametreleri için parametre kestirimlerinin doğruluğu incelenmiştir. Belirlenen her bir koşulda yapılan 100 tekrar için 2PLM'ye dayalı olarak kestirilen güçlük ve ayırt edicilik parametreleri ile gerçek parametreler arasında bir farklılık olup olmadığını değerlendirmek amacıyla, mutlak yanlılık ve RMSE hesaplanmıştır. Yanlılık, mutlak yanlılık ve RMSE için hesaplama formülleri Eşitlik 19, 20 ve 21'de verilmiştir.

$$Yanlılık (Bias) = \frac{\sum_{t=1}^R (\hat{\tau}_{jt} - \tau_j)}{R} \quad (19)$$

$$Mutlak Yanlılık (Absolute Bias) = \frac{\sum_{t=1}^R |\hat{\tau}_{jt} - \tau_j|}{R} \quad (20)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^R (\hat{\tau}_{jt} - \tau_j)^2}{R}} \quad (21)$$

Eşitlik 19, 20 ve 21'de R, yapılan tekrar sayısını; τ , gerçek parametre değerlerini ve $\hat{\tau}$ ise kestirilen parametre değerlerini göstermektedir (Morris vd., 2019). Ayrıca eşitliklerdeki τ_j , her bir madde için gerçek parametre değerlerini; $\hat{\tau}_{jt}$ ise her bir madde için kestirilen parametre değerlerini belirtmektedir (Atar, 2007; Bilir, 2009; DeMars, 2003; Sandilands, 2014). Mutlak yanlılık, kestirilen ve gerçek parametre değeri arasındaki mutlak farkın toplamının tekrar sayısına bölünmesi sonucunda elde edilirken; RMSE, kestirilen ve gerçek parametre değeri arasındaki farkın karelerine ilişkin toplamın tekrar sayısına bölünmesi ile elde edilen değerin karekökü alınarak hesaplanmaktadır. Eşitliklerdeki formüllerden anlaşılabileceği üzere gerçek parametre değerleri kestirilen parametre değerlerinden büyük olduğunda yanlılık pozitif veya negatif olabilirken, mutlak yanlılık ve RMSE kriterlerine ilişkin sonuçlar yalnızca pozitif olabilmektedir (Wang ve Chen, 2005; Woods ve Grimm, 2011; Yuan vd., 2015).

Ana Etkiler ile Etkileşimler için Faktöriyel ANOVA. Simülasyon çalışması kapsamında değişimlenen faktörlerin ana etkileri ile bu faktörler arasındaki etkileşim etkileri için manidarlığın incelenmesi amacıyla Tip I hata, istatistiksel güç, Tip III hata, mutlak yanlılık ve RMSE sonuçları Faktöriyel ANOVA kullanılarak analiz edilmiştir. Alanyazında yer alan çok sayıdaki çalışmada da (Feinberg ve Rubright, 2016; Harwell vd., 2018; Harwell vd., 1996; Yuan vd., 2015) simülasyon sonuçlarını analiz etmek için Faktöriyel ANOVA yapılması önerilmekte ve DMF belirleme tekniklerinin performansları ile madde parametre kestirimlerinin doğruluğu üzerinde etkili olan faktörlerin belirlenmesi amacıyla Faktöriyel ANOVA yapıldığı görülmüştür (Atar, 2007; Carroll, 2015; Lee ve Geisinger, 2016; Lopez Rivas, 2012; Meade, Lautenschlager ve Johnson, 2007; Rockoff, 2018; Seybert ve Stark, 2012; Wood, 2011). Faktöriyel ANOVA sonuçlarının ise tek bir sunum şeklinin olmaması nedeniyle anlaşılmayı kolaylaştıracak şekilde tablo ve şekiller (koşullar arasındaki bütün farklı etkileşimlerin yer aldığı grafikler) yardımıyla sonuçlar özetlenmiştir.

Faktöriyel ANOVA’da gözlemlerin bağımsız, varyansların homojen ve dağılımların normal olması gibi çeşitli varsayımlara sahip olan Faktöriyel ANOVA’da, belirtilen varsayımların incelenmesi yönünde (Gravetter ve Wallnau, 2012; Field, 2009; Howell, 2014; Kirk, 2008) genel bir fikir birliğinin var olduğu bilinmektedir. ANOVA’da yapılan F testinin normallik ihlaline karşı sağlamlığı sürekli araştırılmasına rağmen birbirinden farklı sonuçların elde edildiği görülmüştür. Alanyazında yapılmış olan birkaç simülasyon çalışmasında (Blanca, Alarcon, Arnau, Bono ve Bendayan, 2017; Knief ve Forstmeier, 2021) ANOVA modellerinin bazı koşullarda normallik varsayımının ihlallerine karşı sağlam olduğu belirtilmesine rağmen bazı araştırmacılar tarafından da (Bishara ve Hittner, 2012; Harwell vd., 1996; Puth, Neuhäuser ve Ruxton, 2014) dağılımların normallikten sapma gösterdiği durumlarda analiz sonuçlarının geçerliğinin azalabilmesi nedeniyle çeşitli veri dönüştürme yaklaşımlarına başvurularak dağılımların normalleştirilmesi gerektiği vurgulanmıştır. Bu nedenle analizlerden önce normallik ve varyansların homojenliği varsayımları kontrol edilmiştir.

Faktöriyel ANOVA’da varyansların homojenliği Levene testi kullanılarak incelenmiştir. Araştırma kapsamında örneklemelerin büyük olmasına bağlı olarak yapılan ANOVA’larda varyansların homojenliği varsayımının sağlandığı görülmüştür. Normallik varsayımı ise çarpıklık ve basıklık katsayıları hesaplanarak veya Kolmogrov-Smirnov ile Shapiro-Wilk testleri gibi istatistiksel yöntemler kullanılarak test edilmiştir (Hair, Black, Babin ve Anderson, 2019; Howell, 2014). George ve Mallery (2020) çarpıklık ve basıklık

katsayılarının ± 1 aralığında olması durumunda dağılımın normal, ± 2 aralığında olması durumunda ise kabul edilebilir düzeyde normal olduğunu belirtmişlerdir. Bu bağlamda yapılan çalışma kapsamında çarpıklık ve basıklık katsayılarının ± 1 aralığında olduğu durumlar için verinin normal dağıldığı kabul edilmiş, aksi durumlarda ise dağılımın normal olmadığı düşünülerek veri dönüştürme yoluna gidilmiştir. Bu süreçte kare, karekök, logit, probit, logaritmik, arcsin ve ters çevirme yöntemleri arasından verinin yapısına uygun olan tercih edilerek, verinin yapısı değiştirilmeden farklı bir birime dönüşüm yapılmıştır (Blanca vd., 2017; Hair vd., 2019; Stevens, 2009; Tabachnick ve Fidell, 2013).

Faktöriyel ANOVA tekniği için yapılan analizler haricinde önemli olan başka bir konu ise etki büyüklüğünün hesaplanmasıdır. Analizler sonucunda elde edilen p değeri ile bir etkinin varlığı belirlenmesine rağmen, bu etkinin büyüklüğü tam olarak bilinmemektedir. Bu nedenle p değeri ile birlikte örneklem büyüklüğünden bağımsız bir etki büyüklüğü katsayısının raporlanması gerekmektedir (Ellis ve Steyn, 2003; Sullivan ve Feinn, 2012). Özellikle büyük örneklerde istatistiksel analiz sonuçlarını raporlarken, gruplar arasında gözlenen farklılıkların istatistiksel manidarlığının yanı sıra sonuçların pratik açıdan da anlamlı olup olmadığını gösterecek etki büyüklüğü katsayılarının yorumlanması önemli görülmektedir (Clark ve LaHuis, 2012; Fan, 2001).

Alanyazında çok sayıda etki büyüklüğü katsayısı geliştirilmiş olmasına rağmen (Lakens, 2013; Olejnik ve Algina, 2000, 2003), Faktöriyel ANOVA yapılan çalışmalarda genellikle eta kare ve kısmi eta kare katsayılarının tercih edildiği belirtilmektedir (Fay ve Boyd, 2010; Fritz, Morris ve Richler, 2012; Norouzian ve Plonsky, 2018). Eta kare etki büyüklüğü katsayısının hesaplanabilmesi için gerekli olan formül Eşitlik 22’de verilmiştir.

$$\eta^2 = \frac{\text{Kareler Toplamı}_{\text{faktöretkisi}}}{\text{Kareler Toplamı}_{\text{toplam}}} \quad (22)$$

Eşitlik 22’de eta kare etki büyüklüğü katsayısının formülü yer almaktadır. Eta kare katsayısı, bütün faktörler için toplanan kareler toplamının bir yüzdesine karşılık gelmekte ve 0 ile 1 arasında değişiklik göstermektedir (Lakens, 2013; Levine ve Hullett, 2002). Bu formüle göre bağımsız değişken bağımlı değişkendeki varyansın tamamını açıkladığında 1 değerini almaktadır. Eta kare değerleri varyansın farklı bileşenlerine karşılık geldiğinden toplanabilmektedir (Richardson, 2011). Eta karenin yanı sıra sıklıkla hesaplanan diğer bir etki büyüklüğü katsayısı kısmi eta karedir. Kısmi eta kare katsayısının hesaplanabilmesi için gerekli formül Eşitlik 23’te verilmiştir.

$$\eta_p^2 = \frac{\text{Kareler Toplamı}_{\text{faktöretkisi}}}{\text{Kareler Toplamı}_{\text{faktöretkisi}} + \text{Kareler Toplamı}_{\text{hata}}} \quad (23)$$

Eşitlik 23'ten görüldüğü üzere kısmi eta kare bir ANOVA çalışmasında ilgilenilen faktöre ilişkin kareler toplamının, ilişkili hata varyansı ile faktörün kendi kareler toplamına bölünmesi sonucunda elde edilen katsayıdır (Brown, 2008). Bu formüle göre kısmi eta karede, diğer faktörlerin etkisi paydadana çıkarılmaktadır (Bakeman, 2005). Böylelikle kısmi eta kare katsayısı için, analizde yer alan tüm etkilerin değerlendirmeden çıkarılması sonucunda ilgilenilen faktöre ilişkin etki büyüklüğü hesaplanabildiğinden kısmi eta karelerin toplamı 1'i geçebilmektedir (Fritz vd., 2012; Norouzian ve Plonsky, 2018; Richardson, 2011). Bu bilgilere dayalı olarak Faktöriyel desenlerde eta-kare ve kısmi eta karenin farklı amaçlara hizmet ettiği (Norouzian ve Plonsky, 2018) ve kısmi eta karenin, eta kareden genellikle daha yüksek olduğu belirtilmektedir (Fritz vd., 2012; Levine ve Hullett, 2002; Pierce, Block ve Aguinis, 2004).

Alanyazında genellikle tek yönlü ANOVA'da eta kare katsayısının, çok yönlü ANOVA'da ise kısmi eta kare katsayısının tercih edildiği ifade edilmiştir (Levine ve Hullett, 2002; Norouzian ve Plonsky, 2018). Kısmi eta kare katsayısının kullanılması durumunda hesaplanan etki büyüklüğü katsayılarına dayalı olarak hem farklı çalışmalar arasında karşılaştırma yapılabileceği (Fay ve Boyd, 2010; Keppel ve Wickens, 2004; Olejnik ve Algina, 2003), hem de aynı araştırmadaki çok sayıda bağımsız değişkenin etkilerini tespit etmek amacıyla kısmi eta kare katsayısının yorumlanabileceği belirtilmiştir (Lakens ve Caldwell, 2019; Richardson, 2011). Bu bilgilere dayalı olarak eta karenin bir çalışma içindeki, kısmi eta karenin ise çalışmalar arasındaki etki büyüklüklerinin karşılaştırılması için daha kullanışlı olduğu anlaşılmıştır (George ve Mallery, 2020). Öte yandan Fritz ve diğerleri (2012) tarafından yapılan araştırmada faktöriyel desenlerde kısmi eta kare yerine eta karenin kullanılmasının daha uygun olabileceği açıklanmıştır. Norouzian ve Plonsky (2018) tarafından yapılan çalışmada ise kısmi eta kare değerlerinin eta kare değerlerinden yüksek olması nedeniyle kısmi eta kare katsayısının tek başına rapor edilmesinin hatalı yorumlara neden olabileceği ve ilgili katsayılar yorumlanırken dikkat edilmesi gerektiği vurgulanmıştır. Böylece yapılan çeşitli simülasyon çalışmalarında farklı koşullar için etki büyüklüğü katsayılarına ilişkin sonuçların değişebildiği ortaya konulmuş (Okada, 2013; Skidmore ve Thompson, 2013) ve araştırmacıların kısmi eta kareye ek olarak eta kareyi de raporlamalarının uygun olduğu sonucuna ulaşılmıştır (Levine ve Hullett, 2002).

Hem betimsel düzeyde yapılmış hem de simülasyona dayalı olan birçok araştırmanın önerileri dikkate alındığında (Fritz vd., 2012; Keppel ve Wickens, 2004), hangi etki büyüklüğü katsayısının yorumlanmasının doğru olacağına karar verilirken, çalışmanın amacının ve koşullarının birlikte dikkate alınmasının önemli olduğu anlaşılmış ve yapılan bu çalışmada ana etkiler için eta kare, etkileşim etkileri içinse kısmi eta kare katsayılarının yorumlanmasına karar verilmiştir.

Eta kare katsayısının değerlendirilmesi için araştırmalarda genellikle Cohen (1988) tarafından yapılan kuramsal çalışmada önerilen ölçütlerin dikkate alındığı görülmüştür. Cohen (1988) çalışmasında küçük, orta ve büyük etkilerin sırasıyla .10, .25 ve .40'a eşit f değerlerine karşılık geldiğini bulmuş ve bu f değerlerine dayalı hesaplanan eta katsayılarını sırasıyla .100, .243 ve .371 olarak hesaplamıştır. Eta katsayılarının karesi alınarak elde edilen eta kare katsayılarının yaklaşık değerleri ise sırasıyla .01 (düşük), .06 (orta) ve .14 (yüksek) olarak belirlenmiştir (ss. 283-287). Richardson'ın (2011) yaptığı araştırmasında kısmi eta kare sonuçlarının da Cohen'in (1988) ölçütlerine dayalı olarak değerlendirilebileceği ifade edilmiş ve kısmi eta karenin yorumlanması için; $.01 \leq \eta_p^2 < .06$ (düşük), $.06 \leq \eta_p^2 < .14$ (orta) ve $\eta_p^2 \geq .14$ (yüksek) olmak üzere üç farklı değer aralığı tanımlanmıştır (Gray ve Kinnear, 2012). Finch ve French (2015) tarafından yapılan bir çalışmada, istatistiksel olarak manidar bulunan faktörler arasından yalnızca .10'dan büyük olan eta kare katsayılarının yorumlanması gerektiği açıklanmasına rağmen, bu araştırma kapsamında hesaplanan katsayıların Gray ve Kinnear (2012) tarafından belirlenen ve alanyazında genel olarak kabul edilen aralıklara göre değerlendirilmesine karar verilmiştir.

Simülasyonun Geçerliği

Bilgisayar teknolojisindeki gelişmelerle birlikte simülasyon çalışmaları sağlık bilimleri, davranış bilimleri, eğitim bilimleri ve psikometri alanlarındaki araştırmalarda giderek daha önemli bir rol oynamaktadır (Berends ve Romme, 1999; Feinberg ve Rubright, 2016; Gilbert, 1999; Gu, 2016; Hallgren, 2013; Harwell vd., 2017; Morris vd., 2019; Sigal ve Chalmers, 2016; Tureson ve Odland, 2018). Bu tür çalışmalarda bilgisayar yazılımları kullanılarak yazılan kodlar yardımıyla veri üretilmesi nedeniyle simülasyon çalışmaları gerçek durumların taklit edilebildiği deneysel araştırmalar olarak da kabul edilebilmektedir (Harwell vd., 1996; Morris vd., 2019; Tureson ve Odland, 2018).

Özellikle yeni geliştirilmiş farklı teknik, model veya yaklaşımların performanslarını karşılaştırmak (Hallgren, 2013; Harwell, Rubinstein, Hayes ve Olds, 1992; Harwell vd., 1996; Sigal ve Chalmers, 2016) veya parametre kestirimlerinin doğruluğunu değerlendirmek (Harwell vd., 1996) amacıyla gerçekleştirilen simülasyon çalışmalarında, deneysel kontrolün sağlanabilmesi önemli bir avantaj olarak görülmektedir.

Genel olarak simülasyon sonuçlarının dikkate alınabilmesi için önce yapılan hesaplamaların doğru, güvenilir ve geçerli olduğunun kanıtlanması gerekmektedir (Bratley, Fox ve Schrage, 1987; Davis vd., 2007; Law, 2003; Wood, 2011). Başka bir ifadeyle gerçekleştirilen deneysel işlem sonuçları analiz edilmeden önce (Davis vd., 2007) yazılan kodların doğruluğunun ve simülasyonun geçerliğinin (Feinberg ve Rubright, 2016; Harwell vd., 1996; Kleijnen, 1995; Law, 2003; Morris vd., 2019; Sargent, 2013) ispatlanması son derece önemli görülmektedir. Çünkü simülasyonlarda yalnızca sürecin hassasiyetle takip edilmesi durumunda sonuçların dikkate değer olabileceği ifade edilmekte, aksi durumda ise elde edilen sonuçlara dayalı olarak yapılan yorumların yanıltıcı olabileceği vurgulanmaktadır (Davis vd., 2007; Law, 2003; Spence, 1983).

Simülasyon çalışmalarında bilimsel ve gerçek bilgilere ulaşmak istenildiğinde simülasyon sonuçlarına ilişkin geçerlik çalışmasının farklı yöntemler kullanılarak yapılması gerektiği (Berends ve Romme, 1999) ve genellikle görünüş geçerliği ile istatistiksel geçerlik olmak üzere iki temel geçerlik kanıtı üzerinde durulduğu belirtilmektedir (Sandilands, 2014; Scott, 2014). Bu çalışmalarda görünüş geçerliğinin oldukça gerekli ancak öznel olması nedeniyle simülasyonun geçerliğini kanıtlamak için tek başına yeterli olmadığı açıklandığından, istatistiksel geçerliğin de sağlanması gerekmektedir. Böylelikle simülasyonun geçerliğinden emin olmak ve elde edilen sonuçlara güvenilip güvenilmemesine karar vermek amacıyla her iki geçerlik türü için de kanıt sunulup bütün kanıtların birlikte değerlendirilmesi önemli görülmektedir (Chung, 2004).

Alanyazındaki bu açıklamalardan anlaşıldığı üzere simülasyon çalışmalarında yapılan hesaplamaların, simülasyonun temelinde yatan kuramsal özelliği doğru temsil etmesi (Davis vd., 2007) gerekmesine rağmen Harwell ve diğerleri (2018) yaptıkları araştırmalarında 6 dergide 1985-2012 yılları arasında simülasyonla ilgili yayımlanmış olan 677 makaleyi incelemişler ve üretilen verinin istenen özelliklere sahip olduğunu kanıtlayan çalışma sayısının yalnızca 15 (%2.2) olduğunu belirlemişlerdir. Ulaşılan bu sonuç simülasyon çalışmalarında genellikle yazılan bilgisayar kodlarının ve yapılan hesaplamaların doğru kabul edildiğini göstermiştir (Wood, 2011).

Yapılan bu simülasyon çalışması kapsamında verilerin doğru üretilip, simülasyonun öngörüldüğü şekilde ve gerçeğe yakın sonuçlar ortaya koyduğundan emin olmak için, öncelikle DMF çalışması kapsamında simülasyonun geçerliğini araştıran çalışmalar detaylıca incelenmiştir (Bratley vd., 1987; Hansen, 2017; Lopez Rivas, 2012; Sandilands, 2014; Scott, 2014). Genel olarak benzer adımların izlendiği bu çalışmalarda olduğu gibi, yapılan araştırma kapsamında da öncelikle tek bir örnek veri seti ile grup sayısının üç ve dört olduğu durumlar için üretilen nihai veri setleri üzerinde birtakım kontroller yapılarak görünüş geçerliğine kanıt sunulmaya çalışılmıştır. Sonrasında ise istatistiksel geçerlik için güçlük ve ayırt edicilik parametrelerine ilişkin kestirimlerin doğruluğu araştırılmıştır.

Görünüş Geçerliği

İlgili alanyazında görünüş geçerliğinin incelenmesi amacıyla yaygın olarak iki aşamalı bir sürecin takip edildiği görülmüştür. İlk aşamada bir tekrara dayalı üretilen örnek bir veri seti üzerinde deneme yapılması önerilmiştir. Deneme çalışması sürecinde sırasıyla kodların gözden geçirilmesi (Davis vd., 2007; Kleijnen, 1995; Law, 2003; Li vd., 2012), simülasyon gerçekleşirken hata uyarısı verilip verilmediğinin tespit edilmesi (Sandilands, 2014), simülasyonun tamamlanma süresinin belirlenmesi (Law, 2003), üretilen veri setinde eksik veya hatalı değer olup olmadığının kontrol edilmesi (Lopez Rivas, 2012), farklı DMF miktarları için veri seti üzerinde referans ve odak grupların madde parametrelerinin incelenmesi (Sandilands, 2014; Scott, 2014) ve DMF’li olan veya olmayan maddelerin güçlük ve ayırt edicilik parametrelerine ilişkin MKE’lerin çizdirilerek grafiksel kontrollerin yapılması (Lopez Rivas, 2012) gibi çeşitli işlemlerin gerçekleştirildiği belirlenmiştir. İkinci aşamada ise 100 tekrar sonucunda üretilen nihai veri setlerindeki her bir koşul için satır ve sütun sayılarının incelenmesi (Sandilands, 2014) ve elde edilen bulguların makul ve alanyazındaki çalışmalarda ortaya konulan sonuçlarla tutarlı olup olmadığının belirlenmesi (Davis vd., 2007; Feinberg ve Rubright, 2016; Li ve Zumbo, 2009; Paxton vd., 2001) gibi incelemelerin yapıldığı tespit edilmiştir. Yapılan bu kontrollerin simülasyon sonuçlarına olan güveni arttıracığının belirtilmesi nedeniyle (Davis vd., 2007) bu çalışmada görünüş geçerliği kapsamında simülasyonun beklenildiği gibi gerçekleştiğinden emin olmak için Feinberg ve Rubright (2016) ile Sandilands (2014) tarafından önerilen adımlar dikkate alınmıştır. Böylelikle araştırma kapsamında görünüş geçerliği için yapılan tüm işlemler sırasıyla sunulmuştur.

1. Odak grup sayısının iki olduğu örnek bir durum için yapılan kontroller kapsamında, öncelikle yazılan bütün kodlar incelenerek bir tekrara dayalı örnek bir simülasyon gerçekleştirilmiştir. Yapılan deneme uygulaması kapsamında veri setinin üretilme süresinin uygun olduğu ve işlem sürecinde simülasyonun hata uyarısı vermediği görülmüştür. Ayrıca görsel incelemelere bağlı olarak, veri setinde eksik veya beklenilmeyen bir değere rastlanılmadığı sonucuna ulaşılmıştır.
2. Odak gruplara eklenen DMF miktarlarının (0, 0.40, 0.70) farklı ve grupların örneklem büyüklüklerinin eşit ($N_R = 600$, $N_{O1} = 600$, $N_{O2} = 600$) olduğu koşullarda referans ve odak grupların güçlük ve ayırt edicilik parametrelerinin incelenmesi amaçlanmış ve tek bir veri seti üzerinde 20 maddenin yalnızca referans gruptaki madde parametrelerini ve DMF'li olan dört maddenin (M5, M6, M9, M14) referans ve odak gruplardaki madde parametrelerini kontrol etmek için MKE'ler çizdirilmiştir (EK 8). Böylece yalnızca referans grubun MKE'leri ile teste DMF'li madde olduğu durumda referans ve odak grupların MKE'leri incelenerek, madde parametrelerinin doğru olduğu sonucuna ulaşılmıştır. Ayrıca DMF'li olarak belirlenen dört maddenin referans gruptaki MKE'leri ile madde parametrelerine ayrı ayrı 0.70 DMF miktarı eklendikten sonra odak gruplardaki MKE'leri çizdirilerek görsel kontroller yapılmıştır (EK 9). Yapılan incelemeler doğrultusunda DMF'li olarak belirlenen dört maddenin odak gruplar için beklenildiği gibi daha zor ve daha ayırt edici olduğu sonucuna ulaşılmıştır.
3. Bu aşamada tekrar sayısının 100 olduğu nihai durum için üretilen veri setlerinin beklenen satır ve sütun sayısına sahip olup olmadığı incelenmiştir. Grup sayısının üç olduğu durum için 2 (referans grubun örneklem büyüklüğü) $\times 9$ (odak grupların örneklem büyüklükleri oranı) $\times 25$ (DMF miktarı) $\times 5$ (DMF tekniği) $\times 100$ (tekrar sayısı) = 225,000 satırdan oluşan veri setinin üretildiği tespit edilmiştir. Grup sayısının dört olduğu durum için ise 2 (referans grubun örneklem büyüklüğü) $\times 27$ (odak grupların örneklem büyüklükleri oranı) $\times 81$ (DMF miktarı) $\times 5$ (DMF tekniği) $\times 100$ (tekrar sayısı) = 2,187,000 satırdan oluşan veri setinin üretildiği belirlenmiştir. Yapılan kontrollerde üretilen nihai veri setlerinde hatalı veya eksik verilerin bulunmadığı da anlaşıldığından, veri setlerinin analiz için hazır olduğu sonucuna ulaşılmıştır.
4. Simülasyonun geçerliğinin kanıtlanması sürecinde son aşamada araştırma bulguları ile alanyazındaki benzer çalışmalardan elde edilen sonuçların tutarlılığı

incelenmiştir. Bu kapsamda iki farklı işlem gerçekleştirilmiştir. İlk olarak araştırma kapsamında dikkate alınan GLR-LRT ve GLR-Wald tekniklerine ilişkin sonuçların öngörüldüğü gibi bütün simülasyon koşullarında birbirlerine paralel olduğu bulunmuştur. İkinci olarak ise testte hiçbir DMF’li madde olmayan durum için hesaplanan Tip I hata oranlarına ilişkin sonuçların beklendiği gibi alanyazındaki ilgili araştırmalarla son derece tutarlı olduğu tespit edilmiştir.

Araştırma kapsamında görünüş geçerliği çalışması için yapılan bütün incelemelere göre; veri üretme sürecinin sağlıklı bir şekilde yürütüldüğü, veri setinin hatasız, sonuçların ise beklenildiği gibi olduğu, önceki ilgili araştırmalarda ortaya konulan sonuçların bu çalışma ile tutarlı olduğu ve öngörüldüğü gibi GLR-LRT ile GLR-Wald tekniklerinin birbirlerine çok yakın performans gösterdikleri sonuçlarına ulaşılmıştır. Elde edilen bu sonuçlar, araştırma kapsamında yürütülen simülasyonlar için görünüş geçerliğinin sağlandığını ortaya koymuştur.

İstatistiksel Geçerlik

Simülasyon çalışmalarında yapılan hesaplamaların doğruluğunu tespit etmek amacıyla gerçekleştirilen ikinci aşamada istatistiksel geçerlik incelenmiştir (Sandilands, 2014; Scott, 2014). MTK’ya dayalı veri üretilen simülasyon çalışmalarında madde parametrelerine ilişkin yapılan kestirimlerin doğruluğunun son derece önemli olması nedeniyle (Bolt vd., 2002; Wang ve Chen, 2005; Wen, 2014) araştırma kapsamında güçlük ve ayırt edicilik parametrelerine ilişkin kestirim değerleri ile gerçek parametre değerleri arasındaki farkın hesaplanması gerektiği vurgulanmıştır (DeMars ve Lau, 2011; Sandilands, 2014). Böylece yapılan hesaplamalara bağlı olarak parametre kestirimlerinin doğru olup olmadığının ortaya konulabildiği belirtilmiştir (Finch, 2016).

Simülasyonun istatistiksel geçerliğine ilişkin kanıt sunmak amacıyla yapılan bu çalışmada elde edilen bulguların alanyazındaki ilgili araştırmalar ile tutarlı ve kabul edilebilir olup olmadığının tespit edilebilmesi amacıyla simülasyon deseninde açıklanan koşullar dikkate alınarak grup sayısının üç olduğu durum için yeniden veri üretilmiştir. Veri setleri üretilirken; yetenek (θ) parametresi için standart normal dağılım $[N(0, 1)]$, ayırt edicilik $[a \sim U(0.84, 1.53)]$ ile güçlük $[b \sim U(-1.06, 1.39)]$ parametreleri içinse belirtilen aralıklara göre tekdüze dağılım dikkate alınarak, referans grubun belirlenen dört maddedeki güçlük ve ayırt edicilik parametrelerine DMF miktarları (0.40, 0.70)

eklenmiştir. Grup sayısının üç olduğu durum için tamamen çaprazlanmış faktöriyel desene dayalı olarak gerçekleştirilen veri üretim sürecinde; iki (referans grubun örneklem büyüklüğü) x dokuz (odak grupların örneklem büyüklüklerinin, referans grubun örneklem büyüklüğüne oranı) x yirmi beş (gruplara eklenen DMF miktarı) x beş (DMF tekniği) x yüz (tekrar sayısı) olmak üzere toplam 225.000 koşul simüle edilmiştir.

Araştırma kapsamında veri üretildikten sonraki süreçte güçlük ve ayırt edicilik parametrelerine ilişkin ayrı ayrı mutlak yanlılık ve RMSE hesaplanmıştır. Öncelikle madde düzeyinde hesaplanan ortalama mutlak yanlılık ve ortalama RMSE ile hem güçlük hem de ayırt edicilik parametreleri için yapılan kestirimlerin kararlılığı belirlenmiştir. Sonrasında ise ilgili madde parametreleri için referans grubun örneklem büyüklüğü, odak grupların örneklem büyüklüğü oranları ve odak gruplara eklenen DMF miktarları faktörlerinden hangisinin ortalama mutlak yanlılık ve ortalama RMSE üzerinde etkili olduğunu belirlemek amacıyla Faktöriyel ANOVA yapılmıştır. Analizler sonucunda yirmi madde için hesaplanan mutlak yanlılık ve RMSE değerleri Tablo 2’de sunulmuştur.

Tablo 2

Madde Düzeyinde Hesaplanan Ortalama Mutlak Yanlılık ve Ortalama RMSE Değerleri

Güçlük Parametresi			Ayırt Edicilik Parametresi		
Madde	Ortalama Mutlak Yanlılık	Ortalama RMSE	Madde	Ortalama Mutlak Yanlılık	Ortalama RMSE
b_1	0.0870	0.1112	a_1	0.1212	0.1549
b_2	0.1092	0.1412	a_2	0.0965	0.1224
b_3	0.1185	0.1537	a_3	0.0935	0.1183
b_4	0.0894	0.1142	a_4	0.1109	0.1411
b_5	0.0855	0.1085	a_5	0.1118	0.1449
b_6	0.1495	0.2007	a_6	0.1148	0.1497
b_7	0.0903	0.1156	a_7	0.0864	0.1099
b_8	0.1114	0.1429	a_8	0.1012	0.1289
b_9	0.1188	0.1571	a_9	0.1430	0.1873
b_{10}	0.1326	0.1733	a_{10}	0.0981	0.1245
b_{11}	0.0754	0.0957	a_{11}	0.0951	0.1203
b_{12}	0.0832	0.1064	a_{12}	0.0918	0.1159
b_{13}	0.0727	0.0922	a_{13}	0.1005	0.1277
b_{14}	0.0661	0.0832	a_{14}	0.1478	0.1929
b_{15}	0.0762	0.0972	a_{15}	0.1162	0.1474
b_{16}	0.0555	0.0701	a_{16}	0.1201	0.1531
b_{17}	0.0589	0.0741	a_{17}	0.1112	0.1416
b_{18}	0.1030	0.1332	a_{18}	0.1029	0.1307
b_{19}	0.0837	0.1070	a_{19}	0.1140	0.1452
b_{20}	0.0887	0.1127	a_{20}	0.0974	0.1238

Tablo 2’de 20 maddenin güçlük ve ayırt edicilik parametrelerine ilişkin 2PLM’ye dayalı parametre kestirimleri için hesaplanan ortalama mutlak yanlılık ve ortalama RMSE değerleri verilmiştir. Ortalama mutlak yanlılık değerleri; güçlük parametresi için 0.0555 ile 0.1495 aralığında, ayırt edicilik parametresi için ise 0.0864 ile 0.1478 aralığında değişmiştir. Ortalama RMSE değerleri ise; güçlük parametresi için 0.0701 ile 0.2007 aralığında, ayırt edicilik parametresi için 0.1099 ile 0.1929 aralığında farklılaşmıştır. Elde edilen bulgular güçlük ve ayırt edicilik parametreleri için madde düzeyinde hesaplanan ortalama mutlak yanlılık ile ortalama RMSE değerlerinin birbirine paralel ve RMSE değerlerinin biraz daha yüksek olduğunu göstermiştir. Atar (2007) tarafından DMF ile ilgili yapılan bir simülasyon çalışmasında 20 madde için hesaplanan ortalama RMSE değerlerinin; güçlük parametresinde 0.1422 ile 0.4211 aralığında, ayırt edicilik parametresinde ise 0.1126 ile 0.2770 aralığında olduğu belirtilmiştir. Benzer şekilde Sandilands’ın (2014) çalışmasında 90 maddede güçlük parametresi için ortalama RMSE değerlerinin 0.064 ile 0.290 aralığında tespit edilmiştir. Başka bir araştırmada ise güçlük parametresi için hesaplanan ortalama RMSE değerinin 0.120 ile 0.210 aralığında olduğu belirlenmiştir (Lu ve Jiao, 2009; Akt. DeMars ve Lau, 2011). Bu bağlamda yapılan simülasyon çalışmasından elde edilen bulguların alanyazındaki araştırmalarla son derece tutarlı olduğu gözlemlenmiştir. Ayrıca her iki madde parametresi için madde düzeyindeki hem ortalama RMSE hem de ortalama mutlak yanlılık değerlerinin sıfıra yakın olması nedeniyle kestirim hatasının oldukça az olduğu ortaya konulmuştur. Böylelikle madde düzeyinde kestirilen parametreler ile gerçek parametre değerleri arasında ciddi düzeyde bir farklılaşmanın olmaması, parametre kestirimlerinin etkili olduğunu göstermiştir.

Yapılan bu simülasyon çalışmasında üç grup (referans-odak₁-odak₂) için referans grubun örneklem büyüklüğünün, odak grupların örneklem büyüklüğü oranlarının ve odak gruplardaki DMF miktarlarının bağımsız değişken, ortalama mutlak yanlılık ve ortalama RMSE değerlerinin ise bağımlı değişken olduğu durumlarda Faktöriyel ANOVA yapılmıştır. Bağımsız değişkenler arasında ikili etkileşimler hesaplanmış, ancak elde edilen sonuçlarda istatistiksel manidarlık gözlenmediğinden tablolarda yalnızca ana etkilere yer verilmiştir. Güçlük parametresi için ortalama mutlak yanlılık dikkate alınarak yapılan Faktöriyel ANOVA sonuçları Tablo 3’te sunulmuştur.

Tablo 3

Güçlük Parametresi İçin Faktöriyel ANOVA Sonucu (Ortalama Mutlak Yanlılık)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Referans örneklem büyüklüğü (R)	45.499	1	237.298	.000	.643	.841
Odak1 örneklem oranı (O ₁)	4.603	2	12.003	.000	.065	.348
Odak2 örneklem oranı (O ₂)	9.840	2	25.661	.000	.139	.533
Odak1 DMF miktarı (b ₁)	0.086	1	0.451	.505	.001	.010
Odak2 DMF miktarı (b ₂)	0.522	1	2.725	.106	.007	.057
Hata	8.628	45			.122	
Toplam	70.737					

Tablo 3'teki güçlük parametresi için ortalama mutlak yanlılık sonucuna göre referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,45)} = 237.298, p < .001, \eta^2 = .643$) manidar ve yüksek düzey bir etkiye sahip olduğu belirlenmiştir. Benzer şekilde birinci odak ($F_{(2,45)} = 12.003, p < .001, \eta^2 = .065$) ile ikinci odak ($F_{(2,45)} = 25.661, p < .001, \eta^2 = .139$) grubun örneklem büyüklüğü oranları için ana etkilerin istatistiksel olarak manidar, ancak orta düzey bir etkiye sahip oldukları tespit edilmiştir. Araştırmada odak gruplara eklenen DMF miktarları ana etkisi ile ikili etkileşimler için elde edilen sonuçlar ise manidar bulunmamıştır. Elde edilen bulgulara göre madde güçlük parametresi için ortalama mutlak yanlılık değerleri açısından referans grubun örneklem büyüklüğü ve odak grupların örneklem büyüklüğü oranlarının düzeyleri arasında sırasıyla yüksek ve orta düzey bir farklılaşma olduğu sonucuna ulaşılmıştır. Faktörlerin hangi düzeyleri arasında farklılaşma olduğunu belirlemek amacıyla çizilen grafikler EK 10'da hesaplanan ortalama mutlak yanlılık değerleri ise EK 11'de verilmiştir.

EK 11'deki tablodan görüldüğü üzere en yüksek mutlak yanlılık değeri 0.1422 iken, en düşük mutlak yanlılık değeri 0.0545'tir. EK 10'daki grafiklerde güçlük parametresine ilişkin ortalama mutlak yanlılık değerlerine göre; $N_R = 600$ olduğu durumdaki ortalama mutlak yanlılık değerlerinin, $N_R = 1800$ olduğu durumdaki ortalama mutlak yanlılık değerlerine göre istatistiksel olarak manidar bir şekilde yüksek olduğu belirlenmiştir ($\eta^2 = .643$). $N_R = 600$ olduğu durum için odak grupların örneklem büyüklükleri arttıkça hesaplanan ortalama mutlak yanlılık değerlerinin manidar bir şekilde azaldığı, $N_R = 1800$ olduğu durum için ise odak grupların örneklem büyüklükleri artmasına rağmen toplam örneklem yeterince büyük olması nedeniyle ortalama mutlak yanlılık değerlerinin görece sabit kaldığı görülmüştür. EK 10 ve EK 11 birlikte incelendiğinde ise odak gruplara eklenen DMF miktarlarının ortalama mutlak yanlılık üzerinde manidar etki yaratmadığı anlaşılmıştır. Böylece güçlük parametresi için toplam örneklem büyüklüğü arttıkça ortalama mutlak yanlılığın azaldığı sonucuna ulaşılmıştır.

Parametre kestirimlerinin doğruluğu hakkında kesin bilgiye ulaşmak amacıyla mutlak yanlılığın yanı sıra RMSE’de hesaplanarak sonuçlar birlikte değerlendirilmiştir. Araştırma kapsamında güçlük parametresi için ortalama RMSE’yi etkileyen faktörleri tespit etmek için yapılan Faktöriyel ANOVA sonuçları Tablo 4’te verilmiştir.

Tablo 4

Güçlük Parametresi İçin Faktöriyel ANOVA Sonucu (Ortalama RMSE)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Referans örneklem büyüklüğü (R)	63.602	1	1255.154	.000	.896	.965
Odak1 örneklem oranı (O ₁)	1.456	2	14.366	.000	.021	.390
Odak2 örneklem oranı (O ₂)	3.015	2	29.748	.000	.042	.569
Odak1 DMF miktarı (b ₁)	0.029	1	0.565	.456	.000	.012
Odak2 DMF miktarı (b ₂)	0.112	1	2.212	.144	.002	.047
Hata	2.280	45			.032	
Toplam	70.999					

Tablo 4’teki güçlük parametresi için ortalama RMSE sonucuna göre referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,45)} = 1255.154$, $p < .001$, $\eta^2 = .896$) manidar ve yüksek düzey bir etkiye sahip olduğu belirlenmiştir. Benzer şekilde birinci odak ($F_{(2,45)} = 14.366$, $p < .001$, $\eta^2 = .021$) ile ikinci odak ($F_{(2,45)} = 29.748$, $p < .001$, $\eta^2 = .042$) grubun örneklem büyüklüğü oranları için ana etkilerin istatistiksel olarak manidar, ancak düşük düzey bir etkiye sahip oldukları tespit edilmiştir. Araştırmada odak gruplara eklenen DMF miktarları ana etkisi ile ikili etkileşimler için elde edilen sonuçlar ise manidar bulunmamıştır. Elde edilen bulgulara göre madde güçlük parametresi için ortalama RMSE değerleri açısından referans grubun örneklem büyüklüğü ve odak grupların örneklem büyüklüğü oranları faktörlerinin düzeyleri arasında sırasıyla yüksek düzey ve düşük düzey bir farklılaşma olduğu sonucuna ulaşılmıştır. Faktörlerin hangi düzeyleri arasında farklılaşma olduğunu belirlemek amacıyla çizilen grafikler EK 12’de, her bir simülasyon koşulu için hesaplanan ortalama RMSE değerleri EK 11’de verilmiştir.

Ek 11’deki tablodan görüldüğü üzere en yüksek RMSE değeri 0.1905 iken, en düşük RMSE değeri ise 0.0686’tır. EK 12’deki grafiklerde güçlük parametresine ilişkin ortalama RMSE değerlerine göre; $N_R = 600$ olduğu durumdaki ortalama RMSE değerlerinin, $N_R = 1800$ olduğu durumdaki ortalama RMSE değerlerine göre istatistiksel olarak manidar ($\eta^2 = .896$) bir şekilde yüksek olduğu bulunmuştur. $N_R = 600$ olduğu durum için odak grupların örneklem büyüklükleri arttıkça hesaplanan ortalama RMSE değerlerinin manidar bir şekilde azaldığı, $N_R = 1800$ olduğu durum için ise odak grupların

örneklem büyüklükleri artmasına rağmen toplam örneklemin yeterince büyük olması nedeniyle ortalama RMSE değerlerinin görece sabit kaldığı görülmüştür. Diğer taraftan EK 11 ve EK 12 birlikte incelendiğinde odak gruplara eklenen DMF miktarlarının ortalama RMSE üzerinde manidar bir etki yaratmadığı anlaşılmıştır. Bu bağlamda genel olarak güçlük parametresi için toplam örneklem büyüklüğü arttıkça ortalama RMSE'nin azaldığı sonucuna ulaşılmıştır.

Araştırma kapsamında güçlük parametresinin yanı sıra ayırt edicilik parametresi için de ortalama mutlak yanlılık ve ortalama RMSE hesaplanmıştır. Ayırt Edicilik parametresi için ortalama mutlak yanlılık dikkate alınarak yapılan Faktöriyel ANOVA sonuçları Tablo 5'te sunulmuştur.

Tablo 5

Ayırt Edicilik Parametresi İçin Faktöriyel ANOVA Sonucu (Ortalama Mutlak Yanlılık)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Referans örneklem büyüklüğü (R)	45.499	1	658.464	.000	.643	.936
Odak1 örneklem oranı (O ₁)	10.547	2	76.315	.000	.149	.772
Odak2 örneklem oranı (O ₂)	10.536	2	76.236	.000	.149	.772
Odak1 DMF miktarı (a ₁)	0.133	1	1.926	.172	.002	.041
Odak2 DMF miktarı (a ₂)	0.067	1	0.966	.331	.001	.021
Hata	3.109	45			.044	
Toplam	70.738					

Tablo 5'teki ayırt edicilik parametresi için ortalama mutlak yanlılık sonucuna göre referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,45)} = 658.464$, $p < .001$, $\eta^2 = .643$) manidar ve yüksek düzey bir etkiye sahip olduğu belirlenmiştir. Benzer şekilde birinci odak ($F_{(2,45)} = 76.315$, $p < .001$, $\eta^2 = .149$) ile ikinci odak ($F_{(2,45)} = 76.236$, $p < .001$, $\eta^2 = .149$) grubun örneklem büyüklüğü oranları için ana etkilerin istatistiksel olarak manidar ve yüksek düzey bir etkiye sahip oldukları tespit edilmiştir. Araştırmada odak gruplara eklenen DMF miktarları ana etkisi ile ikili etkileşimler için elde edilen sonuçlar ise manidar bulunmamıştır. Elde edilen bulgulara göre madde ayırt edicilik parametresi için ortalama mutlak yanlılık değerleri açısından referans grubun örneklem büyüklüğü ve odak grupların örneklem büyüklüğü oranları faktörlerinin düzeyleri arasında yüksek düzey bir farklılaşma olduğu sonucuna ulaşılmıştır. Faktörlerin hangi düzeyleri arasında farklılaşma olduğunu belirlemek amacıyla çizilen grafikler EK 13'te, her bir simülasyon koşulu için hesaplanan ortalama RMSE değerleri EK 14'te verilmiştir.

EK 14'teki tablodan görüldüğü üzere en yüksek mutlak yanlılık değeri 0.1618 iken, en düşük mutlak yanlılık değeri ise 0.0678'dir. EK 13'teki grafiklerde ayırt edicilik parametresine ilişkin ortalama mutlak yanlılık değerlerine göre; $N_R = 600$ için ortalama mutlak yanlılık değerlerinin, $N_R = 1800$ için ortalama mutlak yanlılık değerlerine göre istatistiksel olarak manidar bir şekilde yüksek olduğu bulunmuştur ($\eta^2 = .643$). Ayrıca $N_R = 600$ olduğu durumda odak grupların örneklem büyüklükleri arttıkça hesaplanan ortalama mutlak yanlılık değerlerinin manidar bir şekilde azaldığı, $N_R = 1800$ olduğu durumda ise odak grupların örneklem büyüklükleri artmasına rağmen toplam örneklem yeterince büyük olması nedeniyle ortalama mutlak yanlılık değerlerinin görece sabit kaldığı görülmüştür. Diğer taraftan EK 13 ve EK 14 birlikte incelendiğinde odak gruplara eklenen DMF miktarlarının ortalama mutlak yanlılık üzerinde manidar bir etki yaratmadığı anlaşılmıştır. Bu bağlamda genel olarak ayırt edicilik parametresi için toplam örneklem büyüklüğü arttıkça ortalama mutlak yanlılığın azaldığı sonucuna ulaşılmıştır.

Araştırma kapsamında parametre kestirimlerinin doğruluğu hakkında kesin bilgiye ulaşmak amacıyla mutlak yanlılığın yanı sıra RMSE'de hesaplanarak sonuçlar birlikte değerlendirilmiştir. Ayırt edicilik parametresi için ortalama RMSE'yi etkileyen faktörleri belirlemek amacıyla yapılan Faktöriyel ANOVA sonuçları Tablo 6'da sunulmuştur.

Tablo 6

Ayırt Edicilik Parametresi İçin Faktöriyel ANOVA Sonucu (Ortalama RMSE)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Referans örneklem büyüklüğü (R)	45.499	1	637.780	.000	.643	.934
Odak1 örneklem oranı (O ₁)	10.835	2	75.943	.000	.153	.771
Odak2 örneklem oranı (O ₂)	10.434	2	73.126	.000	.147	.765
Odak1 DMF miktarı (a ₁)	0.100	1	1.397	.243	.001	.030
Odak2 DMF miktarı (a ₂)	0.056	1	0.779	.382	.001	.017
Hata	3.210	45			.045	
Toplam	70.737					

Tablo 6'daki ayırt edicilik parametresi için ortalama RMSE sonucuna göre referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,45)} = 637.780$, $p < .001$, $\eta^2 = .643$) manidar ve yüksek düzey bir etkiye sahip olduğu belirlenmiştir. Benzer şekilde birinci odak ($F_{(2,45)} = 75.943$, $p < .001$, $\eta^2 = .153$) ile ikinci odak ($F_{(2,45)} = 73.126$, $p < .001$, $\eta^2 = .147$) grubun örneklem büyüklüğü oranları için ana etkilerin istatistiksel olarak manidar ve yüksek düzey bir etkiye sahip oldukları tespit edilmiştir. Araştırmada odak gruplara

eklenen DMF miktarları ana etkisi ile ikili etkileşimler için elde edilen sonuçlar ise manidar bulunmamıştır. Elde edilen bulgulara göre madde ayırt edicilik parametresi için ortalama RMSE değerleri açısından referans grubun örneklem büyüklüğü ve odak grupların örneklem büyüklüğü oranları faktörlerinin düzeyleri arasında yüksek düzey bir farklılaşma olduğu sonucuna ulaşılmıştır. Dikkate alınan faktörlerin hangi düzeyleri arasında farklılaşma olduğunu belirlemek amacıyla çizilen grafikler EK 15'te sunulmuştur. Ayrıca her bir simülasyon koşulu için hesaplanan ortalama RMSE değerleri EK 14'te verilmiştir.

EK 14'teki tablodan görüldüğü üzere en yüksek RMSE değeri 0.2100 iken, en düşük RMSE değeri ise 0.0850'dir. EK 15'teki grafiklerde ayırt edicilik parametresine ilişkin ortalama RMSE değerlerine göre; $N_R = 600$ olduğu durumdaki ortalama RMSE değerlerinin, $N_R = 1800$ olduğu durumdaki ortalama RMSE değerlerine göre istatistiksel olarak manidar bir şekilde yüksek olduğu bulunmuştur ($\eta^2 = .643$). $N_R = 600$ olduğu durum için odak grupların örneklem büyüklükleri arttıkça hesaplanan ortalama RMSE değerlerinin manidar bir şekilde azaldığı, $N_R = 1800$ olduğu durum için ise odak grupların örneklem büyüklükleri artmasına rağmen toplam örneklemin yeterince büyük olması nedeniyle ortalama RMSE değerlerinin görece sabit kaldığı görülmüştür. Diğer taraftan EK 14 ve EK 15 birlikte incelendiğinde odak gruplara eklenen DMF miktarlarının ortalama RMSE üzerinde manidar bir etki yaratmadığı anlaşılmıştır. Bu bağlamda genel olarak ayırt edicilik parametresi için toplam örneklem büyüklüğü arttıkça ortalama RMSE'nin azaldığı sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışması kapsamında simülasyonun geçerli olup olmadığını ortaya koymak amacıyla parametre kestirimlerinin gerçeği yansıtmadığı kontrol edildikten sonra güçlük ve ayırt edicilik parametrelerine ilişkin ortalama mutlak yanlılık ve ortalama RMSE'yi etkileyen faktörler belirlenmiştir. Madde parametrelerine ilişkin 2PLM ile yapılan kestirimlerin hatasız olmasının son derece önemli olması nedeniyle (Tay vd., 2015), araştırmada güçlük ve ayırt edicilik parametreleri için ortalama mutlak yanlılık ve ortalama RMSE değerleri birlikte incelenmiştir. Böylece analizler sonucunda elde edilen bulguların hem mutlak yanlılık ile RMSE hem de madde parametreleri açısından son derece paralel olduğu sonucuna ulaşılmıştır (bkz. EK 10-15).

Güçlük ve ayırt edicilik parametreleri için ortalama mutlak yanlılık ve ortalama RMSE değerlerinin düşük ve sıfıra yakın olması, kestirilen ve gerçek parametre değerleri arasındaki farkın daha az, yani kestirimlerdeki hataların daha düşük ve dolayısıyla

sonuçların daha doğru olduğunu göstermiştir (Liu, Zhang ve Luo, 2015; Yuan vd., 2015). Araştırma kapsamında bütün koşullar için hesaplanan RMSE'lerin mutlak yanlılıktan nispeten daha yüksek olduğu belirlenmiştir. Ayrıca ayırt edicilik parametresi için hesaplanan mutlak yanlılık ve RMSE'lerin de güçlük parametresine göre daha yüksek olduğu sonucuna ulaşılmıştır (bkz. EK 10-15). Hem güçlük hem de ayırt edicilik parametresi için toplam örneklem büyüklüğü arttıkça hesaplanan ortalama mutlak yanlılık ve ortalama RMSE değerlerinin sıfıra yaklaşma eğiliminde olduğu görülmüştür.

Yapılan bu araştırmaya benzer şekilde alanyazındaki çok sayıda araştırmada da her iki madde parametresi için mutlak yanlılık ve RMSE'nin örneklem büyüklüğünden etkilendiği belirtilmiştir (Bulut ve Sünbül, 2017; DeMars, 2003; Hulin, Lissak ve Drasgow, 1982; Sandilands, 2014; Sigal ve Chalmers, 2016; Socha vd., 2015; Tay vd., 2015; Wang ve Chen, 2005; Woods, 2009b). Özellikle MTK çalışmalarında madde parametrelerine ilişkin kestirimlerin kararlı olabilmesi için genel olarak 500'ün üzerinde bir örneklem büyüklüğü önerilmiştir (Hulin vd., 1982). Hulin ve diğerleri (1982) tarafından gerçekleştirilen çalışmada 2PLM'ye dayalı olarak yapılan kestirimler sonucunda örneklem büyüklüğü ile testteki madde sayısı arttığında daha düşük RMSE değerleri elde edildiği ortaya konulmuştur. Wang ve Chen (2005) testteki madde sayısının 20 veya daha fazla olması durumunda, örneklem büyüklüğü arttıkça yanlılığın azaldığını açıklamışlardır. Benzer şekilde Sandilands (2014) ile Bulut ve Sünbül (2017) yaptıkları çalışmalarında test uzunluğu ve örneklem büyüklüğü arttığında yanlılık ve RMSE değerlerinin azaldığını belirtmişlerdir. Socha ve diğerleri (2015) tarafından yapılan araştırmada güçlük parametresi için gruptaki örneklem büyüklüğü arttığında yanlılık ve RMSE değerlerinin azaldığı ifade edilmiştir. Atar (2007) ise çalışmasında RMSE ile yanlılığın karesini hesaplanmış ve örneklem büyüklüğü, odak grupların örneklem büyüklüğü oranı (1:1; 1:2) ve DMF miktarı faktörlerinin ana etki ile ikili etkileşimlerini incelemiştir. Araştırmacı güçlük ve ayırt edicilik parametreleri için örneklem büyüklüğü arttıkça hesaplanan RMSE ($p < .001$) ve yanlılığın karesi ($p < .001$) değerlerinin azaldığını ortaya koymuştur. Diğer taraftan DMF miktarı ve toplam örneklem büyüklüğü sabit kalmak şartıyla grupların örneklem büyüklüğü dengesiz olduğunda ayırt edicilik parametresi için RMSE ($p < .001$) ve yanlılığın karesi ($p < .01$) değerlerinin arttığı, güçlük parametresi için ise RMSE ($p = .015$) değerleri artmasına rağmen yanlılığın karesi ($p = .042$) değerlerinin ciddi miktarda değişmediği gözlemlenmiştir. Öte yandan DMF miktarındaki farklılaşmanın manidar bir etki yaratmadığı belirtilmiştir.

Alanyazındaki ilgili alıřmalarda olduėu gibi bu arařtırma kapsamında da grupların rneklem byklklerdeki artıřın RMSE ve yanlılık deėerlerinin sıfıra yaklařmasına yol aması nedeniyle kestirimlerin daha doėru olduėu ortaya konulmuřtur. Ayrıca hem glk hem de ayırt edicilik parametreleri iin manidar olan faktrlere iliřkin hesaplanan mutlak yanlılık ve RMSE deėerlerinin ilgili alanyazındaki (Atar, 2007; Hulin vd., 1982; Sandilands, 2014; Socha vd., 2015) alıřmalarla son derece tutarlı olduėu sonucuna ulařılmıřtır. Bylece simlasyonun geerliėini kanıtlamak amacıyla yapılan btn incelemelere baėlı olarak, bu arařtırma kapsamında retilen verilerin son derece gvenilir ve arařtırmanın amacına uygun olduėu kanaatine varılmıřtır.



BÖLÜM 4

BULGULAR VE YORUMLAR

Araştırma kapsamında ikili puanlanan maddelerde çoklu grup DMF belirleme tekniklerinden GLK, GLR-LRT, GLR-Wald, GMH ve Lordif'in farklı koşullar altındaki performanslarının incelenmesi amaçlanmıştır. Çalışmanın amacı doğrultusunda grup sayısı, referans grubun örneklem büyüklüğü, odak grubun örneklem büyüklüğü oranı, DMF miktarı ve DMF türü faktörlerine ilişkin koşullar değiştirilerek, çoklu grup tekniklerinin performansları Tip I hata, istatistiksel güç ve Tip III hata oranları açısından karşılaştırılmıştır. Hesaplanan temel Tip I hatanın yüksek olduğu durumlarda istatistiksel gücün yorumlanmasının hatalı olabilmesi nedeniyle (Cheng vd., 2013; Elosua ve Wells, 2013; Fidalgo ve Scalón, 2010; Finch, 2005; Kim, 2003; Lopez Rivas, 2012; Rockoff, 2018; Socha vd., 2015) araştırmadaki istatistiksel güç oranlarına ilişkin sonuçlar temel Tip I hata oranlarına bağlı olarak değerlendirilmiştir.

- 4.1. Grup sayısının üç ve dört olduğu durumda; referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3), DMF miktarı [$a(0.0, 0.4, 0.7)$, $b(0.0, 0.4, 0.7)$] ve DMF türü [DMF Yok, TB, TBO(a), TBO(ab)] koşulları için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin;
- a) Tip I hata oranları,
 - b) İstatistiksel güç oranları,
 - c) Tip III hata oranları nasıl değişmektedir?

Çalışmanın birinci araştırma sorusu kapsamında hem üç hem de dört grup için dikkate alınan koşullarda GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin performanslarını belirlemek amacıyla TB, TBO(a) ve TBO(ab) olmak üzere üç DMF türü için ayrı ayrı Faktöriyel ANOVA yapılmıştır. Faktöriyel ANOVA'larda çoklu grup DMF belirleme teknikleri, referans grubun örneklem büyüklüğü, odak grupların örneklem büyüklüğü oranları ve odak grupların madde parametrelerine eklenen DMF miktarları olmak üzere beş faktör için ana etkiler ile bu faktörlerin bütün ikili etkileşimleri hesaplanmasına rağmen, tablolarda yalnızca istatistiksel teknik faktörünün diğer faktörlerle olan ikili etkileşimlerine ilişkin sonuçlar raporlanmış ve elde edilen sonuçlar

arasından istatistiksel olarak manidar olan ana etki ile etkileşim etkileri yorumlanmıştır. Birinci araştırma sorusu kapsamında yanıt aranan bütün alt araştırma soruları için üç ve dört grupta gerçekleştirilen Faktöriyel ANOVA tabloları bu bölümde sunulmuşken, betimsel istatistiklere ilişkin tablo ve grafiklere EK bölümünde yer verilmiştir.

4.1.1 Üç grupta TB DMF türü için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı $[b(0.0, 0.4, 0.7)]$ koşullarındaki Tip I hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TB DMF türü için çoklu grup DMF belirleme tekniklerinin performansları Tip I hata açısından incelenmiştir. Üç grup ve TB DMF türü için Tip I hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 7’de sunulmuştur.

Tablo 7

Üç Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	226.530	4	1423.767	.000	.280	.887
Referans örneklem büyüklüğü (R)	154.442	1	3882.738	.000	.191	.842
Odak1 örneklem oranı (O ₁)	3.533	2	44.413	.000	.004	.109
Odak2 örneklem oranı (O ₂)	2.984	2	37.512	.000	.004	.093
Odak1 DMF miktarı (b ₁)	111.541	2	1402.095	.000	.138	.794
Odak2 DMF miktarı (b ₂)	111.660	2	1403.596	.000	.138	.794
T x R	39.693	4	249.475	.000	.049	.578
T x O ₁	0.719	8	2.261	.022	.001	.024
T x O ₂	0.793	8	2.492	.011	.001	.027
T x b ₁	26.235	8	82.446	.000	.032	.475
T x b ₂	26.000	8	81.707	.000	.032	.473
Hata	28.957	728			.036	
Toplam	808.999					

Tablo 7’deki üç grup ve TB DMF türü için verilen Tip I hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(4,728)} = 1423.767$, $p < .001$, $\eta^2 = .280$) manidar ve yüksek düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,728)} = 3882.738$, $p < .001$, $\eta^2 = .191$) manidar ve yüksek düzey, birinci odak ($F_{(2,728)} = 1402.095$, $p < .001$, $\eta^2 = .138$) ile ikinci odak ($F_{(2,728)} = 1403.596$, $p < .001$, $\eta^2 = .138$) gruba eklenen TB DMF miktarı ana etkilerinin manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan birinci odak ($F_{(2,728)} = 44.413$, $p < .001$, $\eta^2 = .004$) ile ikinci odak ($F_{(2,728)} = 37.512$, $p < .001$, $\eta^2 = .004$) grubun örneklem büyüklüğü oranlarına ilişkin ana etkilerin istatistiksel olarak manidar, ancak çok düşük etkiye sahip oldukları görülmektedir.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(4,728)} = 249.475, p < .001, \eta_p^2 = .578$) manidar ve yüksek düzey, istatistiksel teknik x birinci odak ($F_{(8,728)} = 82.446, p < .001, \eta_p^2 = .475$) ve istatistiksel teknik x ikinci odak ($F_{(8,728)} = 81.707, p < .001, \eta_p^2 = .473$) gruba eklenen TB DMF miktarı ikili etkileşimlerinin manidar ve yüksek düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan istatistiksel teknik x birinci odak ($F_{(8,728)} = 2.261, p < .05, \eta_p^2 = .024$) ve istatistiksel teknik x ikinci odak ($F_{(8,728)} = 2.492, p < .05, \eta_p^2 = .027$) grubun örneklem büyüklüğü oranı ikili etkileşimlerinin manidar ancak düşük düzey bir etkiye sahip oldukları gözlemlenmektedir. Böylece üç grup için TB DMF türü dikkate alındığında, Tip I hata üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .578$). Öte yandan odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de istatistiksel teknik ile etkileşimleri için hesaplanan etki büyüklüklerinin düşük olması nedeniyle, bu faktörlerin Tip I hata oranları üzerinde etkili olmadıkları ifade edilebilir.

4.1.2. Dört grupta TB DMF türü için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1, 1:1/2 ve 1:2/3) ve DMF miktarı [$b(0.0, 0.4, 0.7)$] koşullarındaki Tip I hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TB DMF türü için çoklu grup DMF belirleme tekniklerinin performansları Tip I hata açısından incelenmiştir. Dört grup ve TB DMF türü için Tip I hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 8’de sunulmuştur.

Tablo 8

Dört Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	2338.902	4	9919.025	.000	.321	.847
Referans örneklem büyüklüğü (R)	1624.693	1	27560.579	.000	.223	.794
Odak1 örneklem oranı (O ₁)	32.548	2	276.062	.000	.004	.072
Odak2 örneklem oranı (O ₂)	33.605	2	285.026	.000	.005	.074
Odak3 örneklem oranı (O ₃)	35.095	2	297.664	.000	.005	.077
Odak1 DMF miktarı (b ₁)	489.466	2	4151.545	.000	.067	.537
Odak2 DMF miktarı (b ₂)	488.202	2	4140.818	.000	.067	.537
Odak3 DMF miktarı (b ₃)	491.946	2	4172.580	.000	.068	.539

(Devam ediyor)

Tablo 8 (Devam)

Dört Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
T x R	260.941	4	1106.623	.000	.036	.382
T x O ₁	6.656	8	14.113	.000	.001	.016
T x O ₂	5.677	8	12.037	.000	.001	.013
T x O ₃	4.914	8	10.421	.000	.001	.012
T x b ₁	72.413	8	153.547	.000	.010	.147
T x b ₂	71.385	8	151.368	.000	.010	.145
T x b ₃	70.830	8	150.191	.000	.010	.144
Hata	421.374	7148			.058	
Toplam	7283.183					

Tablo 8'deki dört grup ve TB DMF türü için verilen Tip I hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(4,7148)} = 9919.025$, $p < .001$, $\eta^2 = .321$) manidar ve yüksek düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,7148)} = 27560.579$, $p < .001$, $\eta^2 = .223$) manidar ve yüksek düzey, birinci odak ($F_{(2,7148)} = 4151.545$, $p < .001$, $\eta^2 = .067$), ikinci odak ($F_{(2,7148)} = 4140.818$, $p < .001$, $\eta^2 = .067$) ve üçüncü odak ($F_{(2,7148)} = 4172.580$, $p < .001$, $\eta^2 = .068$) gruba eklenen TB DMF miktarı ana etkilerinin ise manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan birinci odak ($F_{(2,7148)} = 276.062$, $p < .001$, $\eta^2 = .004$), ikinci odak ($F_{(2,7148)} = 285.026$, $p < .001$, $\eta^2 = .005$) ve üçüncü odak ($F_{(2,7148)} = 297.664$, $p < .001$, $\eta^2 = .005$) grubun örneklem büyüklüğü oranlarına ilişkin ana etkilerin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(4,7148)} = 1106.623$, $p < .001$, $\eta_p^2 = .382$) manidar ve yüksek düzey, istatistiksel teknik x birinci odak ($F_{(8,7148)} = 153.547$, $p < .001$, $\eta_p^2 = .147$), istatistiksel teknik x ikinci odak ($F_{(8,7148)} = 151.368$, $p < .001$, $\eta_p^2 = .145$) ve istatistiksel teknik x üçüncü odak ($F_{(8,7148)} = 150.191$, $p < .001$, $\eta_p^2 = .144$) gruba eklenen TB DMF miktarı ikili etkileşimlerinin manidar ve yüksek düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan istatistiksel teknik x birinci odak ($F_{(8,7148)} = 14.113$, $p < .05$, $\eta_p^2 = .016$), istatistiksel teknik x ikinci odak ($F_{(8,7148)} = 12.037$, $p < .05$, $\eta_p^2 = .013$) ve istatistiksel teknik x üçüncü odak ($F_{(8,7148)} = 10.421$, $p < .05$, $\eta_p^2 = .012$) grubun örneklem büyüklüğü oranı ikili etkileşimlerinin manidar, ancak düşük düzey bir etkiye sahip oldukları gözlemlenmektedir. Böylece dört grup için TB DMF türü dikkate alındığında, Tip I hata üzerinde en önemli etkileşim etkisinin, istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 =$

.382). Bununla birlikte odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de istatistiksel teknik ile etkileşimleri için hesaplanan etki büyüklüklerinin düşük olması nedeniyle, bu faktörlerin Tip I hata oranları üzerinde etkili olmadıkları ifade edilebilir.

Araştırma kapsamında hem üç hem de dört grup için EK 16’da verilen bulgulara göre; çoklu grup teknikleri için hesaplanan temel Tip I hata oranlarının .018 ile .056 aralığında değiştiği ortaya konulmuştur. Hesaplanan temel Tip I hata oranlarının genel olarak nominal Tip I hata oranı olan .05’ten düşük veya bu değere makul ölçüde yakın olduğu (Tay vd., 2016) ve Bradley (1978) tarafından önerilen liberal kritere (.075) göre iyi düzeyde kontrol edildiği sonucuna ulaşılmıştır. Böylece 100 tekrarın yapıldığı bu araştırmada; grup sayısı önemli olmaksızın testteki hiçbir madde gerçekte DMF’li olmadığına, bu maddelerin GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri tarafından %1.8 ile %5.6 arasındaki oranlarda TB DMF’li olarak belirlenebildiği gözlemlenmiştir. Ayrıca yapılan bu araştırmada grup sayısı sabit tutulduğunda referans grubun örneklem büyüklüğü artmasına rağmen temel Tip I hata oranlarında ciddi bir farklılaşmanın söz konusu olmadığı tespit edilmiştir. Bu bilgiler doğrultusunda grup sayısının üç veya dört, DMF türünün ise TB olduğu durumda, çoklu grup tekniklerine ilişkin hesaplanan temel Tip I hata oranlarının .075’in altında yer alması nedeniyle (Cheng vd., 2013; Elosua ve Wells, 2013; Fidalgo ve Scalón, 2010; Finch, 2005; Kim, 2003; Lopez Rivas, 2012; Rockoff, 2018; Socha vd., 2015) TB DMF türünün incelendiği bütün koşullarda istatistiksel güç oranlarının yorumlanabileceği sonucuna ulaşılmıştır.

Alanyazında TB DMF türü dikkate alınarak çoklu grup tekniklerine ilişkin performansların incelendiği bazı araştırmalarda; GMH için üç gruptaki temel Tip I hata oranlarının .050 (Penfield, 2001) ve .048 (Fidalgo ve Scalón, 2010), GLR ve GMH için dört gruptaki temel Tip I hata oranlarının ise .060 (Penfield, 2001), .049 (Fidalgo ve Scalón, 2010), .058 (Jin vd., 2015) ve .051 (Rollins III, 2018) olarak hesaplandığı görülmüştür. Battauz (2019) tarafından yapılan başka bir çalışmada ise GLK tekniği için temel Tip I hata oranı .017 olarak hesaplanmış ve alanyazında üç grup için yapılan araştırmalardan elde edilen sonuçların bu çalışma ile son derece tutarlı olduğu belirlenmiştir. Benzer şekilde beş grubun dikkate alındığı araştırmalardaki temel Tip I hata oranlarının; GMH için .054 (Tutz ve Berger, 2016) ve .052 (Schauberger ve Tutz, 2016); GLR için .058 (Tutz ve Berger, 2016) ve .051 (Schauberger ve Tutz, 2016); GLK içinse .017 (Schauberger ve Tutz, 2016) ve .025 (Tutz ve Berger, 2016) olarak hesaplandığı tespit edilmiştir. Finch (2016) tarafından yapılan başka bir araştırmada ise

örneklem büyüklükleri eşit olduğunda testte hiçbir maddenin TB DMF içermediği koşul için temel Tip I hata oranları hesaplanmış ve hem üç grup [GLK (.015), GLR (.050), GMH (.065)] hem de dört grup [GLK (.010), GLR (.055), GMH (.070)] için GLK'nın belirgin bir şekilde daha düşük temel Tip I hata oranına sahip olduğu ifade edilmiştir. Yapılan bu simülasyon çalışmasında da grup sayısı önemli olmaksızın TB DMF türü için hesaplanan temel Tip I hata oranlarına göre hem GLR-LRT, GLR-Wald ve GMH teknikleri için elde edilen değerlerin birbirlerine son derece yakın hem de en düşük temel Tip I hata oranına sahip olan tekniğin GLK olması nedeniyle, bu sonuçların alanyazındaki birçok araştırma (Fidalgo ve Scalón, 2010; Finch, 2016; Penfield, 2001; Rollins III, 2018; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016) ile desteklendiği tespit edilmiştir. Ayrıca yapılan bu araştırma kapsamında temel Tip I hata oranlarının referans grubun örneklem büyüklüğüne bağlı olarak değişmediği belirlenmiş ve elde edilen bu bulgunun benzer özellikteki araştırmalar (Battauz, 2019; Finch, 2016; Penfield, 2001) ile tutarlı olduğu gözlemlenmiştir. İkili grupların ele alındığı başka simülasyon çalışmalarında da testin DMF'li madde içermemesi durumunda grupların örneklem büyüklüğü artmasına rağmen hesaplanan temel Tip I hata oranlarının değişmediği vurgulandığından (Finch ve French, 2007; Kopf vd., 2015b; Li ve Zumbo, 2009; Stark vd., 2006) grup sayısının üç veya dört, DMF türünün ise TB olduğu bütün koşullarda hesaplanan temel Tip I hata oranlarının alanyazındaki araştırmaların sonuçları ile desteklendiği sonucuna ulaşılmıştır.

Araştırma kapsamında temel bir Tip I hatanın yanı sıra belirlenen dört maddenin güçlük parametrelerinin değişimlendiği koşullarda da Tip I hata hesaplanmıştır. Bu bağlamda referans grubun örneklem büyüklüğü sabit kalmak şartıyla odak gruplara eklenen TB DMF miktarı arttığında veya yine odak gruplara eklenen TB DMF miktarı sabit kalmak şartıyla referans grubun örneklem büyüklüğü arttığında GLK, GLR-LRT, GLR-Wald ve GMH için hesaplanan Tip I hata oranlarının da yükselerek, üç grup için .040 ile .441 arasında, dört grup içinse .033 ile .438 arasında değişiklik gösterdiği belirlenmiştir (bkz. EK 16). Öte yandan bütün koşullarda GLK'nın GLR-LRT, GLR-Wald ve GMH tekniklerinden daha düşük Tip I hata oranlarına sahip olduğu, Lordif için elde edilen değerlerin ise üç grup için .061 ile .066 arasında, dört grup için .064 ile .070 arasında yer aldığı tespit edilmiştir. Böylelikle MTK'ya dayalı olan GLK ve Lordif'in, KTK'ya dayalı olan GLR-LRT, GLR-Wald ve GMH tekniklerinden daha iyi performans gösterdikleri ortaya konulmuştur. Ayrıca Lordif için saflaştırma işleminin otomatik olarak gerçekleştirilmesi neticesinde, bu tekniğin DMF miktarı ve referans grubun örneklem büyüklüğündeki artıştan etkilenmeyerek, dikkate alınan bütün koşullarda en

düşük Tip I hata oranına sahip olan teknik olduğu görülmüştür. Simülasyon çalışmalarında hesaplanan yüksek Tip I hata oranları, tekniklerin performanslarının düşük olduğunu gösterdiğinden (Kim, 2003; Schauburger ve Tutz, 2016) odak gruplara eklenen TB DMF miktarı ve grupların örneklem büyüklükleri arttıkça Tip I hata oranlarının yükselmesine bağlı olarak GLK, GLR-LRT, GLR-Wald ve GMH tekniklerinin performanslarının azaldığı sonucuna ulaşılmıştır.

Yapılan bu araştırma kapsamında analiz sürecinde ek bir saflaştırma işleminin gerçekleştirilmemesi nedeniyle hem üç hem de dört grup için referans grubun örneklem büyüklüğü ve TB DMF miktarı arttıkça GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için hesaplanan Tip I hata oranlarının yükseldiği tespit edilmiştir. TB DMF türünün incelenerek hem ikili hem de çoklu grupların dikkate alındığı alanyazındaki benzer çalışmalarda, DMF miktarı ve örneklem büyüklüğü ayrı ayrı veya birlikte arttıkça saflaştırma işlemi yapılmadığı takdirde hesaplanan Tip I hata oranlarının yükseldiği belirtilmiş (Finch ve French, 2008; Hidalgo vd., 2014; Jin vd., 2015; Li ve Zumbo, 2009; Scott, 2014; Stark vd., 2006; Tay vd., 2013; Tutz ve Schauburger, 2015) ve elde edilen bulguların alanyazındaki sonuçlar ile desteklendiği gözlemlenmiştir.

Araştırma kapsamında grup sayısının üç veya dört, DMF türünün ise TB olduğu bütün koşullar altında GLR-LRT, GLR-Wald ve GMH tekniklerine ilişkin Tip I hata oranlarının birbirlerine son derece yakın olduğu tespit edilmiştir. Alanyazındaki hem ikili grup (MH ve LR) (Kim, 2010) hem de çoklu grup (GMH ve GLR) (Finch, 2016; Rollins III, 2018; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) tekniklerinin dikkate alındığı birçok araştırmada da oldukça benzer sonuçlara ulaşılması nedeniyle elde edilen bulguların ilgili araştırmalarla tutarlı olduğu anlaşılmıştır. Diğer taraftan bazı çalışmalarda da (Gao, 2019; Magis vd., 2011) açıklandığı gibi, TB DMF türünün ele alındığı bütün koşullarda GLR-LRT ve GLR-Wald tekniklerinin Tip I hata açısından paralel performans gösterdikleri belirlenmiştir.

İkili puanlanan maddeler için grup sayısının üç veya dört, DMF türünün ise TB olduğu durumda Lordif tekniğinin performansının ilk kez incelendiği bu simülasyon çalışmasında; referans grubun örneklem büyüklüğü ve odak gruplara eklenen TB DMF miktarı arttıkça GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için Tip I hata oranlarının yükseldiği tespit edilmesine rağmen, Lordif için hesaplanan Tip I hata oranlarının ciddi miktarda değişmediği ve liberal kritere (.075) göre iyi düzeyde kontrol edildiği görülmüştür. Bu bulguyu destekler nitelikte olan Tunç (2016) ile Schauburger ve Mair'in (2020) çalışmalarında da iki grup ve ikili puanlanan maddelerde TB DMF miktarı

ve örneklem büyüklüğü belirli bir büyüklüğe kadar artsa dahi, Lordif için hesaplanan Tip I hata oranlarının ciddi miktarda farklılaşmadığı ortaya konulmuştur. Bu anlamda grup sayısı önemli olmaksızın testteki maddelerin TB DMF’li olduğu bütün koşullar altında Lordif’in Tip I hata açısından kabul edilebilir düzeyde performans göstermesi nedeniyle elde edilen bulguların sınırlı sayıda araştırma ile desteklendiği belirlenmiştir.

Araştırma kapsamında performansı incelenen Lordif tekniğinin analiz sürecinde yinelemeli bir saflaştırma işleminin otomatik olarak geliştirildiğinden, testteki TB DMF içeren maddelerin eşleştirme kriteri üzerindeki etkisi arındırılmakta ve bu durumun Lordif için önemli bir avantaj olduğu belirtilmektedir (Choi vd., 2011, 2016; Crane vd., 2006). Yapılan bu simülasyon çalışmasında dört maddenin TB DMF içermesi durumunda, DMF miktarı ve örneklem büyüklüğü artsa dahi Lordif için hesaplanan Tip I hata oranlarının bu artıştan etkilenmeyerek diğer tekniklere kıyasla daha düşük ve liberal kritere göre iyi düzeyde kontrol edilmesi, analiz esnasında otomatik olarak gerçekleştirilen saflaştırma işleminin Lordif’in performansına olumlu yönde katkı sağladığını kanıtlamıştır. Araştırmanın bu bulgusuna dayalı olarak hem üç hem de dört grupta TB DMF’li madde içeren testler için ek bir saflaştırma işlemine gerek duyulmaksızın ele alınan bütün koşullarda Tip I hata oranlarının düşük olması nedeniyle, Lordif’in tercih edilebileceği sonucuna ulaşılmıştır.

4.1.3 Üç grupta TB DMF türü için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı $[b(0.4, 0.7)]$ koşullarındaki istatistiksel güç oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TB DMF türü için çoklu grup DMF belirleme tekniklerinin performansları istatistiksel güç açısından incelenmiştir. Üç grup ve TB DMF türü için istatistiksel güç oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 9’da sunulmuştur.

Tablo 9

Üç Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	20.594	4	90.813	.000	.057	.547
Referans örneklem büyüklüğü (R)	115.910	1	2044.531	.000	.323	.872
Odak1 örneklem oranı (O ₁)	3.190	2	28.132	.000	.009	.157

(Devam ediyor)

Tablo 9 (Devam)

<i>Üç Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları)</i>						
Odak2 örneklem oranı (O ₂)	2.161	2	19.062	.000	.006	.112
Odak1 DMF miktarı (b ₁)	46.904	1	827.346	.000	.131	.733
Odak2 DMF miktarı (b ₂)	46.810	1	825.677	.000	.130	.733
T x R	12.629	4	55.691	.000	.035	.425
T x O ₁	0.128	8	0.283	.971	.000	.007
T x O ₂	0.166	8	0.365	.938	.000	.010
T x b ₁	2.336	4	10.301	.000	.007	.120
T x b ₂	1.814	4	8.000	.000	.005	.096
Hata	17.064	301			.048	
Toplam	359.002					

Tablo 9'daki üç grup ve TB DMF türü için verilen istatistiksel güç sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(4,301)} = 90.813$, $p < .001$, $\eta^2 = .057$) manidar ve düşük düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,301)} = 2044.531$, $p < .001$, $\eta^2 = .323$) manidar ve yüksek düzey, birinci odak ($F_{(1,301)} = 827.346$, $p < .001$, $\eta^2 = .131$) ile ikinci odak ($F_{(1,301)} = 825.677$, $p < .001$, $\eta^2 = .130$) gruba eklenen TB DMF miktarı ana etkilerinin manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan birinci odak ($F_{(2,301)} = 28.132$, $p < .001$, $\eta^2 = .009$) ile ikinci odak ($F_{(2,301)} = 19.062$, $p < .001$, $\eta^2 = .006$) grubun örneklem büyüklüğü oranlarına ilişkin ana etkilerin manidar, ancak çok düşük etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(4,301)} = 55.691$, $p < .001$, $\eta_p^2 = .425$) manidar ve yüksek düzey, istatistiksel teknik x birinci odak ($F_{(4,301)} = 10.301$, $p < .001$, $\eta_p^2 = .120$) ve istatistiksel teknik x ikinci odak ($F_{(4,301)} = 8.000$, $p < .001$, $\eta_p^2 = .096$) gruba eklenen TB DMF miktarı ikili etkileşimlerinin manidar ve orta düzey bir etkiye sahip oldukları görülmektedir. Böylece üç grup için TB DMF türü dikkate alındığında, istatistiksel güç üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu ifade edilebilir ($\eta_p^2 = .425$). Diğer yandan odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de istatistiksel teknik ile etkileşimlerinin ya manidar olmaması ya da hesaplanan etki büyüklüklerinin son derece düşük olması nedeniyle, TB DMF'li maddelerin istatistiksel güç oranları üzerinde bu faktörlerin etkili olmadıkları belirtilebilir.

4.1.4 Dört grupta TB DMF türü için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı $[b(0.4, 0.7)]$ koşullarındaki istatistiksel güç oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TB DMF türü için çoklu grup DMF belirleme tekniklerinin performansları istatistiksel güç açısından incelenmiştir. Dört grup ve TB DMF türü için istatistiksel güç oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 10'da sunulmuştur.

Tablo 10

Dört Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	1.314	4	375.879	.000	.075	.422
Referans örneklem büyüklüğü (R)	4.562	1	5217.641	.000	.262	.717
Odak1 örneklem oranı (O ₁)	0.042	2	23.864	.000	.002	.023
Odak2 örneklem oranı (O ₂)	0.058	2	33.456	.000	.003	.031
Odak3 örneklem oranı (O ₃)	0.054	2	30.749	.000	.003	.029
Odak1 DMF miktarı (b ₁)	1.261	1	1442.722	.000	.072	.412
Odak2 DMF miktarı (b ₂)	1.126	1	1288.010	.000	.065	.384
Odak3 DMF miktarı (b ₃)	1.243	1	1421.901	.000	.071	.408
T x R	1.075	4	307.399	.000	.062	.373
T x O ₁	0.009	8	1.240	.272	.000	.005
T x O ₂	0.008	8	1.195	.298	.000	.005
T x O ₃	0.007	8	0.941	.481	.000	.004
T x b ₁	0.124	4	35.399	.000	.007	.064
T x b ₂	0.122	4	34.949	.000	.007	.063
T x b ₃	0.119	4	33.890	.000	.007	.062
Hata	1.804	2063			.104	
Toplam	17.415					

Tablo 10'daki dört grup ve TB DMF türü için verilen istatistiksel güç sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(4,2063)} = 375.879$, $p < .001$, $\eta^2 = .075$) manidar ve orta düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,2063)} = 5217.641$, $p < .001$, $\eta^2 = .262$) manidar ve yüksek düzey, birinci odak ($F_{(1,2063)} = 1442.722$, $p < .001$, $\eta^2 = .072$), ikinci odak ($F_{(1,2063)} = 1288.010$, $p < .001$, $\eta^2 = .065$) ve üçüncü odak ($F_{(1,2063)} = 1421.901$, $p < .001$, $\eta^2 = .071$) gruba eklenen TB DMF miktarı ana etkilerinin ise manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan birinci odak ($F_{(2,2063)} = 23.864$, $p < .001$, $\eta^2 = .002$), ikinci odak ($F_{(2,2063)} = 33.456$, $p < .001$, $\eta^2 = .003$) ve üçüncü odak ($F_{(2,2063)} = 30.749$, $p < .001$, $\eta^2 = .003$) grubun örneklem

büyükülüğü oranlarına ilişkin ana etkilerin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(4,2063)} = 307.399, p < .001, \eta_p^2 = .373$) manidar ve yüksek düzey, istatistiksel teknik x birinci odak ($F_{(4,2063)} = 35.399, p < .001, \eta_p^2 = .064$), istatistiksel teknik x ikinci odak ($F_{(4,2063)} = 34.949, p < .001, \eta_p^2 = .063$) ve istatistiksel teknik x üçüncü odak ($F_{(4,2063)} = 33.890, p < .001, \eta_p^2 = .062$) gruba eklenen TB DMF miktarı ikili etkileşimlerinin manidar ve orta düzey bir etkiye sahip oldukları görülmektedir. Böylece dört grup için TB DMF türü dikkate alındığında, istatistiksel güç üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .373$). Öte yandan odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de istatistiksel teknik ile etkileşimlerinin ya manidar olmaması ya da hesaplanan etki büyüklüklerinin son derece düşük olması nedeniyle, TB DMF’li maddelerin istatistiksel güç oranları üzerinde bu faktörlerin etkili olmadıkları ifade edilebilir.

Araştırma kapsamında hem üç hem de dört grup için EK 17’de verilen bulgulara göre; testteki maddelerin TB DMF içerdiği durumlarda çoklu grup teknikleri için hesaplanan istatistiksel güç oranlarının .502 ile 1.000 aralığında değiştiği belirlenmiştir. Hesaplanan istatistiksel güç oranları alanyazında önerilen (Lei ve Li, 2013; Scott vd., 2009) kriterlere göre değerlendirildiğinde; $N_R = 600$ ve her iki odak gruba eklenen TB DMF miktarlarının 0.40 olduğu koşuldaki istatistiksel güç oranlarının GLK (.513) için yetersiz, GLR-LRT (.702), GLR-Wald (.700), GMH (.686) ve Lordif (.785) teknikleri için kabul edilebilir, diğer bütün koşullarda ise belirtilen tekniklerin TB DMF’li maddeleri doğru belirleme oranlarının yüksek (%80’in üzerinde) olduğu anlaşılmıştır. Benzer şekilde $N_R = 600$ ve üç odak gruba eklenen TB DMF miktarlarının 0.40 olduğu koşuldaki istatistiksel güç oranlarının GLK (.502) için yetersiz, GLR-LRT (.699), GLR-Wald (.697), GMH (.692) ve Lordif (.798) teknikleri için kabul edilebilir, diğer bütün koşullarda ise belirtilen tekniklerin TB DMF’li maddeleri doğru belirleme oranlarının yüksek (%80’in üzerinde) olduğu tespit edilmiştir. Böylece 100 tekrarın yapıldığı bu araştırmada; $N_R=600$ ve odak gruplara eklenen TB DMF miktarlarının 0.40 olduğu durum haricindeki bütün koşullar için, genel olarak referans grubun örneklem büyüklüğü ve TB DMF miktarı arttıkça gerçekte TB DMF içeren maddelerin çoklu grup teknikleri tarafından %80 ve üzeri olasılıkla DMF’li olarak işaretlenebildiği ortaya konulmuştur.

Simülasyon çalışmalarında .80'in üzerindeki istatistiksel güç oranlarının, tekniklerin performanslarının yüksek düzeyde olduğunu göstermesi nedeniyle (Finch, 2016; Lei ve Li, 2013; Scott vd., 2009; Svetina ve Rutkowski, 2014) yapılan bu araştırma kapsamında ya $N_R = 1800$ olduğu bütün koşullarda ya da $N_R = 600$ iken en az bir odak gruba eklenen TB DMF miktarlarının 0.70 olması durumunda çoklu grup teknikleri için hesaplanan istatistiksel güç oranlarının yükselerek 0.80'in üzerine çıktığı sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında hem üç hem de dört grup için $N_R = 600$ olduğunda, odak gruplara eklenen TB DMF miktarı arttıkça bütün tekniklerin istatistiksel güç oranlarında önemli bir artışın söz konusu olduğu belirlenmiş ve elde edilen bu sonucun alanyazındaki araştırmalar (Battauz, 2019; Finch, 2016; Jin vd., 2015; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) ile desteklendiği gözlemlenmiştir. Öte yandan TB DMF miktarı düşük olsa dahi $N_R = 1800$ olduğunda hem üç grupta [GLK (.941), GLR-LRT (.983), GLR-Wald (.983), GMH (.981), Lordif (.996)] hem de dört grupta [GLK (.950), GLR-LRT (.987), GLR-Wald (.987), GMH (.985), Lordif (.998)] çoklu grup teknikleri için hesaplanan istatistiksel güç oranlarının son derece yüksek olduğu tespit edilmiştir (bkz. EK 17). Böylece grup sayısı üç veya dört olduğunda TB DMF türü için referans grubun örneklem büyüklüğü arttıkça istatistiksel güç oranlarının yükseldiği belirlenmiş ve elde edilen sonucun alanyazındaki araştırmalar ile (Battauz, 2019; Finch, 2016; Oshima vd., 2015; Penfield, 2001) desteklendiği görülmüştür. Bu bulguya dayalı olarak, grup sayısı ve TB DMF miktarı önemli olmaksızın büyük örneklemelerin dikkate alındığı geniş ölçekli gerçek test uygulamalarında, gerçekte DMF'li olan maddelerin çoklu grup teknikleri tarafından genel olarak doğru tespit edilebildiği sonucuna ulaşılmıştır.

Yapılan bu araştırmanın sonuçlarına benzer şekilde alanyazında ikili puanlanan maddeler ve çoklu grupların dikkate alındığı simülasyon çalışmalarında TB DMF miktarı ile referans grubun örneklem büyüklüğünün ayrı ayrı veya birlikte artması durumunda; GLK, GLR ve GMH tekniklerinin istatistiksel güç oranlarındaki artışın yanı sıra GLR ve GMH tekniklerinin benzer ve GLK'dan daha yüksek istatistiksel güç oranlarına sahip oldukları ortaya konulmuştur. Bu bilgiler doğrultusunda çoklu grupların dikkate alındığı önceki araştırma bulgularının (Battauz, 2019; Finch, 2016; Jin vd., 2015; Rollins III, 2018; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) yapılan bu simülasyon çalışması ile desteklendiği sonucuna ulaşılmıştır. Elde edilen bu sonuçları destekler nitelikte olan iki grubun dikkate alındığı başka simülasyon çalışmalarında da TB DMF miktarı ve grupların örneklem büyüklükleri arttığında Lordif

(Schauberger ve Mair, 2020; Tunç, 2016), LK (Sandilands, 2014), MH ve LR (Rockoff, 2018) tekniklerine ilişkin istatistiksel güç oranlarının da yükseldiği tespit edilmiştir.

Araştırma kapsamında grup sayısı önemli olmaksızın TB DMF türü için $N_R=1800$ olduğunda çoklu grup tekniklerine ilişkin istatistiksel güç oranlarının birbirlerine ve 1.00'a oldukça yakın olduğu belirlenirken, $N_R = 600$ olan bütün koşullarda istatistiksel güç oranları açısından Lordif'in en yüksek, GLK'nın ise en düşük performansı gösterdiği tespit edilmiştir (bkz. EK 17). Nitekim $N_R = 600$ ve gruplara eklenen TB DMF miktarlarının 0.40 olması durumunda GLK'nın performansının yetersiz olduğu gözlemlenmiş ve GLK'nın yüksek düzeyde performans gösterebilmesi için ya TB DMF miktarı önemli olmaksızın $N_R = 1800$ ya da $N_R = 600$ olsa dahi odak gruplardan en az birine eklenen TB DMF miktarlarının 0.70 olması gerektiği belirlenmiştir. GLK tekniği ile ilgili yapılan bir çalışmada üç grup için TB DMF miktarı önemli olmaksızın örneklem büyüklüğünün 2000'in üzerinde olması durumunda hesaplanan istatistiksel güç oranlarının 1.00'a yaklaştığı, örneklem büyüklüğü düşük olduğunda ise GLK'nın performansının ciddi miktarda azaldığı açıklanmıştır (Battauz, 2019). Bu bağlamda örneklem büyüklüğünün yetersiz olduğu durumlarda GLK tekniğinin performansının düşük olduğunu destekler nitelikte olan Kim ve diğerlerinin (1995) çalışmalarında da vurgulandığı üzere, bu tekniğin düşük DMF miktarı ile küçük örneklemdeki performansının daha fazla incelenebilmesi amacıyla farklı simülasyon koşullarının dikkate alındığı yeni çalışmaların yapılmasına ihtiyaç olduğu sonucuna ulaşılmıştır.

Araştırma kapsamında grup sayısı önemli olmaksızın, DMF türünün TB olduğu bütün koşullar altında GLR-LRT, GLR-Wald ve GMH tekniklerine ilişkin istatistiksel güç oranlarının birbirlerine son derece yakın olduğu tespit edilmiştir. Alanyazındaki hem ikili grup (MH ve LR) (Kim, 2010) hem de çoklu grup (GMH ve GLR) (Finch, 2016; Rollins III, 2018; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) tekniklerinin dikkate alındığı birçok çalışmada da oldukça benzer sonuçlara ulaşılması nedeniyle elde edilen bulguların ilgili araştırmalarla tutarlı olduğu anlaşılmıştır. Diğer taraftan bazı çalışmalarda da (Gao, 2019; Magis vd., 2011) belirtildiği gibi, TB DMF türünün dikkate alındığı bütün koşullarda GLR-LRT ve GLR-Wald tekniklerinin istatistiksel güç açısından son derece paralel performans gösterdikleri ortaya konulmuştur.

İkili puanlanan maddeler için grup sayısının üç veya dört, DMF türünün ise TB olduğu durumda Lordif tekniğinin performansının incelendiği ilk araştırma niteliğinde olan bu simülasyon çalışmasında; bütün koşullar için .785 ve üzerinde istatistiksel güç

oranlarına sahip olan Lordif'in dikkate alınan diğer tekniklere kıyasla nispeten daha iyi performans sergilediği tespit edilmiştir. Lordif tekniği için elde edilen bu bulguların, ikili puanlanan maddeler ve ikili grupta TB DMF türünü dikkate alan araştırma sonuçları (Schauberger ve Mair, 2020; Tunç, 2016) ile desteklendiği görülmüştür. İlgili araştırmalarda Lordif tekniği için analiz sürecinde yinelemeli bir saflaştırma işleminin otomatik olarak gerçekleştirilmesinin tekniğin performansına olumlu yönde katkı sağladığı vurgulandığından (Choi vd., 2011, 2016; Crane vd., 2006), yapılan bu araştırmadaki bütün koşullarda Lordif tekniğine ilişkin istatistiksel güç oranlarının diğer tekniklere göre daha yüksek olması beklenen bir sonuç olarak kabul edilmiştir.

4.1.5 Üç grupta TB DMF türü için GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı $[b(0.4, 0.7)]$ koşullarındaki Tip III hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TB DMF türü için çoklu grup tekniklerinin performansları Tip III hata açısından incelenmiştir. Üç grup ve TB DMF türü için Tip III hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 11'de sunulmuştur.

Tablo 11

Üç Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	281.086	3	8340.655	.000	.979	.991
Referans örneklem büyüklüğü (R)	0.782	1	69.655	.000	.003	.227
Odak1 örneklem oranı (O ₁)	0.047	2	2.097	.125	.000	.017
Odak2 örneklem oranı (O ₂)	0.036	2	1.601	.204	.000	.013
Odak1 DMF miktarı (b ₁)	0.137	1	12.237	.000	.000	.049
Odak2 DMF miktarı (b ₂)	0.130	1	11.608	.000	.000	.047
T x R	0.498	3	14.774	.000	.002	.158
T x O ₁	0.010	6	0.146	.990	.000	.004
T x O ₂	0.026	6	0.389	.886	.000	.010
T x b ₁	0.393	3	11.664	.000	.001	.129
T x b ₂	0.405	3	12.012	.000	.001	.132
Hata	2.662	237			.009	
Toplam	286.999					

Tablo 11'deki üç grup ve TB DMF türü için verilen Tip III hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(3,237)} = 8340.655$, $p < .001$, $\eta^2 = .979$) manidar ve çok yüksek düzey bir etkiye sahip olduğu anlaşılmaktadır. Diğer taraftan referans grubun örneklem büyüklüğü ($F_{(1,237)} = 69.655$, $p < .001$, $\eta^2 = .003$), birinci odak ($F_{(1,237)} = 12.237$,

$p < .001$, $\eta^2 = .000$) ile ikinci odak ($F_{(1,237)} = 11.608$, $p < .001$, $\eta^2 = .000$) gruba eklenen TB DMF miktarı ana etkilerinin manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(3,237)} = 14.774$, $p < .001$, $\eta_p^2 = .158$) manidar ve yüksek düzey, istatistiksel teknik x birinci odak ($F_{(3,237)} = 11.664$, $p < .001$, $\eta_p^2 = .129$) ve istatistiksel teknik x ikinci odak ($F_{(3,237)} = 12.012$, $p < .001$, $\eta_p^2 = .132$) gruba eklenen TB DMF miktarı ikili etkileşimlerinin manidar ve orta düzey bir etkiye sahip oldukları görülmektedir. Böylece üç grup için TB DMF türü dikkate alındığında, Tip III hata üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .158$). Öte yandan odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de istatistiksel teknik ile etkileşimleri manidar olmaması nedeniyle, TB DMF’li maddelerin Tip III hata oranları üzerinde bu faktörlerin etkili olmadıkları ifade edilebilir.

4.1.6 Dört grupta TB DMF türü için GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı [$b(0.4, 0.7)$] koşullarındaki Tip III hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TB DMF türü için çoklu grup DMF belirleme tekniklerinin performansları Tip III hata açısından incelenmiştir. Dört grup ve TB DMF türü için Tip III hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 12’de sunulmuştur.

Tablo 12

Dört Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	153.823	3	25499.388	.000	.795	.979
Referans örneklem büyüklüğü (R)	7.141	1	3551.067	.000	.037	.684
Odak1 örneklem oranı (O ₁)	0.038	2	9.572	.000	.000	.012
Odak2 örneklem oranı (O ₂)	0.045	2	11.264	.000	.000	.014
Odak3 örneklem oranı (O ₃)	0.044	2	10.938	.000	.000	.013
Odak1 DMF miktarı (b ₁)	0.542	1	269.734	.000	.003	.141
Odak2 DMF miktarı (b ₂)	0.518	1	257.713	.000	.003	.136
Odak3 DMF miktarı (b ₃)	0.539	1	267.857	.000	.003	.140

(Devam ediyor)

Tablo 12 (Devam)

Dört Grupta TB DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
T x R	21.424	3	3551.477	.000	.111	.866
T x O ₁	0.118	6	9.755	.000	.001	.034
T x O ₂	0.132	6	10.950	.000	.001	.038
T x O ₃	0.126	6	10.430	.000	.001	.037
T x b ₁	1.609	3	266.644	.000	.008	.328
T x b ₂	1.553	3	257.380	.000	.008	.320
T x b ₃	1.607	3	266.471	.000	.008	.327
Hata	3.302	1642			.017	
Toplam	193.435					

Tablo 12'deki dört grup ve TB DMF türü için verilen Tip III hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(3,1642)} = 25499.388$, $p < .001$, $\eta^2 = .795$) manidar ve yüksek düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,1642)} = 3551.067$, $p < .001$, $\eta^2 = .037$) manidar ve düşük düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan birinci odak ($F_{(1,1642)} = 269.734$, $p < .001$, $\eta^2 = .003$), ikinci odak ($F_{(1,1642)} = 257.713$, $p < .001$, $\eta^2 = .003$) ve üçüncü odak ($F_{(1,1642)} = 267.857$, $p < .001$, $\eta^2 = .003$) gruba eklenen TB DMF miktarı ana etkilerinin manidar ve çok düşük bir etkiye sahip oldukları görülmektedir. Benzer şekilde birinci odak ($F_{(2,1642)} = 9.572$, $p < .001$, $\eta^2 = .000$), ikinci odak ($F_{(2,1642)} = 11.264$, $p < .001$, $\eta^2 = .000$) ve üçüncü odak ($F_{(2,1642)} = 10.938$, $p < .001$, $\eta^2 = .000$) grubun örneklem büyüklüğü oranlarına ilişkin ana etkilerin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(3,1642)} = 3551.477$, $p < .001$, $\eta_p^2 = .866$) manidar ve yüksek düzey etkiye sahip olduğu görülmektedir. Benzer şekilde istatistiksel teknik x birinci odak ($F_{(3,1642)} = 266.644$, $p < .001$, $\eta_p^2 = .328$), istatistiksel teknik x ikinci odak ($F_{(3,1642)} = 257.380$, $p < .001$, $\eta_p^2 = .320$) ve istatistiksel teknik x üçüncü odak ($F_{(3,1642)} = 266.471$, $p < .001$, $\eta_p^2 = .327$) gruba eklenen TB DMF miktarı ikili etkileşiminin manidar ve yüksek düzey etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan istatistiksel teknik x birinci odak ($F_{(6,1642)} = 9.755$, $p < .05$, $\eta_p^2 = .034$), istatistiksel teknik x ikinci odak ($F_{(6,1642)} = 10.950$, $p < .05$, $\eta_p^2 = .038$) ve istatistiksel teknik x üçüncü odak ($F_{(6,1642)} = 10.430$, $p < .05$, $\eta_p^2 = .037$) grubun örneklem büyüklüğü oranı ikili etkileşiminin manidar, ancak düşük düzey etkiye sahip oldukları gözlemlenmektedir. Böylece dört grup için TB DMF türü dikkate alındığında, Tip III hata üzerindeki en

önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .866$). Bununla birlikte odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de istatistiksel teknik ile etkileşimleri için hesaplanan etki büyüklüklerinin düşük olması nedeniyle, TB DMF’li maddelerin Tip III hata oranları üzerinde bu faktörlerin etkili olmadıkları ifade edilebilir.

Araştırma kapsamında hem üç hem de dört grup için EK 18’de verilen bulgulara göre; TB DMF içeren maddelerin TBO DMF’yi belirlemek amacıyla geliştirilen çoklu grup teknikleri ile analiz edilmesi durumunda hesaplanan Tip III hata oranlarının .041 ile 1.000 aralığında değiştiği ortaya konulmuştur. Hesaplanan Tip III hata oranları açısından yalnızca GLR-LRT ve GLR-Wald için elde edilen değerlerin Bradley (1978) tarafından önerilen liberal kritere (.075) göre iyi kontrol edildiği, GLK ve Lordif için hesaplanan Tip III hata oranlarının ise .075’ten oldukça yüksek olduğu sonucuna ulaşılmıştır. Böylece 100 tekrarın yapıldığı bu çalışmada, grup sayısı önemli olmaksızın TB DMF içeren maddelerin TBO DMF’yi belirlemek amacıyla geliştirilen teknikler tarafından %4.1 ile %100 arasındaki oranlarda DMF’li olarak belirlenebildiği gözlemlenmiştir. Simülasyon çalışmalarında hesaplanan yüksek Tip III hata oranları; gerçekte DMF’li olan maddeler için tespit edilen DMF türünün analizin gerçekleştirildiği istatistiksel teknik tarafından hatalı tanımlandığını gösterdiğinden (Lopez Rivas, 2012), bu çalışmadaki TB DMF’li maddelerin DMF türünün çoklu grup tekniklerinin bazıları ile doğru belirlenemediği sonucu ulaşılmıştır.

Yapılan bu simülasyon çalışmasında üç grup için güçlük parametrelerinin değişimlendiği bütün koşullardaki Tip III hata oranlarının; GLK için .372 ile 1.000 arasında, Lordif içinse .179 ile .209 arasında değişiklik gösterdiği belirlenmiştir. Benzer şekilde dört grupta hesaplanan Tip III hata oranlarının; GLK için .367 ile 1.000 arasında, Lordif içinse .211 ile .243 arasında değişiklik gösterdiği tespit edilmiştir. Bununla birlikte GLR-LRT ve GLR-Wald tekniklerine ilişkin hesaplanan Tip III hata oranlarının da üç grup için .042 ile .054, dört grup için .041 ile .050 arasında değiştiği anlaşılmış ve böylelikle TBO DMF’yi belirlemek amacıyla geliştirilen GLR-LRT ve GLR-Wald tekniklerinin, gerçekte TB DMF’li olan maddeleri de DMF’li olarak belirleyebilme olasılığının yaklaşık %5 olduğu gözlemlenmiştir (bkz. EK 18). Böylece ikili puanlanan maddeler ve çoklu gruplarda istatistiksel tekniklerin performanslarının Tip III hata açısından incelendiği ilk araştırma niteliğinde olan bu çalışma kapsamında; diğer koşulların sabit tutulması şartıyla referans grubun örneklem büyüklüğü ve odak gruplara

eklenen TB DMF miktarı arttığında GLR-LRT ve GLR-Wald için Tip III hata oranlarının liberal kritere (.075) göre kabul edilebilir olduğu ve bütün koşullar için genel olarak sabit kaldığı sonucuna ulaşılmıştır. Bu bulguyu destekler nitelikte olan simülasyon çalışmalarında da (Finch ve French, 2008; Lopez Rivas, 2012) iki grup ve ikili puanlanan maddelerde TB DMF miktarı ve örneklem büyüklüğü artsa dahi KTK'ya dayalı teknikler (örn., LR) için Tip III hata oranlarının ciddi şekilde farklılaşmadığı açıklanmıştır.

Araştırma kapsamında MTK'ya dayalı olan tekniklerin performansları dikkate alındığında ise; grup sayısı önemli olmaksızın GLK ve Lordif için hesaplanan yüksek Tip III hata oranları, TBO DMF'yi belirlemek amacıyla geliştirilen bu tekniklerin TB DMF içeren maddeleri de yüksek oranda DMF'li olarak işaretleyebildiklerini göstermiştir. Ulaşılan sonuçlara göre; Lordif'in performansının bütün koşullarda sabit kaldığı tespit edilmesine rağmen, GLK için elde edilen değerlerin referans grubun örneklem büyüklüğü ve TB DMF miktarından etkilenecek yükseldikleri gözlemlenmiştir. Öte yandan yapılan bu çalışmada TBO DMF türünü belirlemek amacıyla geliştirilen GLR-LRT ve GLR-Wald tekniklerinin TB DMF içeren maddelerdeki Tip III hata oranlarına ilişkin performanslarının bütün koşullarda birbirlerine son derece yakın olduğu tespit edilmiş ve bu teknikler için elde edilen performansların alanyazındaki açıklamalarla (Gao, 2019; Magis vd., 2011) desteklendiği sonucuna ulaşılmıştır.

DMF türünün TB, grup sayısının ise üç olduğu bu simülasyon çalışmasında Lordif için hesaplanan Tip III hata oranlarının, $N_R = 600$ olduğu koşullarda .179 ile .193 arasında, $N_R = 1800$ olduğu koşullarda ise .192 ile .209 arasında değiştiği görülmüştür. Benzer şekilde grup sayısı dört olduğunda hesaplanan Tip III hata oranlarının, $N_R = 600$ olduğunda .211 ile .243 arasında, $N_R = 1800$ olduğunda ise .227 ile .242 arasında farklılaştığı tespit edilmiştir. Bu sonuçlara dayalı olarak grup sayısı önemli olmaksızın referans grubun örneklem büyüklüğü ve TB DMF miktarı artsa dahi Lordif için hesaplanan Tip III hata oranlarının ciddi miktarda değişmediği belirlenmesine rağmen, yine de elde edilen değerlerin liberal kritere göre kabul edilebilir olmadığı görülmüştür. Böylece TBO DMF'yi tespit etmek amacıyla geliştirilen Lordif tekniğinin gerçekte TB DMF'li olan maddeleri de yaklaşık %23 olasılıkla DMF'li olarak işaretleyebilmesi nedeniyle, bu tekniğin kullanıldığı araştırmalarda DMF türünün güvenilir bir şekilde tespit edilemeyebileceği ortaya konulmuştur. Bu anlamda gerçek test uygulamalarında Lordif ile gerçekleştirilen analiz sonuçlarına göre DMF'li olarak belirlenen maddelerin DMF türünün tespit edilebilmesi amacıyla MKE'lere başvurulması gerektiği belirlenmiştir.

Araştırma kapsamında dikkate alınan bir diğer teknik olan GLK için DMF türünün TB, grup sayısının ise üç olduğu koşullarda hesaplanan Tip III hata oranlarının, $N_R = 600$ olduğu koşullarda .372 ile .882 arasında, $N_R = 1800$ olduğu koşullarda ise .911 ile 1.000 arasında değiştiği belirlenmiştir. Benzer şekilde grup sayısı dört olduğunda hesaplanan Tip III hata oranlarının, $N_R = 600$ iken .367 ile .890 arasında $N_R = 1800$ olduğunda ise .917 ile 1.000 arasında farklılaştığı tespit edilmiştir. Bu sonuçlara dayalı olarak grup sayısı önemli olmaksızın referans grubun örneklem büyüklüğü ve odak gruplara eklenen TB DMF miktarı arttıkça GLK için hesaplanan Tip III hata oranlarının yükseldiğinin belirlenmesinin yanı sıra, DMF miktarı önemli olmaksızın $N_R = 1800$ olduğunda TBO DMF'yi tespit etmek amacıyla geliştirilen GLK'nın TB DMF içeren maddeleri de çok yüksek (%90 ve üzeri) olasılıkla DMF'li olarak işaretleyebildiği anlaşılmıştır. Bu bilgiler doğrultusunda TBO(α) DMF'yi belirlemek amacıyla geliştirilen GLK ve Lordif teknikleri ile TB DMF içeren maddelerin de DMF'li olarak işaretlenebildiği anlaşıldığından, büyük örneklemelerin dikkate alındığı geniş ölçekli gerçek test uygulamalarında DMF türünün bu teknikler ile güvenilir bir şekilde belirlenemeyebileceği ortaya konulmuştur. Böylelikle belirtilen teknikler tarafından DMF'li olarak işaretlenen maddelerin DMF türünün doğru tespit edilebilmesi amacıyla MKE'lere başvurulmasının gerekli olduğu sonucuna ulaşılmıştır.

4.1.7 Üç grupta TBO(α) DMF türü için GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı [$\alpha(0.0, 0.4, 0.7)$] koşullarındaki Tip I hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(α) DMF türü için çoklu grup DMF belirleme tekniklerinin performansları Tip I hata açısından incelenmiştir. Üç grup ve TBO(α) DMF türü için Tip I hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 13'te sunulmuştur.

Tablo 13

Üç Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	508.788	3	11226.123	.000	.786	.983
Referans örneklem büyüklüğü (R)	72.093	1	4772.101	.000	.111	.892
Odak1 örneklem oranı (O ₁)	1.791	2	59.274	.000	.003	.171
Odak2 örneklem oranı (O ₂)	1.036	2	34.283	.000	.002	.106
Odak1 DMF miktarı (a ₁)	0.631	2	20.869	.000	.001	.068
Odak2 DMF miktarı (a ₂)	0.787	2	26.059	.000	.001	.083
T x R	46.723	3	1030.917	.000	.072	.843
T x O ₁	0.576	6	6.360	.000	.001	.062
T x O ₂	0.715	6	7.884	.000	.001	.076
T x a ₁	1.314	6	14.491	.000	.002	.131
T x a ₂	1.202	6	13.261	.000	.002	.121
Hata	8.702	576			.013	
Toplam	647.003					

Tablo 13'deki üç grup ve TBO(a) DMF türü için verilen Tip I hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(3,576)} = 11226.123$, $p < .001$, $\eta^2 = .786$) manidar ve yüksek düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,576)} = 4772.101$, $p < .001$, $\eta^2 = .111$) manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan birinci odak ($F_{(2,576)} = 59.274$, $p < .001$, $\eta^2 = .003$) ile ikinci odak ($F_{(2,576)} = 34.283$, $p < .001$, $\eta^2 = .002$) grubun örneklem büyüklüğü oranları ve birinci odak ($F_{(2,576)} = 20.869$, $p < .001$, $\eta^2 = .001$) ile ikinci odak ($F_{(2,576)} = 26.059$, $p < .001$, $\eta^2 = .001$) gruba eklenen TBO(a) DMF miktarı ana etkilerinin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(3,576)} = 1030.917$, $p < .001$, $\eta_p^2 = .843$) manidar ve yüksek düzey bir etkiye sahip olduğu görülmektedir. Diğer taraftan istatistiksel teknik x birinci odak ($F_{(6,576)} = 6.360$, $p < .001$, $\eta_p^2 = .062$), istatistiksel teknik x ikinci odak ($F_{(6,576)} = 7.884$, $p < .001$, $\eta_p^2 = .076$) grup örneklem büyüklüğü oranı, istatistiksel teknik x birinci odak ($F_{(6,576)} = 14.491$, $p < .001$, $\eta_p^2 = .131$) ve istatistiksel teknik x ikinci odak ($F_{(6,576)} = 13.261$, $p < .001$, $\eta_p^2 = .121$) gruba eklenen TBO(a) DMF miktarı ikili etkileşimlerinin manidar ve orta düzey bir etkiye sahip oldukları gözlemlenmektedir. Böylece üç grup için TBO(a) DMF türü dikkate alındığında, Tip I hata üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .843$). Bununla birlikte odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimleri için hesaplanan etki

büyükliklerinin nispeten düşük olması nedeniyle, bu faktörlerin Tip I hata oranları üzerinde etkili olmadıkları ifade edilebilir.

4.1.8 Dört grupta TBO(a) DMF türü için GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı [$a(0.0, 0.4, 0.7)$] koşullarındaki Tip I hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(a) DMF türü için çoklu grup DMF belirleme tekniklerinin performansları Tip I hata açısından incelenmiştir. Dört grup ve TBO(a) DMF türü için Tip I hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 14’te sunulmuştur.

Tablo 14

Dört Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	4822.602	3	256780.045	.000	.827	.993
Referans örneklem büyüklüğü (R)	592.894	1	94706.199	.000	.102	.943
Odak1 örneklem oranı (O ₁)	6.041	2	482.460	.000	.001	.145
Odak2 örneklem oranı (O ₂)	7.144	2	570.551	.000	.001	.167
Odak3 örneklem oranı (O ₃)	6.190	2	494.362	.000	.001	.148
Odak1 DMF miktarı (a ₁)	2.171	2	173.388	.000	.000	.057
Odak2 DMF miktarı (a ₂)	2.657	2	212.170	.000	.000	.069
Odak3 DMF miktarı (a ₃)	3.083	2	246.210	.000	.001	.079
T x R	319.104	3	16990.708	.000	.055	.899
T x O ₁	1.580	6	42.056	.000	.000	.042
T x O ₂	1.374	6	36.588	.000	.000	.037
T x O ₃	1.720	6	45.797	.000	.000	.046
T x a ₁	4.540	6	120.856	.000	.001	.113
T x a ₂	4.651	6	123.822	.000	.001	.115
T x a ₃	4.970	6	132.305	.000	.001	.122
Hata	35.709	5704			.006	
Toplam	5831.002					

Tablo 14’teki dört grup ve TBO(a) DMF türü için verilen Tip I hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(3,5704)} = 256780.045$, $p < .001$, $\eta^2 = .827$) manidar ve yüksek düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,5704)} = 94706.199$, $p < .001$, $\eta^2 = .102$) manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan birinci odak ($F_{(2,5704)} = 482.460$, $p < .001$, $\eta^2 = .001$), ikinci odak ($F_{(2,5704)} = 570.551$, $p < .001$, $\eta^2 = .001$) ve üçüncü odak ($F_{(2,5704)} = 494.362$, $p < .001$, $\eta^2 = .001$) grubun örneklem büyüklüğü oranları ile birinci odak ($F_{(2,5704)} = 173.388$, $p < .001$, $\eta^2 = .000$), ikinci odak ($F_{(2,5704)} = 212.170$, $p < .001$, $\eta^2 = .000$) ve üçüncü odak

($F_{(2,5704)} = 246.210, p < .001, \eta^2 = .001$) gruba eklenen TBO(a) DMF miktarları ana etkilerinin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları görülmektedir.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(3,5704)} = 16990.708, p < .001, \eta_p^2 = .899$) manidar ve yüksek düzey bir etkiye sahip olduğu görülmektedir. İstatistiksel teknik x birinci odak ($F_{(6,5704)} = 120.856, p < .001, \eta_p^2 = .113$), istatistiksel teknik x ikinci odak ($F_{(6,5704)} = 123.822, p < .001, \eta_p^2 = .115$) ve istatistiksel teknik x üçüncü odak ($F_{(6,5704)} = 132.305, p < .001, \eta_p^2 = .122$) gruba eklenen TBO(a) DMF miktarı ikili etkileşiminin manidar ve orta düzey bir etkiye sahip oldukları gözlemlenmektedir. Diğer taraftan istatistiksel teknik x birinci odak ($F_{(6,5704)} = 42.056, p < .001, \eta_p^2 = .042$), istatistiksel teknik x ikinci odak ($F_{(6,5704)} = 36.588, p < .001, \eta_p^2 = .037$) ve istatistiksel teknik x üçüncü odak ($F_{(6,5704)} = 45.797, p < .001, \eta_p^2 = .046$) grubun örneklem büyüklüğü oranı ikili etkileşimlerinin manidar ve düşük düzey bir etkiye sahip oldukları sonucuna ulaşılmaktadır. Böylece dört grup için TBO(a) DMF türü dikkate alındığında, Tip I hata üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .899$). Bununla birlikte odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimleri için hesaplanan etki büyüklüklerinin nispeten düşük olması nedeniyle, bu faktörlerin Tip I hata oranları üzerinde etkili olmadıkları ifade edilebilir.

Araştırma kapsamında hem üç hem de dört grup için EK 19'da verilen bulgulara göre; çoklu grup teknikleri için hesaplanan temel Tip I hata oranlarının .028 ile .061 aralığında değiştiği ortaya konulmuştur. Hesaplanan temel Tip I hata oranlarının genel olarak nominal Tip I hata oranı olan .05'ten düşük veya bu değere makul ölçüde yakın olduğu (Tay vd., 2016) ve Bradley (1978) tarafından önerilen liberal kritere (.075) göre iyi düzeyde kontrol edildiği sonucuna ulaşılmıştır. Böylece 100 tekrarın yapıldığı bu araştırmada; grup sayısı önemli olmaksızın testteki hiçbir madde gerçekte DMF'li olmamasına rağmen, bu maddelerin GLK, GLR-LRT, GLR-Wald ve Lordif teknikleri tarafından %2.8 ile %6.1 arasındaki oranlarda TBO(a) DMF'li olarak belirlenebildiği gözlemlenmiştir. Ayrıca yapılan bu araştırmada grup sayısı sabit tutulduğunda referans grubun örneklem büyüklüğü artmasına rağmen temel Tip I hata oranlarında ciddi bir farklılaşmanın söz konusu olmadığı tespit edilmiştir. Bu bilgiler doğrultusunda grup

sayısının üç veya dört, DMF türünün ise TBO(α) olduğu durumda, çoklu grup tekniklerine ilişkin hesaplanan temel Tip I hata oranlarının .075'in altında yer alması nedeniyle (Cheng vd., 2013; Elosua ve Wells, 2013; Fidalgo ve Scalón, 2010; Finch, 2005; Lopez Rivas, 2012; Rockoff, 2018) TBO(α) DMF türünün incelendiği bütün koşullarda istatistiksel güç oranlarının yorumlanabileceği belirlenmiştir.

Alanyazında TBO(α) DMF türü dikkate alınarak çoklu grup tekniklerine ilişkin performansların incelendiği bazı araştırmalarda; GLK için üç gruptaki temel Tip I hata oranlarının .036 (Batauz, 2016), GLR için dört gruptaki temel Tip I hata oranlarının ise .054 (Jin vd., 2015) olarak hesaplandığı görülmüştür. Yapılan bu simülasyon çalışmasında da TBO(α) DMF türü için hem üç grupta [GLK (.030), GLR-LRT (.045), GLR Wald (.044), Lordif (.060)] hem de dört grupta [GLK (.028), GLR-LRT (.043), GLR Wald (.042), Lordif (.061)] GLK'nın en yüksek, Lordif'in ise en düşük performansa sahip olduğu belirlenmiştir. Grup sayısı önemli olmaksızın GLR-LRT ve GLR-Wald teknikleri için hesaplanan temel Tip I hata oranlarının da GLK'dan yüksek olması neticesinde, elde edilen bulguların alanyazındaki sınırlı sayıda araştırma (Batauz, 2019; Jin vd., 2015) ile desteklendiği tespit edilmiştir. Ayrıca yapılan bu araştırma kapsamında temel Tip I hata oranlarının referans grubun örneklem büyüklüğüne bağlı olarak farklılaşmadığı belirlenmiş ve elde edilen sonucun Batauz'un (2019) çalışması ile tutarlı olduğu görülmüştür. Benzer şekilde ikili grupların ele alındığı başka simülasyon çalışmalarında da testin DMF'li madde içermemesi durumunda grupların örneklem büyüklüğü artmasına rağmen hesaplanan temel Tip I hata oranlarının değişmediği açıklandığından (Finch ve French, 2007; Kopf vd., 2015b; Li ve Zumbo, 2009; Stark vd., 2006) grup sayısının üç veya dört, DMF türünün ise TBO(α) olduğu bütün koşullarda hesaplanan temel Tip I hata oranlarının alanyazındaki araştırmaların sonuçları ile desteklendiği sonucuna ulaşılmıştır.

Araştırma kapsamında temel bir Tip I hatanın yanı sıra belirlenen dört maddenin ayırt edicilik parametrelerinin değişimlendiği koşullarda da Tip I hata hesaplanmıştır. Bu bağlamda referans grubun örneklem büyüklüğü sabit kalmak şartıyla odak gruplara eklenen TBO(α) DMF miktarı veya yine odak gruplara eklenen TBO(α) DMF miktarı sabit kalmak şartıyla referans grubun örneklem büyüklüğü arttığında GLK, GLR-LRT, GLR-Wald ve Lordif için hesaplanan Tip I hata oranlarının da üç grup için .027 ile .081 arasında, dört grup içinse .025 ile .079 arasında değişiklik gösterdiği belirlenmiştir (bkz. EK 19). Böylece grup sayısı önemli olmaksızın elde edilen değerlerin genel olarak kabul edilebilir ve GLK'nın nispeten GLR-LRT, GLR-Wald ve Lordif tekniklerinden daha düşük Tip I hata oranlarına sahip olduğu sonucuna ulaşılmıştır.

Alanyazındaki çalışmalarda da (Gao, 2019; Magis vd., 2011) belirtildiği gibi, saflaştırma işleminin gerçekleştirilmediği bu araştırmada TBO(α) DMF türünün dikkate alındığı bütün koşullarda GLR-LRT ve GLR-Wald teknikleri için Tip I hata oranlarının son derece paralel performans göstererek üç grup için .044 ile .053 arasında, dört grup içinse .040 ile .051 arasında değişmesi nedeniyle, bu tekniklerin performanslarının referans grubun örneklem büyüklüğü ve TBO(α) DMF miktarındaki artıştan önemli miktarda etkilenmediği ortaya konulmuştur. Ayrıca TBO(α) DMF türünün dikkate alındığı bütün koşullardaki Tip I hata oranlarının GLK için üç grupta 0.027 ile 0.030 arasında, dört grupta ise .025 ile .028 arasında değişiklik göstermesi nedeniyle referans grubun örneklem büyüklüğü ve odak gruplara eklenen TBO(α) DMF miktarı artsa dahi GLK, GLR-LRT ve GLR-Wald için hesaplanan Tip I hata oranlarının ciddi bir şekilde değişmediği tespit edilmiştir. Bununla birlikte Lordif tekniği için Tip I hata oranlarının üç grupta .065 ile .081 arasında, dört grupta ise .064 ile .079 arasında değişiklik göstermesine rağmen, $N_R=1800$ ve en az iki odak gruba eklenen TBO(α) DMF miktarlarının 0.70 olma durumu haricinde elde edilen değerlerin yine de kabul edilebilir olduğu belirlenmiştir. Simülasyon çalışmalarında hesaplanan yüksek Tip I hata oranları, tekniklerin performanslarının düşük olduğunu gösterdiğinden (Kim, 2003; Schauburger ve Tutz, 2016) odak gruplara eklenen TBO(α) DMF miktarı ve grupların örneklem büyüklükleri birlikte arttığında Tip I hata oranlarının yükselmesine bağlı olarak Lordif tekniğinin performansının nispeten azaldığı gözlemlenmiştir.

Alanyazında çoklu grup ve TBO(α) DMF türünün birlikte dikkate alındığı sınırlı sayıda araştırmaya rastlanılmıştır. Bu araştırmalardan biri Battauz (2019) tarafından yapılmış ve üç grupta grupların örneklem büyüklüğü artsa dahi GLK için hesaplanan Tip I hata oranlarının farklılaşmadığı ortaya konulmuştur. Dört grup ve TBO(α) DMF türünün dikkate alındığı başka bir çalışmada ise Jin ve diğerleri (2015) tarafından testteki maddelere eklenen TBO(α) DMF miktarı arttığında GLR için hesaplanan Tip I hata oranlarının bu artıştan ciddi miktarda etkilenmediği tespit edilmiştir. Benzer şekilde çoklu grupların ele alındığı Oshima ve diğerlerinin (2015) çalışmalarında, diğer koşullar kontrol edilirse TBO(α) DMF miktarı arttığında MG-NCDIF tekniği için hesaplanan Tip I hata oranlarının temel Tip I hata oranından bir miktar yüksek, ancak kabul edilebilir kritik değerin (.075) altında yer aldıkları belirlenmiştir. Bu bilgiler doğrultusunda; grup sayısının üç veya dört, testin ise TBO(α) DMF’li madde içerdiği koşullarda Lordif tekniğine ilişkin performansların Tip I hata açısından incelendiği ilk çalışma özelliği taşıyan bu araştırmada, referans grubun örneklem büyüklüğü veya odak gruplara eklenen

TBO(a) DMF miktarı arttığında GLK, GLR-LRT ve GLR-Wald teknikleri için hesaplanan Tip I hata oranlarının bu artıştan önemli bir şekilde etkilenmeyerek iyi düzeyde kontrol edildiği, Lordif için hesaplanan Tip I hata oranlarının ise bir miktar yükselmesine rağmen elde edilen değerlerin Bradley (1978) tarafından önerilen liberal kritere (.075) göre genel olarak kabul edilebilir olduğu sonucuna ulaşılmıştır.

4.1.9 Üç grupta TBO(a) DMF türü için GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı [$\alpha(0.4, 0.7)$] koşullarındaki istatistiksel güç oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(a) DMF türü için çoklu grup DMF belirleme tekniklerinin performansları istatistiksel güç açısından incelenmiştir. Üç grup ve TBO(a) DMF türü için istatistiksel güç oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 15'te sunulmuştur.

Tablo 15

Üç Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	0.260	3	129.287	.000	.024	.621
Referans örneklem büyüklüğü (R)	6.309	1	9424.860	.000	.580	.975
Odak1 örneklem oranı (O ₁)	0.099	2	74.200	.000	.009	.385
Odak2 örneklem oranı (O ₂)	0.096	2	71.430	.000	.009	.376
Odak1 DMF miktarı (a ₁)	1.360	1	2031.502	.000	.125	.896
Odak2 DMF miktarı (a ₂)	1.251	1	1869.396	.000	.115	.887
T x R	0.111	3	55.478	.000	.010	.413
T x O ₁	0.002	6	0.553	.767	.000	.014
T x O ₂	0.005	6	1.142	.339	.000	.028
T x a ₁	0.015	3	7.510	.000	.001	.087
T x a ₂	0.006	3	3.183	.025	.001	.039
Hata	0.159	237			.015	
Toplam	10.884					

Tablo 15'teki üç grup ve TBO(a) DMF türü için verilen istatistiksel güç sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(3,237)} = 129.287$, $p < .001$, $\eta^2 = .024$) manidar ve düşük düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,237)} = 9424.860$, $p < .001$, $\eta^2 = .580$) manidar ve yüksek düzey, birinci odak ($F_{(1,237)} = 2031.502$, $p < .001$, $\eta^2 = .125$) ile ikinci odak ($F_{(1,237)} = 1869.396$, $p < .001$, $\eta^2 = .115$) gruba eklenen TBO(a) DMF miktarları ana etkilerinin manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan birinci odak ($F_{(2,237)} = 74.200$, $p < .001$, η^2

= .009) ile ikinci odak ($F_{(2,237)} = 71.430, p < .001, \eta^2 = .009$) grubun örneklem büyüklüğü oranlarına ilişkin ana etkilerin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(3,237)} = 55.478, p < .001, \eta_p^2 = .413$) manidar ve yüksek düzey, istatistiksel teknik x birinci odak gruba eklenen TBO(a) DMF miktarı ikili etkileşiminin ($F_{(3,237)} = 7.510, p < .001, \eta_p^2 = .087$) manidar ve orta düzey bir etkiye ve istatistiksel teknik x ikinci odak gruba eklenen TBO(a) DMF miktarı ikili etkileşiminin ($F_{(3,237)} = 3.183, p < .001, \eta_p^2 = .039$) manidar ve düşük düzey bir etkiye sahip oldukları görülmektedir. Böylece üç grup için TBO(a) DMF türü dikkate alındığında, istatistiksel güç üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .413$). Bununla birlikte odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimlerinin ya manidar olmaması ya da hesaplanan etki büyüklüklerinin nispeten düşük olması nedeniyle, bu faktörlerin istatistiksel güç oranları üzerinde etkili olmadıkları ifade edilebilir.

4.1.10 Dört grupta TBO(a) DMF türü için GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı [$a(0.4, 0.7)$] koşullarındaki istatistiksel güç oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(a) DMF türü için çoklu grup DMF belirleme tekniklerinin performansları istatistiksel güç açısından incelenmiştir. Dört grup ve TBO(a) DMF türü için istatistiksel güç oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 16’da sunulmuştur.

Tablo 16

Dört Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	1.757	3	887.497	.000	.033	.619
Referans örneklem büyüklüğü (R)	32.839	1	49754.584	.000	.625	.968
Odaklı örneklem oranı (O ₁)	0.195	2	147.756	.000	.004	.153

(Devam ediyor)

Tablo 16 (Devam)

Dört Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Odak2 örneklem oranı (O ₂)	0.220	2	166.636	.000	.004	.169
Odak3 örneklem oranı (O ₃)	0.204	2	154.486	.000	.004	.158
Odak1 DMF miktarı (a ₁)	3.106	1	4705.274	.000	.059	.741
Odak2 DMF miktarı (a ₂)	3.069	1	4649.432	.000	.058	.739
Odak3 DMF miktarı (a ₃)	3.095	1	4689.615	.000	.059	.741
T x R	0.959	3	484.190	.000	.018	.469
T x O ₁	0.005	6	1.224	.291	.000	.004
T x O ₂	0.006	6	1.471	.184	.000	.005
T x O ₃	0.007	6	1.700	.117	.000	.006
T x a ₁	0.033	3	16.524	.000	.001	.029
T x a ₂	0.026	3	12.970	.000	.000	.023
T x a ₃	0.036	3	18.261	.000	.001	.032
Hata	1.084	1642			.021	
Toplam	52.564					

Tablo 16'daki dört grup ve TBO(a) DMF türü için verilen istatistiksel güç sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(3,1642)} = 887.497, p < .001, \eta^2 = .033$) manidar ve düşük düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,1642)} = 49754.584, p < .001, \eta^2 = .625$) manidar ve yüksek düzey, birinci odak ($F_{(1,1642)} = 4705.274, p < .001, \eta^2 = .059$), ikinci odak ($F_{(1,1642)} = 4649.432, p < .001, \eta^2 = .058$) ve üçüncü odak ($F_{(1,1642)} = 4689.615, p < .001, \eta^2 = .059$) gruba eklenen TBO(a) DMF miktarları ana etkilerinin ise manidar ve düşük düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan birinci odak ($F_{(2,1642)} = 147.756, p < .001, \eta^2 = .004$), ikinci odak ($F_{(2,1642)} = 166.636, p < .001, \eta^2 = .004$) ve üçüncü odak ($F_{(2,1642)} = 154.486, p < .001, \eta^2 = .004$) grubun örneklem büyüklüğü oranlarına ilişkin ana etkilerin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(3,1642)} = 484.190, p < .001, \eta_p^2 = .469$) manidar ve yüksek düzey, istatistiksel teknik x birinci odak ($F_{(3,1642)} = 16.524, p < .001, \eta_p^2 = .029$), istatistiksel teknik x ikinci odak ($F_{(3,1642)} = 12.970, p < .001, \eta_p^2 = .023$) ve istatistiksel teknik x üçüncü odak ($F_{(3,1642)} = 18.261, p < .001, \eta_p^2 = .032$) gruba eklenen TBO(a) DMF miktarı ikili etkileşimlerinin, manidar ve düşük düzey bir etkiye sahip oldukları görülmektedir. Böylece dört grup için TBO(a) DMF türü dikkate alındığında, istatistiksel güç üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .469$). Öte

yandan odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimlerinin ya manidar olmaması ya da hesaplanan etki büyüklüklerinin nispeten düşük olması nedeniyle, bu faktörlerin istatistiksel güç oranları üzerinde etkili olmadıkları ifade edilebilir.

Araştırma kapsamında hem üç hem de dört grup için EK 20’de verilen bulgulara göre; testteki maddelerin TBO(a) DMF içerdiği durumlarda çoklu grup teknikleri için hesaplanan istatistiksel güç oranlarının .378 ile 1.000 aralığında değiştiği ortaya konulmuştur. Hesaplanan istatistiksel güç oranları alanyazında önerilen (Lei ve Li, 2013; Scott vd., 2009) kriterlere göre değerlendirildiğinde; $N_R = 600$ ve odak gruplara eklenen TBO(a) DMF miktarlarının 0.40 olduğu koşuldaki istatistiksel güç oranlarının üç grupta [GLK (.419), GLR-LRT (.391), GLR-Wald (.378), Lordif (.517)] ve dört grupta [GLK (.414), GLR-LRT (.392), GLR-Wald (.380), Lordif (.521)] dikkate alınan çoklu grup teknikleri için yetersiz, diğer bütün koşullarda ise belirtilen tekniklerin TBO(a) DMF içeren maddeleri doğru belirleme oranlarının en az kabul edilebilir (%60’ın üzerinde) olduğu belirlenmiştir. Böylece 100 tekrarın yapıldığı bu araştırmada; $N_R = 600$ ve odak gruplara eklenen TBO(a) DMF miktarlarının 0.40 olduğu durum haricindeki bütün koşullar için, genel olarak referans grubun örneklem büyüklüğü ve TBO(a) DMF miktarı arttıkça gerçekte TBO(a) DMF’li maddelerin çoklu grup teknikleri tarafından %60 ve üzeri olasılıkla DMF’li olarak işaretlenebildiği tespit edilmiştir. Simülasyon çalışmalarında .80’in üzerindeki istatistiksel güç oranlarının, tekniklerin performanslarının yüksek düzeyde olduğunu göstermesi nedeniyle (Finch, 2016; Lei ve Li, 2013; Scott vd., 2009; Svetina ve Rutkowski, 2014) yapılan bu araştırma kapsamında ya $N_R = 1800$ olduğu bütün koşullarda ya da $N_R = 600$ iken odak gruplara eklenen TBO(a) DMF miktarlarının 0.70 olması durumunda çoklu grup teknikleri için hesaplanan istatistiksel güç oranlarının yükselerek 0.80’in üzerine çıktığı sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında hem üç hem de dört grup için $N_R = 600$ olduğunda, odak gruplara eklenen TBO(a) DMF miktarı arttıkça bütün tekniklerin istatistiksel güç oranlarında önemli bir artışın söz konusu olduğu belirlenmiş ve bu sonucun alanyazındaki araştırmalar (Battauz, 2019; Carroll, 2015; Jin vd., 2015; Oshima vd., 2015; Svetina ve Rutkowski, 2014) ile desteklendiği gözlemlenmiştir. Öte yandan TBO(a) DMF miktarı düşük olsa dahi $N_R = 1800$ olduğunda hem üç grupta [GLK (.909), GLR-LRT (.845), GLR-Wald (.843), Lordif (.921)] hem de dört grupta [GLK (.917), GLR-LRT (.852), GLR-Wald (.852), Lordif (.942)] çoklu grup teknikleri için hesaplanan istatistiksel güç oranlarının son derece yüksek olduğu belirlenmiştir (bkz. EK 20).

Böylece grup sayısı üç veya dört olduğunda TBO(*a*) DMF türü için referans grubun örneklem büyüklüğü arttıkça istatistiksel güç oranlarının yükseldiği belirlenmiş ve elde edilen bu sonucun alanyazındaki araştırmalar tarafından (Battauz, 2019; Carroll, 2015; Oshima vd., 2015) desteklendiği görülmüştür. Bu bulguya dayalı olarak, grup sayısı ve TBO(*a*) DMF miktarı önemli olmaksızın büyük örneklemelerin dikkate alındığı geniş ölçekli gerçek test uygulamalarında, gerçekte DMF’li olan maddelerin çoklu grup teknikleri tarafından genel olarak doğru tespit edilebildiği sonucuna ulaşılmıştır.

Yapılan bu araştırmanın sonuçlarına benzer şekilde alanyazında ikili puanlanan maddeler ve çoklu grupların dikkate alındığı simülasyon çalışmalarında TBO(*a*) DMF miktarı ile referans grubun örneklem büyüklüğünün ayrı ayrı veya birlikte artması durumunda; GLK ve GLR tekniklerinin istatistiksel güç oranlarının yükseldiği açıklanmıştır (Battauz, 2019; Carroll, 2015; Jin vd., 2015; Svetina ve Rutkowski, 2014). GLK tekniği ile ilgili Battauz (2019) tarafından yapılan çalışmada üç grup için örneklem büyüklüğü arttıkça örneklem büyüklüğünün yükseldiği ifade edilmiştir. Çoklu grupların dikkate alındığı başka bir simülasyon çalışmasında Svetina ve Rutkowski (2014) genel olarak odak gruplara eklenen TBO(*a*) DMF miktarı arttıkça GLR tekniği için hesaplanan istatistiksel güç oranlarının da yükseldiğini belirtmiştir. Carroll (2015) ise çalışmasında GLK tekniğinin yeterli performans gösterebilmesi için örneklemin büyük olması gerektiğini aktarmıştır. Ayrıca Oshima ve diğerleri (2015) örneklem büyüklüğü ile TBO(*a*) DMF miktarı arttıkça MG-NCDIF tekniği için hesaplanan istatistiksel güç oranlarının da yükseldiğini belirtmiştir. İlgili araştırmalara benzer şekilde yapılan bu simülasyon çalışmasında da $N_R = 600$ olduğunda istatistiksel tekniklerin TBO(*a*) DMF içeren maddeleri DMF’li olarak belirlemede güçlük çekmeleri nedeniyle, istatistiksel gücün örneklem büyüklüğü ile ilişkili olduğu sonucuna ulaşılmış ve elde edilen bu bilginin ikili grupları dikkate alan araştırmacılar (Cohen, 1988; French ve Maller, 2007; Socha vd., 2015; Wood, 2011; Yu vd., 2008) tarafından da vurgulandığı tespit edilmiştir.

Araştırma kapsamında grup sayısı önemli olmaksızın TBO(*a*) DMF türü için $N_R = 1800$ olduğunda çoklu grup teknikleri için hesaplanan istatistiksel güç oranlarının birbirlerine ve 1.00’a oldukça yakın olduğu belirlenirken, $N_R = 600$ olan bütün koşullarda Lordif’in en yüksek, GLR-LRT ve GLR-Wald tekniklerinin ise birbirlerine benzer (Gao, 2019; Magis vd., 2011) ve diğer iki teknikten biraz daha düşük performansı gösterdikleri sonucuna ulaşılmıştır (bkz. EK 20). Nitekim $N_R = 600$ ve odak gruplara eklenen TBO(*a*) DMF miktarlarının 0.40 olması durumunda dikkate alınan dört tekniğe ilişkin istatistiksel güç oranlarının yetersiz olduğu ve tekniklerin yüksek performans gösterebilmeleri için ya

TBO(a) DMF miktarı önemli olmaksızın $N_R = 1800$ ya da $N_R = 600$ olsa dahi en az iki odak gruba eklenen TBO(a) DMF miktarlarının 0.70 olması gerektiği belirlenmiştir. GLK tekniği ile ilgili yapılan bir çalışmada üç grup için TBO(a) DMF miktarı önemli olmaksızın örneklem büyüklüğünün düşük olması durumunda GLK'nın performansının ciddi miktarda azaldığı açıklanmıştır (Battaaz, 2019). Bu bağlamda örneklem büyüklüğünün yetersiz olduğu durumlarda GLK tekniğinin performansının düşük olduğunu destekler nitelikte olan Kim ve diğerlerinin (1995) çalışmalarında da vurgulandığı üzere, GLK, GLR-LRT, GLR-Wald, Lordif tekniklerinin düşük DMF miktarı ile küçük örneklemdeki performansının daha fazla incelenebilmesi amacıyla farklı simülasyon koşullarının dikkate alındığı yeni çalışmaların yapılmasına ihtiyaç olduğu sonucuna ulaşılmıştır.

İkili puanlanan maddeler için grup sayısının üç veya dört, DMF türünün ise TBO(a) olduğu durumda Lordif tekniğinin performansının incelendiği ilk araştırma niteliğinde olan bu simülasyon çalışmasında; $N_R = 600$ ve odak gruplara eklenen TBO(a) DMF miktarlarının 0.40 olduğu durum haricindeki bütün koşullar için .764'ün üzerinde istatistiksel güç oranlarına sahip olan Lordif'in diğer tekniklere kıyasla nispeten daha iyi performans gösterdiği tespit edilmiştir. Lordif tekniği için yinelemeli bir saflaştırma işleminin otomatik olarak gerçekleştirilmesinin tekniğin performansına olumlu yönde katkı sağladığı vurgulandığından (Choi vd., 2011, 2016; Crane vd., 2006), yapılan bu araştırmadaki bütün koşullarda Lordif tekniğine ilişkin istatistiksel güç oranlarının diğer tekniklere göre daha yüksek olması beklenen bir sonuç olarak kabul edilmiştir.

4.1.11 Üç grupta TBO(a) DMF türü için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı [$a(0.4, 0.7)$] koşullarındaki Tip III hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(a) DMF türü için çoklu grup DMF belirleme tekniklerinin performansları Tip III hata açısından incelenmiştir. Üç grup ve TBO(a) DMF türü için Tip III hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 17'de sunulmuştur.

Tablo 17

Üç Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	6.823	4	455.580	.000	.180	.858
Referans örneklem büyüklüğü (R)	20.777	1	5549.157	.000	.547	.949
Odak1 örneklem oranı (O ₁)	0.302	2	40.376	.000	.008	.212
Odak2 örneklem oranı (O ₂)	0.295	2	39.426	.000	.008	.208
Odak1 DMF miktarı (a ₁)	2.495	1	666.274	.000	.066	.689
Odak2 DMF miktarı (a ₂)	2.498	1	667.228	.000	.066	.689
T x R	2.162	4	144.363	.000	.057	.657
T x O ₁	0.003	8	0.111	.999	.000	.003
T x O ₂	0.032	8	1.078	.379	.001	.028
T x a ₁	0.148	4	9.875	.000	.004	.116
T x a ₂	0.168	4	11.213	.000	.004	.130
Hata	1.127	301			.030	
Toplam	37.975					

Tablo 17’deki üç grup ve TBO(a) DMF türü için verilen Tip III hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(4,301)} = 455.580$, $p < .001$, $\eta^2 = .180$) ve referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,301)} = 5549.157$, $p < .001$, $\eta^2 = .547$) manidar ve yüksek düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan birinci odak ($F_{(1,301)} = 666.274$, $p < .001$, $\eta^2 = .066$) ile ikinci odak ($F_{(1,301)} = 667.228$, $p < .001$, $\eta^2 = .066$) gruba eklenen TBO(a) DMF miktarları ana etkilerinin manidar ve orta düzey, birinci odak ($F_{(2,301)} = 40.376$, $p < .001$, $\eta^2 = .008$) ile ikinci odak ($F_{(2,301)} = 39.426$, $p < .001$, $\eta^2 = .008$) grubun örneklem büyüklüğü oranlarına ilişkin ana etkilerin ise manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(4,301)} = 144.363$, $p < .001$, $\eta_p^2 = .657$) manidar ve yüksek düzey, istatistiksel teknik x birinci odak ($F_{(4,301)} = 9.875$, $p < .001$, $\eta_p^2 = .116$) ve istatistiksel teknik x ikinci odak ($F_{(4,301)} = 11.213$, $p < .001$, $\eta_p^2 = .130$) gruba eklenen TBO(a) DMF miktarı ikili etkileşimlerinin manidar ve orta düzey bir etkiye sahip oldukları görülmektedir. Böylece üç grup için TBO(a) DMF türü dikkate alındığında, Tip III hata üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .657$). Öte yandan “odak grupların örneklem büyüklüklerinin referans gruba oranı” faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimlerinin ya manidar olmaması ya da hesaplanan etki büyüklüklerinin nispeten düşük olması nedeniyle, bu faktörlerin Tip III hata oranları üzerinde etkili olmadıkları ifade edilebilir.

4.1.12 Dört grupta TBO(a) DMF türü için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı $[b(0.4, 0.7)]$ koşullarındaki Tip III hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(a) DMF türü için çoklu grup tekniklerinin performansları Tip III hata açısından incelenmiştir. Dört grup ve TBO(a) DMF türü için Tip III hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 18’de sunulmuştur.

Tablo 18

<i>Dört Grupta TBO(a) DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)</i>						
	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	691.366	4	10527.478	.000	.320	.953
Referans örneklem büyüklüğü (R)	1174.845	1	71557.824	.000	.544	.972
Odak1 örneklem oranı (O ₁)	5.042	2	153.543	.000	.002	.130
Odak2 örneklem oranı (O ₂)	4.800	2	146.177	.000	.002	.124
Odak3 örneklem oranı (O ₃)	4.512	2	137.408	.000	.002	.118
Odak1 DMF miktarı (a ₁)	64.716	1	3941.734	.000	.030	.656
Odak2 DMF miktarı (a ₂)	68.117	1	4148.872	.000	.032	.668
Odak3 DMF miktarı (a ₃)	62.771	1	3823.278	.000	.029	.650
T x R	15.730	4	239.517	.000	.007	.317
T x O ₁	0.042	8	0.316	.960	.000	.001
T x O ₂	0.177	8	1.344	.217	.000	.005
T x O ₃	0.028	8	0.217	.988	.000	.001
T x a ₁	0.631	4	9.604	.000	.000	.018
T x a ₂	0.946	4	14.409	.000	.000	.027
T x a ₃	1.043	4	15.886	.000	.000	.030
Hata	33.871	2063			.016	
Toplam	2159.003					

Tablo 18’de verilen Tip III hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(4,2063)} = 10527.478$, $p < .001$, $\eta^2 = .320$) manidar ve yüksek düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,2063)} = 71557.824$, $p < .001$, $\eta^2 = .544$) manidar ve yüksek düzey, birinci odak ($F_{(1,2063)} = 3941.734$, $p < .001$, $\eta^2 = .030$), ikinci odak ($F_{(1,2063)} = 4148.872$, $p < .001$, $\eta^2 = .032$) ve üçüncü odak ($F_{(1,2063)} = 3823.278$, $p < .001$, $\eta^2 = .029$) gruba eklenen TBO(a) DMF miktarları ana etkilerinin ise manidar ve düşük düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan birinci odak ($F_{(2,2063)} = 153.543$, $p < .001$, $\eta^2 = .002$), ikinci odak ($F_{(2,2063)} = 146.177$, $p < .001$, $\eta^2 = .002$) ve üçüncü odak ($F_{(2,2063)} = 137.408$, $p < .001$, $\eta^2 = .002$) grubun örneklem büyüklüğü oranlarına ilişkin ana etkilerin manidar, ancak çok düşük bir etkiye sahip oldukları görülmektedir.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ikili etkileşiminin ($F_{(4,2063)} = 239.517, p < .001, \eta_p^2 = .317$) manidar ve yüksek düzey, istatistiksel teknik x birinci odak ($F_{(4,2063)} = 9.604, p < .001, \eta_p^2 = .018$), istatistiksel teknik x ikinci odak ($F_{(4,2063)} = 14.409, p < .001, \eta_p^2 = .027$) ve istatistiksel teknik x üçüncü odak ($F_{(4,2063)} = 15.886, p < .001, \eta_p^2 = .030$) gruba eklenen TBO(a) DMF miktarı ikili etkileşimlerinin, manidar ve düşük düzey bir etkiye sahip oldukları görülmektedir. Böylece dört grup için TBO(a) DMF türü dikkate alındığında, Tip III hata üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .317$). Öte yandan “odak grupların örneklem büyüklüklerinin referans gruba oranı” faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimlerinin ya manidar olmaması ya da hesaplanan etki büyüklüklerinin oldukça düşük olması nedeniyle, bu faktörlerin Tip III hata oranları üzerinde etkili olmadıkları ifade edilebilir.

Araştırma kapsamında hem üç hem de dört grup için EK 21’de verilen bulgulara göre; TBO(a) DMF içeren maddelerin TB çoklu grup teknikleri ile analiz edilmesi sonucunda hesaplanan Tip III hata oranlarının .098 ile .989 aralığında değiştiği tespit edilmiştir. Hesaplanan Tip III hata oranları açısından GLK, GLR-LRT, GLR-Wald, GMH ve Lordif için elde edilen değerlerin Bradley (1978) tarafından önerilen liberal kriterden oldukça yüksek olduğu sonucuna ulaşılmıştır. Böylece 100 tekrarın yapıldığı bu araştırmada, grup sayısı önemli olmaksızın TBO(a) DMF içeren maddelerin TB DMF’yi belirlemek amacıyla geliştirilen teknikler tarafından %9.8 ile %98.9 arasındaki oranlarda DMF’li olarak belirlenebildiği gözlemlenmiştir. Simülasyon çalışmalarında hesaplanan yüksek Tip III hata oranları; gerçekte DMF’li olan maddeler için tespit edilen DMF türünün analizin gerçekleştirildiği istatistiksel teknik tarafından hatalı tanımlandığını gösterdiğinden (Lopez Rivas, 2012), bu araştırmadaki TBO(a) DMF’li maddelerin DMF türünün genel olarak çoklu grup teknikleri ile doğru belirlenemediği ortaya konulmuştur.

Yapılan bu simülasyon çalışmasında üç grupta belirlenen dört maddenin ayırt edicilik parametrelerinin değişimlendiği bütün koşullardaki Tip III hata oranlarının; GLR-LRT, GLR-Wald, GMH ve Lordif için .347 ile .989 arasında, GLK içinse .101 ile .694 arasında ve diğerlerinden daha düşük olduğu belirlenmiştir. Benzer şekilde dört grupta hesaplanan Tip III hata oranlarının; GLR-LRT, GLR-Wald, GMH ve Lordif için .327 ile .989 arasında, GLK içinse .098 ile .669 arasında ve yine diğerlerinden daha düşük olduğu tespit edilmiştir (bkz. EK 21). Böylece ikili puanlanan maddeler ve çoklu

gruplarda istatistiksel tekniklerin performanslarının Tip III hata açısından incelendiği ilk araştırma niteliğinde olan bu çalışma kapsamında; grup sayısının üç veya dört, DMF türünün ise TBO(a) olduğu durum için hesaplanan yüksek Tip III hata oranları, TB DMF’li maddeleri belirlemek amacıyla geliştirilen GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin gerçekte TBO(a) DMF içeren maddeleri de yüksek oranda DMF’li olarak tespit edebildiğini göstermiştir. Benzer şekilde iki grupta Tip III hatanın hesaplandığı araştırmalarda da (Finch ve French, 2008; Lopez Rivas 2012) TB DMF’yi tespit etmek amacıyla geliştirilen tekniklerin (MH, LR ve MTKOO) testteki TBO(a) DMF’li maddeleri de yanlışlıkla DMF’li olarak belirleme oranlarının yüksek olduğu belirlenmiş ve böylelikle çalışmalardan elde edilen bulguların desteklendiği anlaşılmıştır.

Araştırma kapsamında hesaplanan Tip III hata oranlarına göre; grup sayısı önemli olmaksızın GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için elde edilen değerlerin bütün koşullar altında birbirlerine son derece yakın ve GLK’dan oldukça yüksek olduğu belirlenmiştir. MTK’ya dayalı olan GLK ve Lordif tekniklerine ilişkin Tip III hata oranları dikkate alındığında; Lordif tekniğinin GLK’dan ziyade, KTK’ya dayalı oldukları bilinen GLR-LRT, GLR-Wald ve GMH teknikleri ile benzer performans gösterdiği tespit edilmiştir. Elde edilen bu ilginç bulgu Lordif’in eşleştirme kriteri olarak MTK’ya dayalı kestirimleri kullanmasına rağmen, kuramsal temellerinin LR’ye dayalı geliştirilmesi nedeniyle GLR-LRT, GLR-Wald ve GMH teknikleri ile benzer ve bu üç teknikle son derece paralel performans göstermesi, Lordif ile ilgili yeni çalışmaların yapılmasına ihtiyaç olduğunu ortaya koymuştur. Öte yandan yapılan bu çalışmada TB DMF türünü belirlemek amacıyla geliştirilen GLR-LRT, GLR-Wald ve GMH tekniklerinin TBO(a) DMF içeren maddelerdeki Tip III hata oranlarına ilişkin performanslarının bütün koşullarda birbirlerine son derece yakın olduğu tespit edilmiş ve bu teknikler için elde edilen performansların alanyazındaki açıklamalarla (Finch, 2016; Gao, 2019; Magis vd., 2011) desteklendiği sonucuna ulaşılmıştır.

Grup sayısının üç veya dört, DMF türünün ise TBO(a) olduğu bu simülasyon çalışmasında diğer koşullar sabit kalmak şartıyla referans grubun örneklem büyüklüğü ve odak gruplara eklenen TBO(a) DMF miktarı arttığında, TB DMF’li maddeleri belirlemek amacıyla geliştirilen GLK, GLR-LRT, GLR-Wald, GMH ve Lordif için hesaplanan Tip III hata oranlarının ciddi miktarda yükseldiği tespit edilmiştir (bkz. EK 21). Grup sayısı önemli olmaksızın $N_R = 1800$ ve odak gruplara eklenen TBO(a) DMF miktarlarının 0.70 olduğu koşulda GLR-LRT, GLR-Wald, GMH ve Lordif için hesaplanan Tip III hata oranlarının 1.00’a oldukça yakın olduğu bulunmuştur. Böylece örneklem büyüklüğü ve

DMF miktarının artması durumunda belirtilen dört tekniğin, testteki TBO(*a*) DMF’li maddeleri de çok yüksek (%95 ve üzeri) olasılıkla DMF’li olarak belirleyebildiği ortaya konulmuş ve ulaşılan bu sonucun Finch ve French (2008) tarafından yapılan iki gruplu çalışma ile de desteklendiği görülmüştür. Bu bağlamda TB DMF’yi belirlemek amacıyla geliştirilen GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri ile TBO(*a*) DMF’li maddelerin de DMF’li olarak işaretlenebildiği anlaşıldığından, büyük örneklemelerin dikkate alındığı geniş ölçekli gerçek test uygulamalarında DMF türünün çoklu grup teknikleri ile güvenilir bir şekilde belirlenemeyebileceği ortaya konulmuştur. Böylelikle belirtilen teknikler tarafından DMF’li olarak işaretlenen maddelerin DMF türünün doğru tespit edilebilmesi amacıyla MKE’lere başvurulmasının gerekli olduğu belirlenmiştir.

4.1.13 Üç grupta TBO(*ab*) DMF türü için GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı [*a*(0.0, 0.4, 0.7), *b*(0.0, 0.4, 0.7)] koşullarındaki Tip I hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(*ab*) DMF türü için çoklu grup tekniklerinin performansları Tip I hata açısından incelenmiştir. Üç grup ve TBO(*ab*) DMF türü için Tip I hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 19’da sunulmuştur.

Tablo 19

Üç Grupta TBO(ab) DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)

	Kareler Toplamı	<i>sd</i>	<i>F</i>	<i>p</i>	η^2	η_p^2
Teknik (T)	2225.020	3	8004.751	.000	.476	.841
Referans örneklem büyüklüğü (R)	497.814	1	5372.821	.000	.106	.541
Odak1 örneklem oranı (O ₁)	13.542	2	73.077	.000	.003	.031
Odak2 örneklem oranı (O ₂)	10.274	2	55.444	.000	.002	.024
Odak1 DMF miktarı (a ₁)	4.528	2	24.433	.000	.001	.011
Odak2 DMF miktarı (a ₂)	4.489	2	24.225	.000	.001	.011
Odak1 DMF miktarı (b ₁)	157.687	2	850.944	.000	.034	.272
Odak2 DMF miktarı (b ₂)	140.797	2	759.800	.000	.030	.250
T x R	436.179	3	1569.202	.000	.093	.508
T x O ₁	8.312	6	14.952	.000	.002	.019
T x O ₂	8.986	6	16.164	.000	.002	.021
T x a ₁	7.749	6	13.939	.000	.002	.018
T x a ₂	8.498	6	15.286	.000	.002	.020
T x b ₁	255.142	6	458.951	.000	.055	.377
T x b ₂	263.615	6	474.193	.000	.056	.385
Hata	421.854	4553			.090	
Toplam	4588.565					

Tablo 19'daki üç grup ve TBO(ab) DMF türü için verilen Tip I hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(3,4553)} = 8004.751, p < .001, \eta^2 = .476$) manidar ve yüksek düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,4553)} = 5372.821, p < .001, \eta^2 = .106$) manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Ayrıca birinci odak ($F_{(2,4553)} = 850.944, p < .001, \eta^2 = .034$) ile ikinci odak ($F_{(2,4553)} = 759.800, p < .001, \eta^2 = .030$) grubun madde güçlük parametrelerine eklenen DMF miktarı ana etkilerinin de istatistiksel olarak manidar ve düşük düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan birinci odak ($F_{(2,4553)} = 73.077, p < .001, \eta^2 = .003$) ile ikinci odak ($F_{(2,4553)} = 55.444, p < .001, \eta^2 = .002$) grubun örneklem büyüklüğü oranları ve birinci odak ($F_{(2,4553)} = 24.433, p < .001, \eta^2 = .001$) ile ikinci odak ($F_{(2,4553)} = 24.225, p < .001, \eta^2 = .001$) grubun madde ayırt edicilik parametrelerine eklenen DMF miktarı ana etkilerinin istatistiksel olarak manidar, ancak çok düşük etkiye sahip oldukları görülmektedir.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ($F_{(3,4553)} = 1569.202, p < .001, \eta_p^2 = .508$), istatistiksel teknik x birinci odak grubun ($F_{(6,4553)} = 458.951, p < .001, \eta_p^2 = .377$) ve istatistiksel teknik x ikinci odak grubun ($F_{(6,4553)} = 474.193, p < .001, \eta_p^2 = .385$) madde güçlük parametrelerine eklenen DMF miktarlarına ilişkin ikili etkileşimlerinin istatistiksel olarak manidar ve yüksek düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan istatistiksel teknik x birinci odak grup ($F_{(6,4553)} = 14.952, p < .001, \eta_p^2 = .019$) ile istatistiksel teknik x ikinci odak grup ($F_{(6,4553)} = 16.164, p < .001, \eta_p^2 = .021$) örneklem büyüklüğü oranları ve istatistiksel teknik x birinci odak ($F_{(6,4553)} = 13.939, p < .001, \eta_p^2 = .018$) ile istatistiksel teknik x ikinci odak ($F_{(6,4553)} = 15.286, p < .001, \eta_p^2 = .020$) grubun madde ayırt edicilik parametrelerine eklenen DMF miktarlarına ilişkin ikili etkileşimlerinin istatistiksel olarak manidar ve düşük düzey bir etkiye sahip oldukları gözlemlenmektedir. Böylece üç grup için TBO(ab) DMF türü dikkate alındığında, Tip I hata üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .508$). Bununla birlikte odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimleri için hesaplanan etki büyüklüklerinin nispeten düşük olması nedeniyle, bu faktörlerin Tip I hata oranları üzerinde etkili olmadıkları ifade edilebilir.

4.1.14 Dört grupta TBO(*ab*) DMF türü için GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı [$a(0.0, 0.4, 0.7)$, $b(0.0, 0.4, 0.7)$] koşullarındaki Tip I hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(*ab*) DMF türü için çoklu grup tekniklerinin performansları Tip I hata açısından incelenmiştir. Dört grup ve TBO(*ab*) DMF türü için Tip I hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 20’de sunulmuştur.

Tablo 20

Dört Grupta TBO(ab) DMF Türü için Faktöriyel ANOVA Sonucu (Tip I Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	66330.651	3	360406.205	.000	.599	.907
Referans örneklem büyüklüğü (R)	13547.067	1	220823.110	.000	.122	.666
Odak1 örneklem oranı (O ₁)	155.918	2	1270.762	.000	.001	.022
Odak2 örneklem oranı (O ₂)	165.251	2	1346.829	.000	.001	.024
Odak3 örneklem oranı (O ₃)	153.041	2	1247.318	.000	.001	.022
Odak1 DMF miktarı (a ₁)	35.342	2	288.043	.000	.000	.005
Odak2 DMF miktarı (a ₂)	39.591	2	322.677	.000	.000	.006
Odak3 DMF miktarı (a ₃)	43.581	2	355.191	.000	.000	.006
Odak1 DMF miktarı (b ₁)	1342.352	2	10940.462	.000	.012	.165
Odak2 DMF miktarı (b ₂)	1311.501	2	10689.022	.000	.012	.162
Odak3 DMF miktarı (b ₃)	1290.096	2	10514.566	.000	.012	.160
T x R	10835.482	3	58874.366	.000	.098	.615
T x O ₁	100.656	6	273.455	.000	.001	.015
T x O ₂	95.483	6	259.403	.000	.001	.014
T x O ₃	103.177	6	280.306	.000	.001	.015
T x a ₁	128.343	6	348.674	.000	.001	.019
T x a ₂	123.079	6	334.373	.000	.001	.018
T x a ₃	126.086	6	342.542	.000	.001	.018
T x b ₁	1849.086	6	5023.486	.000	.017	.214
T x b ₂	1855.056	6	5039.705	.000	.017	.215
T x b ₃	1868.865	6	5077.219	.000	.017	.216
Hata	6783.131	110568			.061	
Toplam	110210.900					

Tablo 20’deki dört grup ve TBO(*ab*) DMF türü için verilen Tip I hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(3,110568)} = 360406.205$, $p < .001$, $\eta^2 = .599$) manidar ve yüksek düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,110568)} = 220823.110$, $p < .001$, $\eta^2 = .122$) manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Ayrıca birinci odak ($F_{(2,110568)} = 10940.462$, $p < .001$, $\eta^2 = .012$), ikinci odak ($F_{(2,110568)} = 10689.022$, $p < .001$, $\eta^2 = .012$) ve üçüncü odak ($F_{(2,110568)} = 10514.566$, $p < .001$, $\eta^2 = .012$) grubun madde güçlük parametrelerine eklenen DMF miktarları ana etkilerinin istatistiksel olarak manidar ve düşük düzey bir etkiye sahip oldukları

görülmektedir. Diğer taraftan birinci odak ($F_{(2,110568)} = 1270.762, p < .001, \eta^2 = .001$), ikinci odak ($F_{(2,110568)} = 1346.829, p < .001, \eta^2 = .001$) ve üçüncü odak ($F_{(2,110568)} = 1247.318, p < .001, \eta^2 = .001$) grubun örneklem büyüklüğü oranları ve birinci odak ($F_{(2,110568)} = 288.043, p < .001, \eta^2 = .000$), ikinci odak ($F_{(2,110568)} = 322.677, p < .001, \eta^2 = .000$) ve üçüncü odak ($F_{(2,110568)} = 355.191, p < .001, \eta^2 = .000$) grubun madde ayırt edicilik parametrelerine eklenen DMF miktarları ana etkilerinin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ($F_{(3,110568)} = 58874.366, p < .001, \eta_p^2 = .615$), istatistiksel teknik x birinci odak ($F_{(6,110568)} = 5023.486, p < .001, \eta_p^2 = .214$), istatistiksel teknik x ikinci odak ($F_{(6,110568)} = 5039.705, p < .001, \eta_p^2 = .215$) ve istatistiksel teknik x üçüncü odak ($F_{(6,110568)} = 5077.219, p < .001, \eta_p^2 = .216$) grubun madde güçlük parametrelerine eklenen DMF miktarlarına ilişkin ikili etkileşimlerinin, istatistiksel olarak manidar ve yüksek düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan istatistiksel teknik x birinci odak ($F_{(6,110568)} = 273.455, p < .001, \eta_p^2 = .015$), istatistiksel teknik x ikinci odak ($F_{(6,110568)} = 259.403, p < .001, \eta_p^2 = .014$) ve istatistiksel teknik x üçüncü odak ($F_{(6,110568)} = 280.306, p < .001, \eta_p^2 = .015$) grubun örneklem büyüklüğü oranları ve istatistiksel teknik x birinci odak ($F_{(6,110568)} = 348.674, p < .001, \eta_p^2 = .019$), istatistiksel teknik x ikinci odak ($F_{(6,110568)} = 334.373, p < .001, \eta_p^2 = .018$) ve istatistiksel teknik x üçüncü odak ($F_{(6,110568)} = 342.542, p < .001, \eta_p^2 = .018$) grubun madde ayırt edicilik parametrelerine eklenen DMF miktarlarına ilişkin ikili etkileşimlerinin, istatistiksel olarak manidar ve düşük düzey bir etkiye sahip oldukları gözlemlenmektedir. Böylece dört grup için TBO(ab) DMF türü dikkate alındığında, Tip I hata üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .615$). Öte yandan odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimleri için hesaplanan etki büyüklüklerinin düşük olması nedeniyle, bu faktörlerin Tip I hata oranları üzerinde etkili olmadıkları ifade edilebilir.

Araştırma kapsamında hem üç hem de dört grup için EK 22’de verilen bulgulara göre; çoklu grup teknikleri için hesaplanan temel Tip I hata oranlarının .028 ile .061 aralığında değiştiği ortaya konulmuştur. Hesaplanan temel Tip I hata oranlarının genel

olarak nominal Tip I hata oranı olan .05'ten düşük veya bu değere makul ölçüde yakın olduğu (Tay vd., 2016) ve Bradley (1978) tarafından önerilen liberal kritere (.075) göre iyi düzeyde kontrol edildiği sonucuna ulaşılmıştır. Böylece 100 tekrarın yapıldığı bu araştırmada; grup sayısı önemli olmaksızın testteki hiçbir madde gerçekte DMF'li olmamasına rağmen, bu maddelerin GLK, GLR-LRT, GLR-Wald ve Lordif teknikleri tarafından %2.8 ile %6.1 arasındaki oranlarda TBO(*ab*) DMF'li olarak belirlenebildiği gözlemlenmiştir. Ayrıca yapılan bu araştırmada grup sayısı sabit tutulduğunda referans grubun örneklem büyüklüğü artmasına rağmen temel Tip I hata oranlarında ciddi bir farklılaşmanın söz konusu olmadığı tespit edilmiştir. Çok sayıda araştırmada da belirtildiği gibi (Cheng vd., 2013; Elosua ve Wells, 2013; Fidalgo ve Scalón, 2010; Finch, 2005; Kim, 2003; Lopez Rivas, 2012; Rockoff, 2018; Socha vd., 2015) bu çalışmada grup sayısının üç veya dört, DMF türünün ise TBO(*ab*) olduğu koşullarda çoklu grup tekniklerine ilişkin temel Tip I hata oranlarının .075'in altında yer alması nedeniyle bu DMF türünün incelendiği bütün koşullarda istatistiksel güç oranlarının yorumlanabileceği sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında TBO(*ab*) DMF türü için hesaplanan temel Tip I hata oranlarına göre hem üç grupta [GLK (.030), GLR-LRT (.045), GLR Wald (.044), Lordif (.060)] hem de dört grupta [GLK (.028), GLR-LRT (.043), GLR Wald (.042), Lordif (.061)] GLK'nın en yüksek, Lordif'in ise en düşük performansa sahip olduğu tespit edilmiştir. Alanyazında çoklu gruplar için yapılmış olan DMF belirleme çalışmalarında GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin TBO(*ab*) DMF içeren maddelerdeki performanslarının incelendiği herhangi bir araştırma bulunmamasına rağmen, Dell-Ross'un (2021) çalışmasında GMH tekniğinin TBO(*ab*) DMF türündeki performansının araştırıldığı belirlenmiştir. Belirtilen araştırmada GMH tekniği için hesaplanan temel Tip I hata oranlarının hem üç hem de beş grupta .004'ün altında olduğu gözlemlenmiştir. Çoklu gruplarda TBO(*ab*) DMF türünün dikkate alındığı Oshima ve diğerleri (2015) tarafından yapılan başka bir çalışmada ise MG-NCDIF tekniği için hesaplanan temel Tip I hata oranlarının üç grupta .020, beş grupta ise .030 olduğu görülmüştür. Yapılan bu araştırmada da GLK için elde edilen değerlerin düşük olması nedeniyle MTK'ya dayalı olan GLK ve MG-NCDIF tekniklerinin birbirlerine yakın performans gösterdikleri ortaya konulmuştur. Ayrıca yapılan bu araştırma kapsamında temel Tip I hata oranlarının referans grubun örneklem büyüklüğüne bağlı olarak farklılaşmadığı belirlenmiştir. Benzer şekilde ikili grupların ele alındığı başka simülasyon çalışmalarında da testin DMF'li madde içermemesi durumunda grupların örneklem

büyükliği artmasına rağmen temel Tip I hata oranlarının değişmediği açıklandığından (Finch ve French, 2007; Kopf vd., 2015b; Li ve Zumbo, 2009; Stark vd., 2006) grup sayısının üç veya dört, DMF türünün ise TBO(ab) olduğu bütün koşullardaki temel Tip I hata oranlarının ilgili araştırmaların sonuçları ile desteklendiği sonucuna ulaşılmıştır.

Araştırma kapsamında temel bir Tip I hatanın yanı sıra belirlenen dört maddenin güçlük ve ayırt edicilik parametrelerinin değişimlendiği koşullarda da Tip I hata hesaplanmıştır. Saflaştırma işleminin gerçekleştirilmediği bu araştırmadaki analiz sonuçlarına göre TBO(ab) DMF türünün dikkate alındığı bütün koşullarda GLR-LRT ve GLR-Wald teknikleri için Tip I hata oranlarının üç grupta .040 ile .052 arasında, dört grupta ise .039 ile .051 arasında değişmesi nedeniyle, benzer araştırmalarda da (Gao, 2019; Magis vd., 2011) açıklandığı gibi, bu iki tekniğin son derece paralel performans gösterdiği ve ayrıca bu tekniklerin performanslarının referans grubun örneklem büyüklüğü ile TBO(ab) DMF miktarındaki artıştan önemli miktarda etkilenmediği ortaya konulmuştur. Bununla birlikte TBO(ab) DMF türünün dikkate alındığı bütün koşullardaki Tip I hata oranlarına göre; GLK için elde edilen değerlerin üç grupta 0.045 ile 0.279, dört grupta .042 ile .251 arasında, Lordif için elde edilen değerlerin ise üç grupta 0.077 ile 0.163, dört grupta .082 ile .195 arasında değişiklik göstermesi nedeniyle diğer koşullar sabit tutulduğunda referans grubun örneklem büyüklüğü ve odak gruplara eklenen TBO(ab) DMF miktarı arttıkça GLK ve Lordif'e ilişkin Tip I hata oranlarının genel olarak yükseldiği tespit edilmiştir. Öte yandan GLR-LRT ve GLR-Wald teknikleri için bütün koşullarda hesaplanan Tip I hata oranlarının liberal kritere göre kabul edilebilir olduğu gözlemlenmiş ve bu iki tekniğin GLK ile Lordif'ten daha iyi performans sergilediği belirlenmiştir. Ayrıca Lordif için elde edilen değerlerin bütün koşullarda kabul edilebilir olmamasına rağmen GLK'nın çeşitli koşullar altındaki performansının hem referans grubun örneklem büyüklüğü hem de odak gruplara eklenen TBO(ab) DMF miktarına bağlı olarak farklılaştığı görülmüştür (bkz. EK 22). Simülasyon çalışmalarında hesaplanan yüksek Tip I hata oranları, tekniklerin performanslarının düşük olduğunu gösterdiğinden (Kim, 2003; Schauburger ve Tutz, 2016) odak gruplara eklenen TBO(ab) DMF miktarı ve grupların örneklem büyüklükleri birlikte arttıkça Tip I hata oranlarının yükselmesine bağlı olarak GLK ve Lordif tekniklerinin performanslarının azaldığı sonucuna ulaşılmıştır.

Alanyazında hem çoklu gruplar hem de TBO(ab) DMF türünün dikkate alındığı Oshima ve diğerlerinin (2015) çalışmasında, grup sayısı önemli olmaksızın örneklem büyüklüğü ile TBO(ab) DMF miktarı arttığında MG-NCDIF tekniği için hesaplanan Tip

I hata oranlarının nispeten yükseldiği belirlenmiştir. Benzer şekilde TBO(*ab*) DMF miktarı arttıkça MIMIC, GMH ve MG-NCDIF teknikleri için hesaplanan Tip I hata oranlarının yükselme eğiliminde oldukları tespit edilmiştir (Dell-Ross, 2021). Bu bulguları destekler nitelikte olan sınırlı sayıdaki başka simülasyon çalışmalarında da; iki grup ve ikili puanlanan maddeler için örneklem büyüklüğü arttıkça DFIT (Seybert ve Stark, 2012) tekniğine, örneklem büyüklüğü ile birlikte TBO(*ab*) DMF miktarı da arttığında ise MTKOO (Stark vd., 2006) tekniğine ilişkin hesaplanan Tip I hata oranlarının yükseldiği gözlemlenmiştir. İlgili araştırmalardan elde edilen sonuçlara göre TBO(*ab*) DMF türündeki maddelerde hem ikili (LK, DFIT, MTKOO) hem de çoklu gruplarda (MG-NCDIF) MTK'ya dayalı olan teknikler için hesaplanan Tip I hata oranlarının temel Tip I hata oranlarına göre yüksek olduğu anlaşılmıştır.

Yapılan bu araştırmada testin TBO(*ab*) DMF'li madde içerdiği tüm koşullarda GLK ve Lordif için hesaplanan Tip I hata oranlarının hem örneklem büyüklüğündeki hem de TBO(*ab*) DMF miktarındaki artışa bağlı olarak yükselmesi nedeniyle belirtilen tekniklere ilişkin performansların, MTK'ya dayalı tekniklerin (LK, DFIT, MTKOO, MG-NCDIF) incelendiği çalışmaların sonuçlarıyla benzer olduğu görülmüştür. Bununla birlikte grup sayısı önemli olmaksızın GLK ve Lordif tekniklerine ilişkin hesaplanan Tip I hata oranlarının ayırt edicilik parametresinden ziyade daha çok güçlük parametresindeki artışa dayalı olarak yükseldiği belirlenmiştir. Başka bir ifadeyle odak grupların güçlük ve ayırt edicilik parametrelerine 0.40 ile 0.70 olmak üzere iki farklı DMF miktarı eklenen bu araştırmada, ayırt edicilik parametresine kıyasla güçlük parametresine eklenen DMF miktarı 0.70 olduğunda MTK'ya dayalı olan GLK ve Lordif tekniklerine ilişkin Tip I hata oranlarının daha fazla yükseldiği ortaya konulmuştur (bkz. EK 22). Böylece elde edilen bu bulgulara göre GLK ve Lordif'in Tip I hata açısından performanslarının daha çok güçlük parametresindeki artıştan olumsuz yönde etkilendiği gözlemlenmiştir. Bu bilgiler doğrultusunda; grup sayısının üç veya dört, DMF türünün ise TBO(*ab*) olduğu koşullarda GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerine ilişkin performansların Tip I hata oranları açısından incelendiği ilk çalışma özelliği taşıyan bu araştırmada, GLR-LRT ve GLR-Wald için elde edilen değerlerin bütün koşullarda kabul edilebilir düzeyde olduğu, GLK ve Lordif için hesaplanan Tip I hata oranlarının ise genel olarak her iki madde parametresinden kaynaklı DMF miktarına bağlı olarak yükselmesine rağmen tekniklere ilişkin performansların güçlük parametresinden daha fazla etkilendiği sonucuna ulaşılmıştır.

4.1.15 Üç grupta TBO(*ab*) DMF türü için GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı [*a*(0.4, 0.7), *b*(0.4, 0.7)] koşullarındaki istatistiksel güç oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(*ab*) DMF türü için çoklu grup DMF belirleme tekniklerinin performansları istatistiksel güç açısından incelenmiştir. Üç grup ve TBO(*ab*) DMF türü için istatistiksel güç oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 21’de sunulmuştur.

Tablo 21

Üç Grupta TBO(ab) Türü DMF için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları)

	Kareler Toplamı	<i>sd</i>	<i>F</i>	<i>p</i>	η^2	η_p^2
Teknik (T)	386.010	3	2652.558	.000	.335	.881
Referans örneklem büyüklüğü (R)	476.921	1	9831.845	.000	.414	.901
Odak1 örneklem oranı (O ₁)	7.298	2	75.224	.000	.006	.122
Odak2 örneklem oranı (O ₂)	5.290	2	54.523	.000	.005	.092
Odak1 DMF miktarı (a ₁)	69.283	1	1428.279	.000	.060	.570
Odak2 DMF miktarı (a ₂)	70.348	1	1450.233	.000	.061	.574
Odak1 DMF miktarı (b ₁)	2.269	1	46.775	.000	.002	.042
Odak2 DMF miktarı (b ₂)	2.325	1	47.934	.000	.002	.043
T x R	11.909	3	81.837	.000	.010	.185
T x O ₁	0.243	6	0.836	.542	.000	.005
T x O ₂	0.230	6	0.790	.577	.000	.004
T x a ₁	8.868	3	60.936	.000	.008	.145
T x a ₂	9.043	3	62.138	.000	.008	.147
T x b ₁	12.792	3	87.902	.000	.011	.197
T x b ₂	11.406	3	78.379	.000	.010	.179
Hata	52.291	1078			.045	
Toplam	1151.001					

Tablo 21’deki üç grup ve TBO(*ab*) DMF türü için verilen istatistiksel güç sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(3,1078)} = 2652.558$, $p < .001$, $\eta^2 = .335$) ve referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,1078)} = 9831.845$, $p < .001$, $\eta^2 = .414$) manidar ve yüksek düzey bir etkiye sahip oldukları anlaşılmaktadır. Ayrıca birinci odak ($F_{(1,1078)} = 1428.279$, $p < .001$, $\eta^2 = .060$) ile ikinci odak ($F_{(1,1078)} = 1450.233$, $p < .001$, $\eta^2 = .061$) grubun madde ayırt edicilik parametrelerine eklenen DMF miktarı ana etkilerinin istatistiksel olarak manidar ve orta düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan birinci odak ($F_{(2,1078)} = 75.224$, $p < .001$, $\eta^2 = .006$) ile ikinci odak ($F_{(2,1078)} = 54.523$, $p < .001$, $\eta^2 = .005$) grubun örneklem büyüklüğü oranları

ve birinci odak ($F_{(1,1078)} = 46.775, p < .001, \eta^2 = .002$) ile ikinci odak ($F_{(1,1078)} = 47.934, p < .001, \eta^2 = .002$) grubun madde güçlük parametrelerine eklenen DMF miktarı ana etkilerinin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ($F_{(3,1078)} = 81.837, p < .001, \eta_p^2 = .185$), istatistiksel teknik x birinci odak ($F_{(3,1078)} = 87.902, p < .001, \eta_p^2 = .197$) ile istatistiksel teknik x ikinci odak grubun madde güçlük parametrelerine eklenen DMF miktarı ($F_{(3,1078)} = 78.379, p < .001, \eta_p^2 = .179$), istatistiksel teknik x birinci odak ($F_{(3,1078)} = 60.936, p < .001, \eta_p^2 = .145$) ile istatistiksel teknik x ikinci odak ($F_{(3,1078)} = 62.138, p < .001, \eta_p^2 = .147$) grubun madde ayırt edicilik parametrelerine eklenen DMF miktarlarına ilişkin ikili etkileşimlerinin istatistiksel olarak manidar ve yüksek düzey bir etkiye sahip oldukları görülmektedir. Böylece üç grup için TBO(ab) DMF türü dikkate alındığında, istatistiksel güç üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile birinci odak grubun madde güçlük parametresine eklenen DMF miktarına ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .197$). Öte yandan odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimlerinin ya manidar olmaması ya da hesaplanan etki büyüklüklerinin oldukça düşük olması nedeniyle, bu faktörlerin istatistiksel güç oranları üzerinde etkili olmadıkları ifade edilebilir.

4.1.16 Dört grupta TBO(ab) DMF türü için GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı [$a(0.4, 0.7)$, $b(0.4, 0.7)$] koşullarındaki istatistiksel güç oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(ab) DMF türü için çoklu grup DMF belirleme tekniklerinin performansları istatistiksel güç açısından incelenmiştir. Dört grup ve TBO(ab) DMF türü için istatistiksel güç oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 22’de sunulmuştur.

Tablo 22

Dört Grupta TBO(ab) DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	4916.993	3	37321.610	.000	.356	.891
Referans örneklem büyüklüğü (R)	6213.767	1	141493.680	.000	.450	.912
Odak1 örneklem oranı (O ₁)	32.896	2	374.542	.000	.002	.052
Odak2 örneklem oranı (O ₂)	34.784	2	396.028	.000	.003	.055
Odak3 örneklem oranı (O ₃)	29.741	2	338.619	.000	.002	.047
Odak1 DMF miktarı (a ₁)	377.280	1	8591.032	.000	.027	.386
Odak2 DMF miktarı (a ₂)	379.781	1	8647.991	.000	.027	.387
Odak3 DMF miktarı (a ₃)	386.228	1	8794.796	.000	.028	.391
Odak1 DMF miktarı (b ₁)	7.922	1	180.400	.000	.001	.013
Odak2 DMF miktarı (b ₂)	8.032	1	182.887	.000	.001	.013
Odak3 DMF miktarı (b ₃)	8.202	1	186.766	.000	.001	.013
T x R	241.852	3	1835.736	.000	.017	.287
T x O ₁	1.160	6	4.401	.000	.000	.002
T x O ₂	1.052	6	3.992	.000	.000	.002
T x O ₃	0.883	6	3.353	.003	.000	.001
T x a ₁	52.284	3	396.856	.000	.004	.080
T x a ₂	47.146	3	357.853	.000	.003	.073
T x a ₃	51.216	3	388.750	.000	.004	.078
T x b ₁	62.846	3	477.023	.000	.005	.095
T x b ₂	60.605	3	460.009	.000	.004	.092
T x b ₃	64.297	3	488.034	.000	.005	.097
Hata	601.335	13693			.044	
Toplam	13823.010					

Tablo 22'deki dört grup ve TBO(ab) DMF türü için verilen istatistiksel güç sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(3,13693)} = 37321.610$, $p < .001$, $\eta^2 = .356$) manidar ve yüksek düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,13693)} = 141493.680$, $p < .001$, $\eta^2 = .450$) manidar ve yüksek düzey bir etkiye sahip oldukları anlaşılmaktadır. Ayrıca birinci odak ($F_{(1,13693)} = 8591.032$, $p < .001$, $\eta^2 = .027$), ikinci odak ($F_{(1,13693)} = 8647.991$, $p < .001$, $\eta^2 = .027$) ve üçüncü odak ($F_{(1,13693)} = 8794.796$, $p < .001$, $\eta^2 = .028$) grubun madde ayırt edicilik parametrelerine eklenen DMF miktarları ana etkilerinin istatistiksel olarak manidar ve düşük düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan birinci odak ($F_{(2,13693)} = 374.542$, $p < .001$, $\eta^2 = .002$), ikinci odak ($F_{(2,13693)} = 396.028$, $p < .001$, $\eta^2 = .003$) ve üçüncü odak ($F_{(2,13693)} = 338.619$, $p < .001$, $\eta^2 = .002$) grubun örneklem büyüklüğü oranları ile birinci odak ($F_{(1,13693)} = 180.400$, $p < .001$, $\eta^2 = .000$), ikinci odak ($F_{(1,13693)} = 182.887$, $p < .001$, $\eta^2 = .000$) ve üçüncü odak ($F_{(1,13693)} = 186.766$, $p < .001$, $\eta^2 = .000$) grubun madde güçlük parametrelerine eklenen DMF miktarları ana etkilerinin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ($F_{(3,13693)} = 1835.736, p < .001, \eta_p^2 = .287$) ikili etkileşiminin istatistiksel olarak manidar ve yüksek düzey bir etkiye sahip olduğu görülmektedir. İstatistiksel teknik x birinci odak ($F_{(3,13693)} = 477.023, p < .001, \eta_p^2 = .095$), istatistiksel teknik x ikinci odak ($F_{(3,13693)} = 460.009, p < .001, \eta_p^2 = .092$) ve istatistiksel teknik x üçüncü odak ($F_{(3,13693)} = 488.034, p < .001, \eta_p^2 = .097$) grubun madde güçlük parametrelerine eklenen DMF miktarlarına ve istatistiksel teknik x birinci odak ($F_{(3,13693)} = 396.856, p < .001, \eta_p^2 = .080$), istatistiksel teknik x ikinci odak ($F_{(3,13693)} = 357.853, p < .001, \eta_p^2 = .073$) ve istatistiksel teknik x üçüncü odak ($F_{(3,13693)} = 388.750, p < .001, \eta_p^2 = .078$) grubun madde ayırt edicilik parametrelerine eklenen DMF miktarlarına ilişkin ikili etkileşimlerinin, istatistiksel olarak manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Diğer taraftan istatistiksel teknik x birinci odak ($F_{(6,13693)} = 4.401, p < .001, \eta_p^2 = .002$), istatistiksel teknik x ikinci odak ($F_{(6,13693)} = 3.992, p < .001, \eta_p^2 = .002$) ve istatistiksel teknik x üçüncü odak ($F_{(6,13693)} = 3.353, p < .001, \eta_p^2 = .001$) grubun örneklem büyüklüğü oranlarına ilişkin ikili etkileşimlerinin, istatistiksel olarak manidar ancak çok düşük bir etkiye sahip oldukları gözlemlenmektedir. Böylece dört grup için TBO(ab) DMF türü dikkate alındığında, istatistiksel güç üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilir ($\eta_p^2 = .287$). Bununla birlikte odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimleri için hesaplanan etki büyüklüklerinin oldukça düşük olması nedeniyle, bu faktörlerin istatistiksel güç oranları üzerinde etkili olmadıkları ifade edilebilir.

Araştırma kapsamında hem üç hem de dört grup için EK 23'te verilen bulgulara göre; testteki maddelerin TBO(ab) DMF içerdiği durumlarda çoklu grup teknikleri için hesaplanan istatistiksel güç oranlarının .297 ile 1.000 aralığında değiştiği belirlenmiştir. Hesaplanan istatistiksel güç oranları alanyazında önerilen (Lei ve Li, 2013; Scott vd., 2009) kriterlere göre değerlendirildiğinde; grup sayısı önemli olmaksızın bütün koşullarda GLK ve Lordif tekniklerinin TBO(ab) DMF içeren maddeleri doğru belirleme oranlarının en az kabul edilebilir (%60'ın üzerinde) düzeyde olduğu sonucuna ulaşılmıştır. Öte yandan GLR-LRT ve GLR-Wald teknikleri için, grup sayısı önemli olmaksızın $N_R = 1800$ olduğu bütün koşulların yanı sıra, $N_R = 600$ ve en az iki odak grubun ayırt edicilik parametrelerine eklenen DMF miktarlarının 0.70 olduğu koşullarda

da istatistiksel güç oranlarının kabul edilebilir düzeyin üzerinde olduğu tespit edilmiştir. Böylece 100 tekrarın yapıldığı bu araştırmada; hem üç hem de dört grup için gerçekte TBO(ab) DMF’li olan maddelerin GLK ve Lordif teknikleri tarafından %60 ve üzeri olasılıkla DMF’li olarak işaretlendiği anlaşılmış, GLR-LRT ve GLR-Wald içinse $N_R = 1800$ olduğu bütün koşullarda veya $N_R = 600$ olsa dahi odak grupların ayırt edicilik parametrelerine eklenen DMF miktarlarının 0.70 olması durumunda gerçekte TBO(ab) DMF’li olan maddelerin ilgili teknikler tarafından %60 ve üzeri olasılıkla DMF’li olarak belirlendiği ortaya konulmuştur. Simülasyon çalışmalarında .60’ın üzerindeki istatistiksel güç oranlarının, tekniklere ilişkin performansların kabul edilebilir düzeyde olduğunu göstermesi nedeniyle (Finch, 2016; Lei ve Li, 2013; Scott vd., 2009; Svetina ve Rutkowski, 2014) yapılan bu araştırma kapsamında $N_R = 1800$ olduğunda tekniklerin kabul edilebilir ve üzerinde performans sergilediği, $N_R = 600$ olduğunda ise güçlük parametresi önemli olmaksızın odak grupların ayırt edicilik parametrelerine eklenen DMF miktarlarının 0.70 olması durumunda belirtilen teknikler için hesaplanan istatistiksel güç oranlarının kabul edilebilir düzeyde olduğu sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında hem üç hem de dört grup için çoklu grup tekniklerine ilişkin hesaplanan istatistiksel güç oranlarının referans grubun örneklem büyüklüğü ve odak grupların güçlük veya ayırt edicilik parametrelerine eklenen DMF miktarlarına bağlı olarak farklılaşabildiği ortaya konulmuştur. Bu bağlamda grup sayısı önemli olmaksızın $N_R = 600$ olduğunda, GLR-LRT, GLR-Wald ve Lordif tekniklerinin daha çok odak grupların ayırt edicilik parametrelerine, GLK tekniğinin ise daha çok güçlük parametrelerine eklenen DMF miktarlarından etkilendiği belirlenmiştir. Öte yandan TBO(ab) DMF miktarı düşük olsa dahi $N_R = 1800$ olduğunda; hem üç grupta [GLK (.986), GLR-LRT (.784), GLR-Wald (.780), Lordif (.947)] hem de dört grupta [GLK (.987), GLR-LRT (.794), GLR-Wald (.793), Lordif (.959)] çoklu grup teknikleri için hesaplanan istatistiksel güç oranlarına göre GLK ve Lordif’in yüksek, GLR-LRT ve GLR-Wald’ın ise diğer tekniklerden daha düşük performans gösterdikleri tespit edilmiştir (bkz. EK 23). Bu bulguya dayalı olarak, grup sayısı ve TBO(ab) DMF miktarı önemli olmaksızın büyük örneklemelerin dikkate alındığı geniş ölçekli gerçek test uygulamalarında, gerçekte DMF’li olan maddelerin çoklu grup DMF belirleme teknikleri tarafından genel olarak doğru tespit edilebildiği sonucuna ulaşılmıştır.

Araştırma kapsamında grup sayısı önemli olmaksızın DMF türü TBO(ab) ve $N_R = 1800$ olduğu durumda GLK ve Lordif için hesaplanan istatistiksel güç oranlarının birbirlerine ve 1.00’a oldukça yakın olduğu belirlenirken, $N_R = 600$ olan koşullarda genel

olarak GLK'nın en yüksek yüksek, GLR-LRT ve GLR-Wald tekniklerinin ise diğer iki teknikten biraz daha düşük ve birbirlerine benzer (Finch, 2016; Gao, 2019; Magis vd., 2011) performans gösterdikleri tespit edilmiştir (bkz. EK 23). Özellikle hem üç hem de dört grup için $N_R = 600$ ve yalnızca odak grupların ayırt edicilik parametrelerine eklenen DMF miktarlarının 0.40 olduğu durumda, istatistiksel güç oranlarının GLR-LRT ve GLR-Wald teknikleri için yetersiz olduğu ortaya konulmuştur. Öte yandan Lordif için $N_R = 600$ olduğunda yalnızca odak grupların ayırt edicilik parametrelerine eklenen DMF miktarlarının 0.40 olduğu durumların haricindeki bütün koşullarda elde edilen değerlerin yüksek olduğu gözlemlenmiştir. Son olarak GLK tekniğine ilişkin performansın $N_R = 600$ ve hem güçlük hem de ayırt edicilik parametrelerinden kaynaklı DMF miktarlarının 0.40 olduğu durumda kabul edilebilir, diğer bütün koşullarda ise tekniğin performansının yüksek olduğu ortaya konulmuştur. Böylece araştırma kapsamında elde edilen bulgulara dayalı olarak; grup sayısının üç veya dört, DMF türünün ise TBO(ab) olduğu durumda GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerine ilişkin istatistiksel güç oranlarının hem örneklem büyüklüğü hem de farklı madde parametrelerinden kaynaklı DMF miktarları ile ilişkili olduğu sonucuna ulaşılmıştır.

Alanyazındaki çoklu grup çalışmalarında GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin TBO(ab) DMF'li maddelerdeki performanslarının incelendiği herhangi bir araştırma bulunmamasına rağmen, Dell-Ross'un (2021) çalışmasında grup sayısı önemli olmaksızın genel olarak DMF miktarı arttıkça GMH, MIMIC ve MG-NCDIF teknikleri için hesaplanan istatistiksel güç oranlarının da yükselme eğiliminde olduklarını belirlenmiştir. Oshima ve diğerleri (2015) tarafından yapılan grup sayısının üç veya beş DMF türünün ise TBO(ab) olduğu çalışmada, genel olarak örneklem büyüklüğü ve DMF miktarı arttıkça MG-NCDIF tekniği için hesaplanan istatistiksel güç oranlarının da yükseldiği tespit edilmiştir. Benzer şekilde ikili grup ve ikili puanlanan maddelerde TBO(ab) DMF türünün dikkate alındığı simülasyon çalışmalarında, örneklem büyüklüğü ve DMF miktarı arttığında LK (Seybert vd., 2014) ile MTKOO (Stark vd., 2006) teknikleri için hesaplanan istatistiksel güç oranlarının yükselme eğiliminde olduğu belirtilmiştir. Böylece hem ikili grup hem de çoklu grup DMF belirleme tekniklerinin, testin TBO(ab) DMF'li madde içerdiği koşullardaki istatistiksel güç oranlarının genel olarak örneklem büyüklüğü ve DMF miktarındaki artıştan olumlu yönde etkilendiği ve elde edilen bulguların alanyazındaki TBO(ab) DMF türünü dikkate alan araştırmalar (Oshima vd., 2015; Seybert vd., 2014; Stark vd., 2006) tarafından desteklendiği sonucuna ulaşılmıştır.

4.1.17 Üç grupta TBO(ab) DMF türü için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı $[a(0.4, 0.7), b(0.4, 0.7)]$ koşullarındaki Tip III hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(ab) DMF türü için çoklu grup DMF belirleme tekniklerinin performansları Tip III hata açısından incelenmiştir. Üç grup ve TBO(ab) DMF türü için Tip III hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 23'te sunulmuştur.

Tablo 23

<i>Üç Grupta TBO(ab) DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)</i>						
	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	46.108	4	230.038	.000	.032	.404
Referans örneklem büyüklüğü (R)	175.951	1	3511.333	.000	.122	.721
Odak1 örneklem oranı (O ₁)	3.105	2	30.983	.000	.002	.044
Odak2 örneklem oranı (O ₂)	3.562	2	35.545	.000	.002	.050
Odak1 DMF miktarı (a ₁)	2.354	1	46.969	.000	.002	.033
Odak2 DMF miktarı (a ₂)	2.083	1	41.578	.000	.001	.030
Odak1 DMF miktarı (b ₁)	368.427	1	7352.454	.000	.256	.844
Odak2 DMF miktarı (b ₂)	359.524	1	7174.788	.000	.250	.841
T x R	6.527	4	32.563	.000	.005	.088
T x O ₁	0.104	8	0.260	.978	.000	.002
T x O ₂	0.284	8	0.709	.684	.000	.004
T x a ₁	0.365	4	1.820	.122	.000	.005
T x a ₂	0.343	4	1.711	.145	.000	.005
T x b ₁	18.506	4	92.327	.000	.013	.214
T x b ₂	16.840	4	84.017	.000	.012	.199
Hata	67.948	1356			.047	
Toplam	1438.998					

Tablo 23'te üç grup ve TBO(ab) DMF için verilen Tip III hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(4,1356)} = 230.038, p < .001, \eta^2 = .032$) manidar ve düşük düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,1356)} = 3511.333, p < .001, \eta^2 = .122$) manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Birinci odak ($F_{(1,1356)} = 7352.454, p < .001, \eta^2 = .256$) ile ikinci odak ($F_{(1,1356)} = 7174.788, p < .001, \eta^2 = .250$) grubun madde güçlük parametrelerine eklenen DMF miktarı ana etkilerinin istatistiksel olarak manidar ve yüksek düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan birinci odak ($F_{(2,1356)} = 30.983, p < .001, \eta^2 = .002$) ile ikinci odak ($F_{(2,1356)} = 35.545, p < .001, \eta^2 = .002$) grubun örneklem büyüklüğü oranları ve birinci odak ($F_{(1,1356)} = 46.969, p < .001, \eta^2 = .002$) ile ikinci odak ($F_{(1,1356)} = 41.578,$

$p < .001$, $\eta^2 = .001$) grubun madde ayırt edicilik parametrelerine eklenen DMF miktarı ana etkilerinin istatistiksel olarak manidar, ancak çok düşük etkiye sahip oldukları gözlemlenmektedir.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ($F_{(4,1356)} = 32.563$, $p < .001$, $\eta_p^2 = .088$) ikili etkileşiminin istatistiksel olarak manidar ve orta düzey bir etkiye sahip olduğu görülmektedir. Diğer taraftan istatistiksel teknik x birinci odak ($F_{(4,1356)} = 92.327$, $p < .001$, $\eta_p^2 = .214$) ile istatistiksel teknik x ikinci odak ($F_{(4,1356)} = 84.017$, $p < .001$, $\eta_p^2 = .199$) grubun madde güçlük parametrelerine eklenen DMF miktarlarına ilişkin ikili etkileşimlerinin istatistiksel olarak manidar ve yüksek düzey bir etkiye sahip oldukları anlaşılmaktadır. Böylece üç grup için TBO(ab) DMF türü dikkate alındığında, Tip III hata üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile birinci odak grubun madde güçlük parametresine eklenen DMF miktarına ilişkin ikili etkileşimin olduğu belirtilebilir ($\eta_p^2 = .214$). Öte yandan odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimlerinin ya manidar olmaması ya da hesaplanan etki büyüklüklerinin oldukça düşük olması nedeniyle, bu faktörlerin Tip III hata oranları üzerinde etkili olmadıkları ifade edilebilir.

4.1.18 Dört grupta TBO(ab) DMF türü için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin referans grubun örneklem büyüklüğü (600, 1800), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/1, 1:1/2 ve 1:2/3) ve DMF miktarı [$a(0.4, 0.7)$, $b(0.4, 0.7)$] koşullarındaki Tip III hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda TBO(ab) DMF türü için çoklu grup DMF belirleme tekniklerinin performansları Tip III hata açısından incelenmiştir. Dört grup ve TBO(ab) DMF türü için Tip III hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 24'te sunulmuştur.

Tablo 24

Dört Grupta TBO(ab) DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	555.523	4	1237.436	.000	.032	.224
Referans örneklem büyüklüğü (R)	2039.613	1	18173.082	.000	.118	.515
Odak1 örneklem oranı (O ₁)	26.258	2	116.978	.000	.002	.013

(Devam ediyor)

Tablo 24 (Devam)

Dört Grupta TBO(ab) DMF Türü için Faktöriyel ANOVA Sonucu (Tip III Hata Oranları)

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Odak2 örneklem oranı (O ₂)	26.372	2	117.488	.000	.002	.014
Odak3 örneklem oranı (O ₃)	25.971	2	115.701	.000	.002	.013
Odak1 DMF miktarı (a ₁)	6.894	1	61.425	.000	.000	.004
Odak2 DMF miktarı (a ₂)	6.283	1	55.981	.000	.000	.003
Odak3 DMF miktarı (a ₃)	7.153	1	63.736	.000	.000	.004
Odak1 DMF miktarı (b ₁)	2217.796	1	19760.705	.000	.128	.536
Odak2 DMF miktarı (b ₂)	2212.624	1	19714.620	.000	.128	.535
Odak3 DMF miktarı (b ₃)	2201.178	1	19612.643	.000	.127	.534
T x R	198.779	4	442.784	.000	.012	.094
T x O ₁	2.461	8	2.741	.005	.000	.001
T x O ₂	3.111	8	3.465	.001	.000	.002
T x O ₃	2.524	8	2.811	.004	.000	.001
T x a ₁	1.224	4	2.727	.028	.000	.001
T x a ₂	1.072	4	2.388	.049	.000	.001
T x a ₃	1.417	4	3.157	.013	.000	.001
T x b ₁	133.104	4	296.492	.000	.008	.065
T x b ₂	132.133	4	294.328	.000	.008	.064
T x b ₃	129.295	4	288.008	.000	.007	.063
Hata	1923.106	17135			.111	
Toplam	17279.000					

Tablo 24'teki dört grup ve TBO(ab) DMF türü için verilen Tip III hata sonuçlarına göre; çoklu grup teknikleri ana etkisinin ($F_{(4,17135)} = 1237.436$, $p < .001$, $\eta^2 = .032$) manidar ve düşük düzey, referans grubun örneklem büyüklüğü ana etkisinin ($F_{(1,17135)} = 18173.082$, $p < .001$, $\eta^2 = .118$) ise manidar ve orta düzey bir etkiye sahip oldukları anlaşılmaktadır. Ayrıca birinci odak ($F_{(1,17135)} = 19760.705$, $p < .001$, $\eta^2 = .128$), ikinci odak ($F_{(1,17135)} = 19714.620$, $p < .001$, $\eta^2 = .128$) ve üçüncü odak ($F_{(1,17135)} = 19612.643$, $p < .001$, $\eta^2 = .127$) grubun madde güçlük parametrelerine eklenen DMF miktarları ana etkilerinin istatistiksel olarak manidar ve orta düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan birinci odak ($F_{(2,17135)} = 116.978$, $p < .001$, $\eta^2 = .002$), ikinci odak ($F_{(2,17135)} = 117.488$, $p < .001$, $\eta^2 = .002$) ve üçüncü odak ($F_{(2,17135)} = 115.701$, $p < .001$, $\eta^2 = .002$) grubun örneklem büyüklüğü oranları ve birinci odak ($F_{(1,17135)} = 61.425$, $p < .001$, $\eta^2 = .000$), ikinci odak ($F_{(1,17135)} = 55.981$, $p < .001$, $\eta^2 = .000$) ve üçüncü odak ($F_{(1,17135)} = 63.736$, $p < .001$, $\eta^2 = .000$) grubun madde ayırt edicilik parametrelerine eklenen DMF miktarları ana etkilerinin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ($F_{(4,17135)} = 442.784$, $p < .001$, $\eta_p^2 = .094$), istatistiksel teknik x birinci odak ($F_{(4,17135)} = 296.492$, $p < .001$, $\eta_p^2 = .065$), istatistiksel

teknik x ikinci odak ($F_{(4,17135)} = 294.328, p < .001, \eta_p^2 = .064$) ve istatistiksel teknik x üçüncü odak ($F_{(4,17135)} = 288.008, p < .001, \eta_p^2 = .063$) grubun madde güçlük parametrelerine eklenen DMF miktarlarına ilişkin ikili etkileşimlerinin istatistiksel olarak manidar ve orta düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan istatistiksel teknik x birinci odak ($F_{(4,17135)} = 2.727, p < .001, \eta_p^2 = .001$), istatistiksel teknik x ikinci odak ($F_{(4,17135)} = 2.388, p < .001, \eta_p^2 = .001$) ve istatistiksel teknik x üçüncü odak ($F_{(4,17135)} = 3.157, p < .001, \eta_p^2 = .001$) grubun madde ayırt edicilik parametrelerine eklenen DMF miktarlarına ve istatistiksel teknik x birinci odak ($F_{(8,17135)} = 2.741, p < .001, \eta_p^2 = .001$), istatistiksel teknik x ikinci odak ($F_{(8,17135)} = 3.465, p < .001, \eta_p^2 = .002$) ve istatistiksel teknik x üçüncü odak ($F_{(8,17135)} = 2.811, p < .001, \eta_p^2 = .001$) grubun örneklem büyüklüğü oranlarına ilişkin ikili etkileşimlerinin istatistiksel olarak manidar, ancak çok düşük bir etkiye sahip oldukları sonucuna ulaşılmaktadır. Böylece dört grup için TBO(ab) DMF türü dikkate alındığında, Tip III hata üzerindeki en önemli etkileşim etkisinin istatistiksel teknik ile referans grubun örneklem büyüklüğüne ilişkin ikili etkileşim olduğu belirtilebilmektedir ($\eta_p^2 = .094$). Öte yandan odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerinin hem ana etkileri hem de bu faktörlerin istatistiksel teknik ile etkileşimleri için hesaplanan etki büyüklüklerinin düşük olması nedeniyle, bu faktörlerin Tip III hata oranları üzerinde etkili olmadıkları ifade edilebilmektedir.

Araştırma kapsamında hem üç hem de dört grup için EK 24'te verilen bulgulara göre; TBO(ab) DMF içeren maddelerin TB DMF türünü belirlemek amacıyla geliştirilen çoklu grup teknikleri ile analiz edilmesi sonucunda hesaplanan Tip III hata oranlarının .619 ile 1.000 aralığında değiştiği sonucuna ulaşılmıştır. Hesaplanan Tip III hata oranları açısından GLK, GLR-LRT, GLR-Wald, GMH ve Lordif için elde edilen değerlerin Bradley (1978) tarafından önerilen liberal kriterden oldukça yüksek olduğu tespit edilmiştir. Böylece 100 tekrarın yapıldığı bu çalışmada, grup sayısı önemli olmaksızın TBO(ab) DMF içeren maddelerin TB DMF'yi belirlemek amacıyla geliştirilen teknikler tarafından %61.9 ile %100 arasındaki oranlarda DMF'li olarak belirlenebildiği gözlemlenmiştir. Simülasyon çalışmalarında hesaplanan yüksek Tip III hata oranları; gerçekte DMF'li olan maddeler için tespit edilen DMF türünün analizin gerçekleştirildiği istatistiksel teknik tarafından hatalı tanımlandığını gösterdiğinden (Lopez Rivas, 2012), bu çalışmadaki TBO(ab) DMF'li maddelerin DMF türünün genel olarak çoklu grup teknikleri ile doğru belirlenemediği ortaya konulmuştur.

Yapılan bu simülasyon çalışmasında üç grupta belirlenen dört maddenin madde parametrelerinin değiştiği bütün koşullardaki Tip III hata oranlarının; GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için .619 ile 1.000 arasında olduğu görülmüştür. Lordif için hesaplanan Tip III hata oranlarının ise .792 ile 1.000 arasında ve özellikle odak grupların güçlük parametrelerine eklenen DMF miktarlarının 0.40 olduğu bütün koşullarda diğer tekniklere kıyasla daha yüksek hesaplandığı tespit edilmiştir. Üç gruba benzer şekilde dört grupta da hesaplanan Tip III hata oranlarının; GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için .625 ile 1.000 arasında olduğu gözlemlenmiştir. Ayrıca Lordif için hesaplanan Tip III hata oranlarının .803 ile 1.000 arasında ve özellikle odak grupların güçlük parametrelerine eklenen DMF miktarlarının 0.40 olduğu bütün koşullarda diğer tekniklere göre daha yüksek hesaplandığı belirlenmiştir (bkz. EK 24). Böylece ikili puanlanan maddeler ve çoklu gruplarda istatistiksel tekniklerin performanslarının Tip III hata açısından incelendiği ilk araştırma niteliğinde olan bu çalışma kapsamında; grup sayısının üç veya dört, DMF türünün ise TBO(ab) olduğu durum için hesaplanan yüksek Tip III hata oranları, TB DMF'li maddeleri belirlemek amacıyla geliştirilen GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin gerçekte TBO(ab) DMF içeren maddeleri de yüksek oranda DMF'li olarak tespit edebildiğini göstermiştir. Yapılan bu araştırmaya benzer şekilde iki grup için Tip III hatanın hesaplandığı bir simülasyon çalışmasında da (Lopez Rivas 2012) TB DMF'yi tespit etmek amacıyla geliştirilen tekniklerin testteki TBO(ab) DMF'li maddeleri yanlışlıkla DMF'li olarak belirleme oranlarının yüksek olduğu belirlenmiş ve böylelikle alanyazındaki sınırlı sayıdaki çalışmaların bulgularının desteklendiği sonucuna ulaşılmıştır.

Araştırma kapsamında hesaplanan Tip III hata oranlarına göre; grup sayısı önemli olmaksızın GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için elde edilen değerlerin bütün koşullar altında genellikle birbirlerine yakın olduğu tespit edilmiştir. Ayırt edicilik parametrelerine eklenen DMF miktarları önemli olmaksızın özellikle odak grupların güçlük parametrelerine eklenen DMF miktarlarının 0.40 olduğu koşullarda Lordif için hesaplanan Tip III hata oranlarından oldukça düşük olduğu belirlenmiştir. MTK'ya dayalı olan GLK ve Lordif tekniklerine ilişkin Tip III hata oranları dikkate alındığında; GLK tekniğinin Lordif'ten ziyade, KTK'ya dayalı oldukları bilinen GLR-LRT, GLR-Wald ve GMH teknikleri ile benzer performans gösterdiği anlaşılmıştır. Elde edilen bu bulgu kuramsal olarak hem KTK hem de MTK'ya dayalı tekniklerin özelliklerini gösteren Lordif ile ilgili yeni çalışmaların yapılmasına ihtiyaç olduğunu ortaya koymuştur. Öte

yandan yapılan bu çalışmada TB DMF türünü belirlemek amacıyla geliştirilen GLR-LRT, GLR-Wald ve GMH tekniklerinin TBO(*ab*) DMF içeren maddelerdeki Tip III hata oranlarına ilişkin performanslarının bütün koşullarda birbirlerine son derece yakın olduğu tespit edilmiş ve bu teknikler için elde edilen performansların alanyazındaki açıklamalarla (Finch, 2016; Gao, 2019; Magis vd., 2011) desteklendiği sonucuna ulaşılmıştır.

Grup sayısının üç veya dört, DMF türünün ise TBO(*ab*) olduğu bu simülasyon çalışmasında ayırt edicilik parametrelerine eklenen DMF miktarları önemli olmaksızın, odak grupların güçlük parametrelerine eklenen DMF miktarlarının 0.40 olduğu koşullarda, diğer koşullar sabit kalmak şartıyla referans grubun örneklem büyüklüğü arttıkça, TB DMF’li maddeleri belirlemek amacıyla geliştirilen GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının ciddi miktarda yükseldiği belirlenmiştir. Bununla birlikte odak grupların ayırt edicilik parametrelerine eklenen DMF miktarları arttıkça GLK, GLR-LRT, GLR-Wald ve GMH tekniklerinin Tip III hata oranlarının bir miktar azaldığı, Lordif için hesaplanan Tip III hata oranlarının ise ciddi bir şekilde farklılaşmadığı tespit edilmiştir (bkz. EK 24). Özellikle grup sayısı ve odak grupların madde parametrelerine eklenen DMF miktarları önemli olmaksızın $N_R = 1800$ iken Lordif için hesaplanan Tip III hata oranlarının 1.00’a oldukça yakın olduğu gözlemlenmiştir. Bununla birlikte odak grupların yalnızca güçlük parametrelerine eklenen DMF miktarlarının 0.40 olduğu durum haricindeki bütün koşullarda GLK, GLR-LRT, GLR-Wald, GMH ve Lordif için hesaplanan Tip III hata oranlarının çok yüksek olması nedeniyle güçlük parametrelerine ilişkin DMF miktarları arttıkça TB DMF’yi tespit etmek amacıyla geliştirilen beş tekniğin, testte yer alan TBO(*ab*) DMF’li maddeleri de %90 ve üzeri olasılıkla DMF’li olarak belirleyebildiği ortaya konulmuştur. Bu bilgiler doğrultusunda TB DMF’yi belirlemek amacıyla geliştirilen GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri ile TBO(*ab*) DMF’li maddelerin de DMF’li olarak işaretlenebildiği anlaşıldığından, büyük örneklemelerin dikkate alındığı geniş ölçekli gerçek test uygulamalarında DMF türünün çoklu grup teknikleri ile güvenilir bir şekilde belirlenemeyebileceği ortaya konulmuştur. Böylelikle belirtilen teknikler tarafından DMF’li olarak işaretlenen maddelerin DMF türünün doğru tespit edilebilmesi amacıyla MKE’lere başvurulmasının gerekli olduğu sonucuna ulaşılmıştır.

Araştırma kapsamında grup sayısının üç ve dört olduğu durum için TB, TBO(*a*) ve TBO(*ab*) DMF türlerinde Tip I hata, istatistiksel güç ve Tip III hata oranları açısından gerçekleştirilen bütün Faktöriyel ANOVA sonuçlarına göre hem odak grupların örneklem büyüklüklerinin referans gruba oranı (Odak₁-Odak₂-Odak₃) faktörlerine ilişkin ana etkiler

hem de çoklu grup DMF belirleme teknikleri ile odak grupların örneklem büyüklüklerinin referans gruba oranına ilişkin etkileşim etkileri için hesaplanan etki büyüklüklerinin çok düşük olduğu sonucuna ulaşılmıştır. Çalışmanın ikinci araştırma sorusunda belirtildiği üzere üç ve dört grup için üretilen veri setleri birleştirilerek grup sayısının da değişimlenmesi amaçlandığından, yapılan analizlerde odak grupların örneklem büyüklüklerinin referans gruba oranı faktörlerine ilişkin etki büyüklüklerinin çok düşük olması nedeniyle belirtilen faktörün ikinci araştırma sorusu için gerçekleştirilen analizlere dahil edilmemesine karar verilmiştir. Ayrıca çalışmada grup sayısına bağlı olarak odak gruplara eklenen toplam DMF miktarları farklılaştırıldığından ikinci araştırma sorusunu yanıtlamak amacıyla “Düşük DMF”, “Orta Düzey DMF” ve “Yüksek DMF” olmak üzere üç farklı DMF miktarı yeniden tanımlanmıştır. Bu çerçevede testteki hiçbir maddenin DMF içermeme durumu (DMF Yok) ile birlikte DMF miktarı için dört farklı koşul dikkate alınmıştır.

4.2. Grup sayısı (3, 4), referans grubun örneklem büyüklüğü (600, 1800), DMF miktarı (DMF Yok, Düşük DMF, Orta Düzey DMF, Yüksek DMF) ve DMF türü [DMF Yok, TB, TBO(a), TBO(ab)] koşullarında, GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin;

- a) Tip I hata oranları,
- b) İstatistiksel güç oranları,
- c) Tip III hata oranları nasıl değişmektedir?

Çalışmanın ikinci araştırma sorusu kapsamında GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin farklı koşullardaki performanslarını belirlemek amacıyla Teknik, Referans Grubun Örneklem Büyüklüğü, Grup Sayısı, DMF Türü ve DMF Miktarı olmak üzere beş faktör dikkate alınarak Faktöriyel ANOVA yapılmıştır. Yapılan Faktöriyel ANOVA’lar sonucunda belirtilen beş faktör için ana etkiler ile bu faktörlerin ikili, üçlü, dördü ve beşli etkileşimleri hesaplanmasına rağmen, tablolarda yalnızca çoklu grup teknikleri ile diğer faktörlerin etkileşimlerine yer verilmiş ve elde edilen sonuçlar arasından hem istatistiksel olarak manidar hem de etki büyüklüğü 0.010’a eşit veya bu değerden yüksek olan etkileşim etkileri yorumlanmıştır. İkinci araştırma sorusunda yanıt aranan alt araştırma soruları için Faktöriyel ANOVA tabloları bu bölümde, betimsel istatistiklere ilişkin bazı grafikler ise EK bölümünde sunulmuştur.

4.2.1. Grup sayısı (3, 4), referans grubun örneklem büyüklüğü (600, 1800), DMF miktarı (DMF Yok, Düşük DMF, Orta Düzey DMF, Yüksek DMF) ve DMF türü [DMF Yok, TB, TBO(a), TBO(ab)] koşullarında, GLK, GLR-LRT, GLR-Wald, GMH

ve Lordif için Tip I hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda çoklu grup DMF belirleme tekniklerinin performansları Tip I hata açısından incelenmiştir. Tip I hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 25’te sunulmuştur.

Tablo 25

Tip I Hata Oranları İçin Faktöriyel ANOVA Sonucu

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	14.466	4	23669.813	.000	.134	.824
Referans Örneklem Büyüklüğü (R)	13.313	1	87135.880	.000	.123	.811
Grup Sayısı (G)	0.056	1	364.709	.000	.001	.018
DMF Türü (K)	13.959	3	30453.879	.000	.129	.818
DMF Miktarı (M)	3.284	2	10747.932	.000	.030	.515
T x R	5.240	4	8574.936	.000	.048	.629
T x G	0.154	4	252.271	.000	.001	.047
T x K	22.645	10	14821.667	.000	.210	.880
T x M	1.538	8	1258.263	.000	.014	.332
T x R x G	0.042	4	69.108	.000	.000	.013
T x R x K	5.769	10	3775.757	.000	.053	.651
T x R x M	0.691	8	565.064	.000	.006	.182
T x G x K	0.039	10	25.819	.000	.000	.013
T x G x M	0.017	8	14.030	.000	.000	.006
T x K x M	1.556	12	848.643	.000	.014	.334
T x R x G x K	0.012	10	7.576	.000	.000	.004
T x R x G x M	0.006	8	5.128	.000	.000	.002
T x R x K x M	0.428	12	233.388	.000	.004	.121
T x G x K x M	0.008	12	4.451	.000	.000	.003
T x R x G x K x M	0.003	12	1.461	.131	.000	.001
Hata	3.097	20272			.029	
Toplam	107.955					

Tablo 25’te verilen Tip I hata sonuçlarına göre; çoklu grup teknikleri ($F_{(4,20272)} = 23669.813$, $p < .001$, $\eta^2 = .134$), referans grubun örneklem büyüklüğü ($F_{(1,20272)} = 87135.880$, $p < .001$, $\eta^2 = .123$) ve DMF türü ($F_{(3,20272)} = 30453.879$, $p < .001$, $\eta^2 = .129$) ana etkilerinin manidar ve orta düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan DMF miktarı ($F_{(2,20272)} = 10747.932$, $p < .001$, $\eta^2 = .030$) ana etkisinin manidar ve düşük düzey, grup sayısı ($F_{(1,20272)} = 364.709$, $p < .001$, $\eta^2 = .001$) ana etkisinin ise manidar, ancak çok düşük düzey bir etkiye sahip oldukları anlaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ($F_{(4,20272)} = 8574.936$, $p < .001$, $\eta_p^2 = .629$), istatistiksel teknik x DMF türü ($F_{(10,20272)} = 14821.667$, $p < .001$, $\eta_p^2 = .880$) ve istatistiksel teknik x DMF miktarı ($F_{(8,20272)} = 1258.263$, $p < .001$, $\eta_p^2 = .332$) ikili etkileşimlerinin

manidar ve yüksek düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan istatistiksel teknik x grup sayısı ($F_{(4,20272)} = 252.271, p < .001, \eta_p^2 = .047$) ikili etkileşiminin manidar, ancak düşük düzey bir etkiye sahip olduğu gözlemlenmektedir.

Faktörler arasındaki üçlü etkileşimlere göre; istatistiksel teknik x referans grubun örneklem büyüklüğü x DMF türü ($F_{(10,20272)} = 3775.757, p < .001, \eta_p^2 = .651$), istatistiksel teknik x referans grubun örneklem büyüklüğü x DMF miktarı ($F_{(8,20272)} = 565.064, p < .001, \eta_p^2 = .182$) ve istatistiksel teknik x DMF türü x DMF miktarı ($F_{(12,20272)} = 848.643, p < .001, \eta_p^2 = .334$) etkileşimlerinin manidar ve yüksek düzey bir etkiye sahip oldukları anlaşılmaktadır. Öte yandan istatistiksel teknik x referans grubun örneklem büyüklüğü x grup sayısı ($F_{(4,20272)} = 69.108, p < .001, \eta_p^2 = .013$) ve istatistiksel teknik x grup sayısı x DMF türü ($F_{(10,20272)} = 25.819, p < .001, \eta_p^2 = .013$) etkileşimlerinin manidar ve düşük düzey, istatistiksel teknik x grup sayısı x DMF miktarı ($F_{(8,20272)} = 14.030, p < .001, \eta_p^2 = .006$) etkileşiminin ise manidar, ancak düşük düzey bir etkiye sahip oldukları görülmektedir.

Dörtlü etkileşim etkileri dikkate alındığında ise; istatistiksel teknik x referans grubun örneklem büyüklüğü x DMF türü x DMF miktarı ($F_{(12,20272)} = 233.388, p < .001, \eta_p^2 = .121$) etkileşiminin manidar ve orta düzey bir etkiye sahip olduğu belirtilebilirken, istatistiksel teknik x referans grubun örneklem büyüklüğü x grup sayısı x DMF türü ($F_{(10,20272)} = 7.576, p < .001, \eta_p^2 = .004$), istatistiksel teknik x referans grubun örneklem büyüklüğü x grup sayısı x DMF miktarı ($F_{(8,20272)} = 5.128, p < .001, \eta_p^2 = .002$) ve istatistiksel teknik x grup sayısı x DMF türü x DMF miktarı ($F_{(12,20272)} = 4.451, p < .001, \eta_p^2 = .003$) etkileşimlerinin manidar, ancak çok düşük düzey bir etkiye sahip oldukları ifade edilebilir. Son olarak, istatistiksel teknik x referans grubun örneklem büyüklüğü x grup sayısı x DMF türü x DMF miktarı ($F_{(12,20272)} = 1.461, p > .05, \eta_p^2 = .001$) beşli etkileşiminin istatistiksel olarak manidar olmadığı sonucuna ulaşılmaktadır.

Araştırma kapsamında gerçekleştirilen ANOVA sonucuna göre çoklu grup teknikleri ile diğer dört faktörün ikili, üçlü, dörtlü ve beşli etkileşimleri için grafikler çizdirilmiş ve GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin Tip I hata açısından performansları betimsel olarak sunulmuştur. Bu amaç doğrultusunda EK 25'te verilen grafikler incelenerek, hesaplanan değerlerin Bradley (1978) tarafından önerilen liberal kritere göre kabul edilebilir olup olmadığı belirlenmiştir.

Çoklu grup teknikleri ve referans grubun örneklem büyüklüğünün ikili etkileşimi incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının, referans grubun örneklem büyüklüğüne bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .629$). Yapılan analizler sonucunda, GLK, GMH ve Lordif için hesaplanan Tip I hata oranlarının referans grubun örneklem büyüklüğüne bağlı olarak yükseldiği, GLR-LRT ve GLR-Wald tekniklerinin gösterdikleri performansların ise referans grubun örneklem büyüklüğüne paralel olarak bir miktar arttığı belirlenmiştir. Alanyazında hem ikili hem de çoklu grupların dikkate alındığı araştırmalarda (Hidalgo vd., 2014; Li ve Zumbo, 2009; Oshima vd., 2015) diğer koşullar sabitken örneklem büyüklüğü arttıkça istatistiksel tekniklere ilişkin hesaplanan Tip I hata oranlarının yükselme eğiliminde olduğu açıklandığından, elde edilen sonucun ilgili çalışmalarla desteklendiği görülmüştür. Yapılan incelemelerde referans grubun örneklem büyüklüğü önemli olmaksızın en yüksek Tip I hata oranına sahip olan tekniğin GMH olduğu anlaşılmıştır. Ayrıca $N_R = 600$ iken GLK, GLR-LRT ve GLR-Wald için, $N_R = 1800$ olduğunda ise GLR-LRT ve GLR-Wald teknikleri için hesaplanan Tip I hata oranlarının liberal kritere göre kabul edilebilir oldukları, diğer teknikler için elde edilen değerlerin ise her iki örneklem büyüklüğünde de kabul edilebilir sınır değerinin üstünde yer aldıkları tespit edilmiştir (bkz. EK 25.1). Bu bilgiler doğrultusunda diğer koşullar sabitken çoklu grup tekniklerine ilişkin performansların genellikle referans grubun örneklem büyüklüğüne bağlı olarak farklılaştığı ortaya konulmuş ve tekniklerin bu performanslarının daha iyi anlaşılabilmesi amacıyla diğer faktörlerle olan üçlü etkileşimlerin incelenmesi gerektiği kanaatine varılmıştır.

Çoklu grup teknikleri ve grup sayısının ikili etkileşimi ele alındığında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının, grup sayısına bağlı olarak manidar bir şekilde farklılaştığı, ancak elde edilen değerlerin grup sayısından ciddi bir şekilde etkilenmediği sonucuna ulaşılmıştır ($\eta_p^2 = .047$). Alanyazında çoklu grupların dikkate alındığı araştırmalarda (Battauz, 2019; Fidalgo ve Scalón, 2010; Penfield, 2001; Oshima vd., 2015) grup sayısı arttıkça istatistiksel tekniklere ilişkin Tip I hata oranlarının önemli bir şekilde farklılaşmadığı açıklandığından, elde edilen sonuçların alanyazınla tutarlı olduğu görülmüştür. Yapılan incelemelerde grup sayısı önemli olmaksızın en yüksek Tip I hata oranına sahip olan tekniğin GMH olduğu ortaya konulmuştur. Grup sayısının hem üç hem de dört olduğu durum için GLR-LRT ve GLR-Wald tekniklerine ilişkin hesaplanan Tip I hata oranlarının

liberal kritere göre kabul edilebilir oldukları belirlenmişken GLK, GMH ve Lordif için elde edilen değerlerin kabul edilebilir sınır değerin üstünde yer aldıkları tespit edilmiştir. Ayrıca grup sayısı arttığında GLR-LRT ve GLR-Wald için Tip I hata oranlarının bir miktar azaldığı, GMH ve Lordif içinse elde edilen değerlerin ise bir miktar arttığı belirlenmiştir (bkz. EK 25.2). Bu bağlamda diğer koşullar sabitken çoklu grup tekniklerine ilişkin performansların genellikle grup sayısına bağlı olarak ciddi bir şekilde farklılaşmadığı ortaya konulmuş ve tekniklerin bu performanslarının daha iyi anlaşılabilmesi amacıyla diğer faktörlerle olan üçlü etkileşimlerin incelenmesi gerektiği sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında çoklu grup teknikleri ve DMF türünün ikili etkileşimi ele alındığında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının, DMF türüne bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .880$). Elde edilen bulgulara göre belirtilen teknikler için hesaplanan temel Tip I hata oranlarının liberal kritere göre kabul edilebilir olduğu gözlemlenmiştir. Hem en düşük temel Tip I hata oranına sahip olan tekniğin GLK olması, hem de GLR-LRT, GLR-Wald ve GMH tekniklerinin birbirlerine son derece yakın performans göstermeleri nedeniyle, bu sonuçların alanyazında yer alan çok sayıda araştırma (Battaüz, 2019; Carroll, 2015; Fidalgo ve Scalón, 2010; Finch, 2016; Jin vd., 2015; Penfield, 2001; Rollins III, 2018; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016) tarafından desteklendiği ortaya konulmuştur. Lordif tekniği için hesaplanan temel Tip I hata oranının ise diğer tekniklerden daha yüksek olmasına rağmen yine de liberal kritere göre kabul edilebilir olduğu belirlenmiştir (bkz. EK 25.3).

Testteki maddeler TB DMF içerdiğinde Lordif için hesaplanan Tip I hata oranının liberal kritere göre kabul edilebilir olduğu belirlenmişken GLK, GLR-LRT, GLR-Wald ve GMH için elde edilen değerlerin kabul edilebilir sınır değerin üstünde yer aldıkları tespit edilmiştir. Belirtilen dört teknik için hesaplanan Tip I hata oranlarının temel Tip I hata oranlarına göre yüksek, Lordif için elde edilen değerlerin ise temel Tip I hata oranına yakın olduğu gözlemlenmiştir. Ayrıca analiz edilen veriler ışığında GLR-LRT, GLR-Wald ve GMH teknikleri için elde edilen değerlerin birbirlerine son derece yakın ve GLK tekniğinden daha yüksek oldukları saptanmış ve bu sonuçların alanyazında yer alan çok sayıda araştırma (Finch, 2016; Gao, 2019; Magis vd., 2011; Rollins III, 2018; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) tarafından desteklendiği sonucuna ulaşılmıştır (bkz. EK 25.3).

Testteki maddeler $TBO(a)$ DMF içerdiğinde GLK, GLR-LRT, GLR-Wald ve Lordif için hesaplanan Tip I hata oranlarının temel Tip I hata oranlarına yakın ve elde edilen değerlerin liberal kritere göre kabul edilebilir olduğu belirlenmiştir. Ayrıca yapılan analizler sonucunda GLR-LRT ve GLR-Wald teknikleri için elde edilen değerlerin birbirlerine son derece yakın (Gao, 2019; Magis vd., 2011) ve GLK tekniğinden daha yüksek oldukları anlaşılmıştır. $TBO(a)$ DMF türünde dikkate alındığı koşullarda Lordif için saflaştırma işlemi otomatik olarak gerçekleştirildiğinden Tip I hata oranının temel Tip I hata oranına yakın olması beklenen bir sonuç olarak kabul edilmiştir. Öte yandan testte birden fazla $TBO(a)$ DMF’li madde bulunması durumunda saflaştırma işlemi gerçekleştirilmemesine rağmen GLK, GLR-LRT ve GLR-Wald tekniklerine ilişkin performansların kabul edilebilir düzeyde ve elde edilen değerlerin temel Tip I hata oranlarına yakın olduğu tespit edilmiştir. Böylelikle testteki dört maddenin $TBO(a)$ DMF içermesi durumunda, belirtilen tekniklerin performanslarının şaşırtıcı bir şekilde DMF’li maddelerin varlığından ciddi bir şekilde etkilenmediği saptanmıştır (bkz. EK 25.3). Alanyazındaki çoklu gruplarda $TBO(a)$ DMF türünün dikkate alındığı araştırmalarda da (Jin vd., 2015; Oshima vd., 2015) saflaştırmanın yapılıp yapılmaması önemli olmaksızın istatistiksel teknikler için hesaplanan Tip I hata oranlarının temel Tip I hata oranlarına yakın olduğu ortaya konulmuştur.

Testteki maddeler $TBO(ab)$ DMF içerdiğinde GLK ve Lordif için Tip I hata oranlarının temel Tip I hata oranlarına göre yüksek ve elde edilen değerlerin kabul edilebilir olmadığı, GLR-LRT ve GLR-Wald için hesaplanan Tip I hata oranlarının ise temel Tip I hata oranlarına benzer ve kabul edilebilir düzeyde olduğu tespit edilmiştir. İlgili çalışmalarda da (Gao, 2019; Magis vd., 2011) belirtildiği gibi bu araştırmada GLR-LRT ve GLR-Wald teknikleri için elde edilen değerlerin benzer olduğu belirlenmiştir. Öte yandan GLK için hesaplanan Tip I hata oranının temel Tip I hata oranından yüksek olduğu saptanmış ve $TBO(ab)$ DMF türü için elde edilen değerlerin $TBO(a)$ DMF türünün aksine TB DMF türü için ulaşılan sonuçlar ile benzer olduğu görülmüştür. GLK’nın bu şaşırtıcı performansı, bu teknik için hesaplanan Tip I hata oranının ayırt edicilikten ziyade güçlük parametresinden kaynaklı DMF’den etkilenerek yükseldiğini ortaya koymuştur. Son olarak testteki maddelerin $TBO(ab)$ DMF’li olması durumunda; Lordif için hesaplanan Tip I hata oranının temel Tip I hata oranından yüksek olduğu anlaşılmıştır. Bu anlamda TB ve $TBO(a)$ DMF türlerinde elde edilen sonuçların aksine $TBO(ab)$ DMF türünde Lordif için hesaplanan yüksek Tip I hata oranı, tekniğin

performansının güçlük ve ayırt edicilik parametrelerinin etkileşimine bağlı olarak azaldığını göstermiştir (bkz. EK 25.3).

Yapılan bu simülasyon çalışmasında çoklu grup teknikleri ve DMF miktarının ikili etkileşimi ele alındığında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının, DMF miktarına bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .332$). Testteki hiçbir madde DMF içermediğinde GLK, GLR-LRT, GLR-Wald, GMH ve Lordif için hesaplanan temel Tip I hata oranlarının liberal kritere göre kabul edilebilir olduğu gözlemlenmiş ve odak grupların madde parametrelerine DMF eklendiğinde GLK, GMH ve Lordif için elde edilen değerlerin temel Tip I hata oranlarına göre daha yüksek olduğu saptanmıştır. Ayrıca odak grupların DMF içerdiği durum göz önüne alındığında; DMF miktarı önemli olmaksızın en yüksek Tip I hata oranının GMH, en düşük Tip I hata oranlarının ise GLR-LRT ve GLR-Wald teknikleri için hesaplandığı belirlenmiştir. Söz konusu çalışmada DMF miktarı düşük iken GMH ve Lordif teknikleri için Tip I hata oranlarının liberal kritere göre kabul edilebilir olmadığı ve DMF miktarı arttıkça da elde edilen değerlerin daha da yükseldiği ortaya konulmuştur. Benzer şekilde DMF miktarı düşük olduğunda GLK için elde edilen değerlerin kabul edilebilir olduğu belirlenmişken DMF miktarı arttıkça Tip I hata oranlarının yükselerek kabul edilebilir sınır değerinin üstünde yer aldıkları tespit edilmiştir. Son olarak GLR-LRT ve GLR-Wald tekniklerine ilişkin performansların birbirlerine son derece benzer (Gao, 2019; Magis vd., 2011) ve DMF miktarı önemli olmaksızın elde edilen değerlerin temel Tip I hata oranlarına yakın (liberal kritere göre kabul edilebilir düzeyde) olduğu görülmüştür (bkz. EK 25.4). Araştırmada GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için yapılan analizlerde saflaştırma işleminin gerçekleştirilmemesi nedeniyle DMF miktarı arttıkça Tip I hata oranlarının yükseldiği bulunmuş ve elde edilen sonuçların TB DMF türünün incelendiği benzer özellikteki araştırmalarla (Jin vd. 2015; Tutz ve Schauburger, 2015) desteklendiği gözlemlenmiştir. Benzer şekilde bu sonucu destekler nitelikte olan iki gruplu çalışmalarda da (Li ve Zumbo, 2009; Stark vd., 2006; Tay vd., 2013) TB DMF miktarı arttıkça istatistiksel teknikler için hesaplanan Tip I hata oranlarının yükseldiği açıklanmıştır. Bu simülasyon çalışmasında GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin farklı DMF türlerindeki performanslarının incelenmesi amacıyla veri üretilmesine rağmen yapılan analizlerde DMF türü dikkate alınmadığından, GMH haricindeki teknikler için bu sonuçlara dayalı olarak yapılacak yorumların yanıltıcı olabileceğini hatırlatmakta yarar vardır. Bu bilgilerden hareketle

belirtilen tekniklerin performanslarının daha iyi anlaşılabilmesi amacıyla diğer faktörlerle olan üçlü etkileşimlerin incelenmesi önemli görülmektedir.

Çoklu grup teknikleri, referans grubun örneklem büyüklüğü ve grup sayısının üçlü etkileşimi incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının, referans grubun örneklem büyüklüğü ve grup sayısının etkileşimine bağlı olarak manidar bir şekilde farklılaştığı, ancak elde edilen değerlerin bu faktörlerin etkileşiminden ciddi bir şekilde etkilenmediği sonucuna ulaşılmıştır ($\eta_p^2 = .013$). Alanyazında çoklu grupların dikkate alındığı araştırmalarda diğer koşullar sabitken; grup sayısı arttığında istatistiksel teknikler için hesaplanan Tip I hata oranlarının önemli bir şekilde farklılaşmadığı (Battauz, 2019; Fidalgo ve Scalón, 2010; Penfield, 2001; Oshima vd., 2015), örneklem büyüklüğü arttığında ise Tip I hata oranlarının yükselme eğiliminde olduğu (Oshima vd., 2015) açıklandığından, elde edilen sonuçların ilgili çalışmalarla tutarlı olduğu anlaşılmıştır. Yapılan incelemelerde referans grubun örneklem büyüklüğü ve grup sayısı önemli olmaksızın en yüksek Tip I hata oranının GMH, en düşük Tip I hata oranının ise GLR-LRT ve GLR-Wald teknikleri için hesaplandığı belirlenmiştir. Ayrıca grup sayısı arttıkça GMH ve Lordif için hesaplanan Tip I hata oranlarının nispeten yükseldiği, GLR-LRT ve GLR-Wald için hesaplanan Tip I hata oranlarının bir miktar azaldığı, GLK tekniği için elde edilen değerlerin ise değişmediği tespit edilmiştir. Öte yandan grup sayısı önemli olmaksızın $N_R = 1800$ olduğunda GLK, GMH ve Lordif için hesaplanan Tip I hata oranlarının, $N_R = 600$ olduğunda GMH ve Lordif için hesaplanan Tip I hata oranlarının liberal kritere göre kabul edilebilir sınır değerden yüksek olduğu saptanmıştır (bkz. EK 25.5). Bu anlamda diğer koşullar sabitken çoklu grup tekniklerine ilişkin performansların referans grubun örneklem büyüklüğü ile grup sayısının etkileşimine bağlı olarak ciddi bir şekilde farklılaşmadığı ortaya konulmuş ve tekniklerin gösterdiği bu performansların daha iyi anlaşılabilmesi amacıyla diğer faktörlerle olan dörtlü ve beşli etkileşimlerin incelenmesi gerektiği sonucuna ulaşılmıştır.

Çoklu grup teknikleri, referans grubun örneklem büyüklüğü ve DMF türünün üçlü etkileşimi ele alındığında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının referans grubun örneklem büyüklüğü ve DMF türünün etkileşimine bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .651$). Testteki hiçbir maddenin DMF içermediği durum için yapılan incelemelerde, belirtilen tekniklere ilişkin hesaplanan temel Tip I hata oranlarının liberal kritere göre

kabul edilebilir olmasının yanı sıra elde edilen değerlerin referans grubun örneklem büyüklüğünden etkilenmemesi nedeniyle, ulaşılan sonuçların hem ikili hem de çoklu gruplar için yapılan çok sayıda araştırma (Battauz, 2019; Chun, 2014; Chun vd., 2016; Finch, 2016; Finch ve French, 2007; Kopf vd., 2015b; Li ve Zumbo, 2009; Penfield, 2001; Stark vd., 2006) ile desteklendiği ortaya konulmuştur (bkz. EK 25.6).

Yalnızca güçlük parametresinin (TB DMF) değişimlendiği koşullarda; referans grubun örneklem büyüklüğü arttıkça GLK, GLR-LRT, GLR-Wald ve GMH tekniklerine ilişkin Tip I hata oranlarının yükseldiği, Lordif için elde edilen değerlerin ise referans grubun örneklem büyüklüğünden etkilenmediği belirlenmiştir. Yapılan analizlerde Lordif için saflaştırmanın otomatik olarak gerçekleştirilmesi nedeniyle (Choi vd., 2011, 2016; Crane, Gibbons vd., 2006; Fischer ve Karl, 2019; Hahn vd., 2014; Opariuc-Dan vd., 2017) elde edilen değerlerin referans grubun örneklem büyüklüğünden etkilenmemesi sonucunda alanyazındaki çalışmaların desteklendiği tespit edilmiştir. Bunun yanı sıra GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için saflaştırma işlemi gerçekleştirilmediğinden, hesaplanan Tip I hata oranları referans grubun örneklem büyüklüğüne bağlı olarak yükselmiş ve elde edilen bu sonucun TB DMF türünün incelendiği benzer özellikteki hem ikili hem de çoklu gruplu araştırmalarla (Carroll, 2015; Chun vd., 2016; Guilera vd., 2013; Stark vd., 2006; Tunç, 2016) tutarlı olduğu tespit edilmiştir. Yapılan incelemelerde referans grubun örneklem büyüklüğü önemli olmaksızın en yüksek Tip I hata oranının GMH, GLR-LRT ve GLR-Wald teknikleri için hesaplandığı belirlenmiştir. Ayrıca $N_R = 600$ olduğunda, GLK ve Lordif için hesaplanan Tip I hata oranlarının liberal kritere göre kabul edilebilir olduğu, GLR-LRT, GLR-Wald ve GMH için elde edilen değerlerin ise kabul edilebilir sınır değerden yüksek ve birbirlerine son derece yakın (Finch, 2016; Rollins III, 2018; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) olduğu tespit edilmiştir. Benzer şekilde $N_R = 1800$ olduğunda, GLK'nın performansının da örneklem büyüklüğündeki artıştan olumsuz etkilenmesi sonucunda Lordif haricindeki teknikler için elde edilen Tip I hata oranlarının yine kabul edilebilir olmadığı saptanmıştır (bkz. EK 25.6).

Yalnızca ayırt edicilik parametresinin [TBO(a) DMF] değişimlendiği koşullarda; GLK, GLR-LRT, GLR-Wald ve Lordif teknikleri için hesaplanan Tip I hata oranlarının referans grubun örneklem büyüklüğünden etkilenmediği ve elde edilen değerlerin liberal kritere göre kabul edilebilir olduğu belirlenmiştir (bkz. EK 25.6). Bu bağlamda Lordif için saflaştırma işleminin otomatik olarak gerçekleştirilmesi nedeniyle ulaşılan bu sonucun beklendiği, GLK, GLR-LRT ve GLR-Wald teknikleri içinse TBO(a) DMF

içeren maddelerde, referans grubun örneklem büyüklüğü artsa dahi saflaştırma işlemine ihtiyaç duyulmadan kabul edilebilir düzeyde Tip I hata oranları elde edilebildiği tespit edilmiştir. Alanyazındaki TBO(*a*) DMF türünün incelenmesi amacıyla yapılan hem ikili hem de çoklu grupların ele alındığı araştırmalarda (Battauz, 2019; Chun vd., 2016; Stark vd., 2006) örneklem büyüklüğü artmasına rağmen teknikler için hesaplanan Tip I hata oranlarının farklılaşmadığı açıklandığından elde edilen sonuçların benzer özellikteki çalışmalarla tutarlı olduğu ortaya konulmuştur.

Hem güçlük hem de ayırt edicilik parametrelerinin [TBO(*ab*) DMF] değişimlendiği koşullarda; GLR-LRT ve GLR-Wald teknikleri için hesaplanan Tip I hata oranlarının referans grubun örneklem büyüklüğünden etkilenmemesi nedeniyle her iki örneklem büyüklüğü için de elde edilen değerlerin kabul edilebilir olduğu belirlenmiştir. Bununla birlikte GLK ve Lordif'in performansları dikkate alındığında; $N_R = 600$ iken, Tip I hata oranlarının yalnızca GLK için kabul edilebilir olduğu, $N_R = 1800$ olduğunda ise her iki teknik için elde edilen değerlerin yükselerek kabul edilebilir sınır değerine çıktığı tespit edilmiştir (bkz. EK 25.6). Alanyazındaki çalışmalara benzer şekilde (Gao, 2019; Magis vd., 2011) yapılan bu araştırmada da GLR-LRT ve GLR-Wald teknikleri için elde edilen değerlerin çok yakın olduğu saptanmıştır. Öte yandan TBO(*ab*) DMF türünün dikkate alındığı bu çalışmada, referans grubun örneklem büyüklüğü arttıkça GLK tekniğine ilişkin Tip I hata oranlarının yükseldiği bulunmuş ve bu sonucun MTK'ya dayalı hem ikili hem de çoklu grup tekniklerinin (MG-NCDIF, LK, DFIT, MTKOO) incelendiği araştırmalarla (Oshima vd., 2015; Seybert ve Stark, 2012; Stark vd., 2006) tutarlı olduğu görülmüştür. Bu bağlamda Lordif için TBO(*ab*) DMF türünde hesaplanan Tip I hata oranlarının, referans grubun örneklem büyüklüğü ile güçlük ve ayırt edicilik parametrelerinin etkileşimine bağlı olarak yükseldiği gözlemlenmiştir.

Yapılan bu araştırmada çoklu grup teknikleri, referans grubun örneklem büyüklüğü ve DMF miktarının üçlü etkileşimi incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının, referans grubun örneklem büyüklüğü ve DMF miktarının etkileşimine bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .182$). Saflaştırma işleminin yapılmadığı bu araştırma kapsamında, testteki dört maddenin DMF içerdiği durumlarda referans grubun örneklem büyüklüğü ve odak gruplara eklenen DMF miktarı önemli olmaksızın en yüksek Tip I hata oranının GMH için hesaplandığı belirlenmiştir. Bununla birlikte diğer koşullar sabit tutulduğunda, genellikle GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının referans grubun örneklem büyüklüğü ve

odak gruplara eklenen DMF miktarının etkileşimine bağlı olarak yükseldiği tespit edilmiştir. DMF miktarı önemli olmaksızın $N_R = 1800$ olduğunda, GLK, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının genel olarak kabul edilebilir olmadığı ortaya konulmuştur (bkz. EK 25.7). Alanyazında hem ikili hem de çoklu grupların dikkate alındığı araştırmalarda da (Hidalgo vd., 2014; Li ve Zumbo, 2009; Oshima vd., 2015; Stark vd., 2006) DMF miktarı ve örneklem büyüklüğü birlikte arttıkça Tip I hata oranlarının yükseldiği bulunmuş ve elde edilen sonuçların benzer araştırmalarla tutarlı olduğu gözlemlenmiştir. Ayrıca referans grubun örneklem büyüklüğü ve DMF miktarı önemli olmaksızın GLR-LRT ile GLR-Wald teknikleri için hesaplanan Tip I hata oranlarının, ilgili araştırmalara benzer şekilde (Gao, 2019; Magis vd., 2011) birbirlerine son derece yakın olduğu saptanmış ve bütün koşullar için elde edilen değerlerin liberal kritere göre kabul edilebilir olduğu belirlenmiştir. Bu simülasyon çalışmasında GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin farklı DMF türlerindeki performanslarının incelenmesi amacıyla veri üretilmesine rağmen yapılan analizlerde DMF türü dikkate alınmadığından, GMH haricindeki teknikler için bu sonuçlara dayalı olarak yapılacak yorumların yanıltıcı olabileceğini hatırlatmakta yarar vardır. Bu bilgilerden hareketle belirtilen tekniklerin performanslarının daha iyi anlaşılabilmesi amacıyla diğer faktörlerle olan dörtlü ve beşli etkileşimlerin incelenmesi önemli görülmektedir.

Çoklu grup teknikleri, DMF türü ve grup sayısının üçlü etkileşimi ele alındığında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının, DMF türü ve grup sayısının etkileşimine bağlı olarak manidar bir şekilde farklılaştığı, ancak elde edilen değerlerin bu faktörlerin etkileşiminden ciddi bir şekilde etkilenmediği sonucuna ulaşılmıştır ($\eta_p^2 = .013$). Yapılan incelemeler sonucunda, GLR-LRT, GLR-Wald ve GMH teknikleri için elde edilen değerlerin DMF türüne göre değiştiği anlaşılmaya rağmen, tekniklerin performanslarının grup sayısından etkilenmediği belirlenmiştir. Aynı zamanda grup sayısı önemli olmaksızın GLR-LRT ve GLR-Wald tekniklerine ilişkin hesaplanan Tip I hata oranlarının hem TBO(a) hem de TBO(ab) DMF türleri için liberal kritere göre kabul edilebilir olduğu tespit edilmiştir. Bununla birlikte TB DMF türünde ise GLR-LRT, GLR-Wald ve GMH teknikleri için hesaplanan Tip I hata oranlarının kabul edilebilir sınır değerden son derece yüksek olduğu saptanmıştır. Analiz edilen veriler ışığında; GLK'nın TB DMF, Lordif'in ise TBO(ab) DMF türündeki performanslarının grup sayısından etkilenmesine bağlı olarak GLK için elde edilen değerlerin bir miktar azaldığı, Lordif için elde edilen değerlerin ise bir miktar arttığı gözlemlenmiştir (bkz. EK 25.8). Böylece özellikle TB DMF türünde dikkate alınan

grup sayıları arasındaki farkın Tip I hata oranları üzerinde etkili olabileceği anlaşılmış ve tekniklerin bu performanslarının daha iyi anlaşılabilmesi amacıyla diğer faktörlerle olan dörtlü ve beşli etkileşimlerin incelenmesi gerektiği sonucuna ulaşılmıştır.

Çoklu grup teknikleri, DMF türü ve DMF miktarının üçlü etkileşimi incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının, DMF türü ve DMF miktarının etkileşimine bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .334$). Yapılan analizler sonucunda, TB DMF türünde GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için hesaplanan Tip I hata oranlarının DMF miktarına bağlı olarak yükseldiği, Lordif için elde edilen değerlerin ise DMF miktarındaki artıştan etkilenmediği belirlenmiştir. Böylece saflaştırma işleminin otomatik olarak gerçekleştirildiği Lordif için, hesaplanan Tip I hata oranlarının TB DMF miktarı artmasına rağmen bütün koşullarda liberal kritere göre kabul edilebilir olduğu tespit edilmiştir. Buna karşın yalnızca DMF miktarının düşük olduğu koşulda GLK için hesaplanan Tip I hata oranının kabul edilebilir olduğu, diğer koşullarda ise saflaştırma işleminin gerçekleştirilmemesi nedeniyle TB DMF miktarı arttıkça Tip I hata oranlarının yükselmesine bağlı olarak elde edilen değerlerin de kabul edilebilir sınır değerini oldukça üzerine çıktığı saptanmıştır. Benzer şekilde GLR-LRT, GLR-Wald ve GMH teknikleri için de saflaştırma yapılmadığından, hesaplanan Tip I hata oranlarının hiçbir koşulda kabul edilebilir olmadığı ve TB DMF miktarı arttıkça elde edilen değerlerin yükseldiği gözlemlenmiştir (bkz. EK 25.9). Alanyazında Lordif tekniğinin çoklu gruplardaki performansının incelendiği herhangi bir araştırmaya rastlanılmamasına rağmen, ikili grupların ele alındığı simülasyon çalışmalarında (Schauberger ve Mair, 2020; Tunç, 2016) TB DMF miktarı belirli bir düzeye kadar arttığında Lordif için hesaplanan Tip I hata oranlarının ciddi bir şekilde yükselmediği belirtilmiştir. Bununla birlikte GLK, GLR-LRT, GLR-Wald ve GMH tekniklerinin benzer koşullardaki performanslarının incelendiği çalışmalarda (Jin vd. 2015; Tutz ve Schauburger, 2015) saflaştırma yapılmadığı taktirde TB DMF miktarı arttıkça Tip I hata oranlarının yükseldiği açıklanmıştır. Benzer şekilde belirtilen teknikler haricindeki ikili ve çoklu grup tekniklerinin dikkate alındığı çalışmalarda da (Finch ve French, 2008; Hidalgo vd., 2014; Li ve Zumbo, 2009; Oshima vd., 2015; Stark vd., 2006) TB DMF miktarı arttıkça Tip I hata oranlarının yükselme eğiliminde olduğu raporlanmış ve bu çalışmadan elde edilen sonuçların ilgili araştırmalar ile oldukça benzer olduğu belirlenmiştir.

Yalnızca ayırt edicilik parametresinin [TBO(*a*) DMF] değişimlendiği koşullarda; GLK, GLR-LRT, GLR-Wald ve Lordif teknikleri için hesaplanan Tip I hata oranlarının

DMF miktarından etkilenmediği ve elde edilen değerlerin liberal kritere göre kabul edilebilir oldukları ortaya konulmuştur (bkz. EK 25.9). Çalışmada Lordif için saflaştırma işleminin otomatik olarak gerçekleştirilmesi nedeniyle ulaşılan bu sonucun beklendiği, GLK, GLR-LRT ve GLR-Wald teknikleri için ise saflaştırma yapılmamasına rağmen TBO(*a*) DMF içeren maddelerde DMF miktarı artsa dahi kabul edilebilir düzeyde Tip I hata oranları elde edilebildiği tespit edilmiştir. Alanyazındaki TBO(*a*) DMF türünün incelendiği benzer özellikteki hem ikili hem de çoklu gruplu araştırmalarda (Jin vd. 2015; Oshima vd., 2015; Stark vd., 2006) da DMF miktarı artmasına rağmen teknikler için hesaplanan Tip I hata oranlarının farklılaşmadığı belirtildiğinden elde edilen sonuçların ilgili çalışmalarla tutarlı olduğu anlaşılmıştır.

Hem güçlük hem de ayırt edicilik parametrelerinin [TBO(*ab*) DMF] değiştiği koşullarda; GLR-LRT ve GLR-Wald için hesaplanan Tip I hata oranlarının DMF miktarından etkilenmemesi nedeniyle üç DMF miktarı için de elde edilen değerlerin kabul edilebilir olduğu belirlenmiştir. Bununla birlikte GLK ve Lordif'in performansları dikkate alındığında; DMF miktarı düşük iken, Tip I hata oranlarının yalnızca GLK için kabul edilebilir olduğu, DMF miktarı orta ve yüksek olduğunda ise her iki teknik için elde edilen değerlerin yükselerek kabul edilebilir sınır değerine çıktığı tespit edilmiştir (bkz. EK 25.9). Alanyazındaki çalışmalarda belirtildiği üzere (Gao, 2019; Magis vd., 2011) yapılan bu araştırmada da GLR-LRT ve GLR-Wald teknikleri için elde edilen değerlerin birbirlerine çok yakın olduğu saptanmıştır. Öte yandan DMF miktarı arttıkça GLK tekniğine ilişkin Tip I hata oranlarının yükseldiği bulunmuş ve bu sonucun MTK'ya dayalı hem ikili (MTKOO) hem de çoklu grup (MG-NCDIF), tekniklerinin incelendiği araştırmalarla (Oshima vd., 2015; Stark vd., 2006) tutarlı olduğu görülmüştür. Ayrıca bu çalışmada Lordif için çarpıcı bir sonuç elde edilerek TBO(*ab*) DMF türünde hesaplanan Tip I hata oranlarının DMF miktarı ile birlikte güçlük ve ayırt edicilik parametrelerinin etkileşimine bağlı olarak yükseldiği gözlemlenmiştir. Bu kapsamda üçlü etkileşim için elde edilen bu bulgulardan hareketle; TB, TBO(*a*) ve TBO(*ab*) DMF türleri için yapılan analizlerde ek bir saflaştırma işlemi gerçekleştirilmezse eğer, çoklu grup tekniklerine ilişkin performansların odak gruplara eklenen DMF'nin kaynağı ile DMF miktarının etkileşimine bağlı olarak farklılaşabildiği sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında çoklu grup teknikleri, referans grubun örneklem büyüklüğü, DMF türü ve DMF miktarının dördümlü etkileşimleri incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata

oranlarının, referans grubun örneklem büyüklüğü, DMF türü ve DMF miktarının etkileşimine bağlı olarak manidar bir şekilde farklılaştığı belirlenmiştir ($\eta_p^2 = .121$). Bu nedenle istatistiksel olarak manidar olan dörtlü etkileşim için betimsel istatistikler hesaplanmış ve hesaplanan Tip I hata oranları Tablo 26’da verilmiştir.

Tablo 26

Çoklu Grup Teknikleri İçin Hesaplanan Tip I Hata Oranlarına İlişkin Ortalama ve Standart Hatalar

Referans Örneklem Büyüklik	DMF Miktarı	DMF Türü	GLK		GLR-LRT		GLR-Wald		GMH		Lordif		
			Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	
N _R =600	Düşük	DMF Yok	.026	.001	.046	.001	.045	.001	.049	.001	.059	.001	
		<i>b</i>	.035	.001	.085	.001	.084	.001	.085	.001	.067	.001	
		<i>a</i>	.027	.001	.045	.001	.042	.001			.064	.001	
		<i>ab</i>	.045	.000	.045	.000	.042	.000			.085	.001	
	Orta	<i>b</i>	.057	.001	.123	.001	.121	.001	.123	.001	.065	.001	
		<i>a</i>	.027	.000	.045	.001	.042	.001			.065	.001	
		<i>ab</i>	.060	.000	.045	.000	.042	.000			.100	.000	
		<i>b</i>	.074	.001	.152	.001	.150	.001	.153	.001	.066	.001	
	Yüksek	<i>a</i>	.026	.000	.045	.001	.042	.000			.067	.001	
		<i>ab</i>	.074	.000	.045	.000	.042	.000			.113	.000	
		Düşük	DMF Yok	.026	.001	.046	.001	.045	.001	.049	.001	.059	.001
			<i>b</i>	.081	.002	.164	.002	.164	.002	.169	.002	.065	.001
<i>a</i>	.027		.001	.046	.001	.045	.001			.067	.001		
<i>ab</i>	.089		.001	.047	.000	.047	.000			.128	.001		
N _R =1800	Orta	<i>b</i>	.169	.003	.293	.004	.292	.004	.301	.004	.067	.001	
		<i>a</i>	.027	.000	.049	.001	.048	.001			.073	.001	
		<i>ab</i>	.156	.001	.047	.000	.046	.000			.162	.001	
		<i>b</i>	.247	.003	.390	.004	.389	.004	.400	.004	.068	.001	
	Yüksek	<i>a</i>	.027	.000	.050	.000	.050	.000			.079	.001	
		<i>ab</i>	.219	.001	.046	.000	.045	.000			.182	.000	

Araştırma kapsamında 20 maddelik testte hem hiçbir maddenin DMF’li olmadığı hem de dört maddenin DMF içerdiği iki farklı durum için Tip I hata hesaplanmıştır. Tablo 26’da verilen Tip I hata oranları incelendiğinde; referans grubun örneklem büyüklüğü önemli olmaksızın çoklu grup teknikleri [GLK (.026), GLR-LRT (.046), GLR-Wald (.045), GMH (.049), Lordif (.059)] için hesaplanan temel Tip I hata oranlarının genel olarak .05’ten düşük veya bu değere makul ölçüde yakın olduğu (Tay vd., 2016) ve elde edilen değerlerin Bradley (1978) tarafından önerilen liberal kritere (.075) göre iyi kontrol edildiği belirlenmiştir. Böylece testteki hiçbir madde gerçekte DMF’li olmamasına rağmen, bu maddelerin dikkate alınan çoklu grup teknikleri tarafından %2.6 ile %5.9 arasındaki oranlarda DMF’li olarak tespit edilmesi, ilgili tekniklerin temel Tip I hata yapma olasılıklarının kabul edilebilir düzeyde olduğunu göstermiş ve hesaplanan temel Tip I hata oranlarının referans grubun örneklem büyüklüğünden etkilenmediği

anlaşılmıştır. Çoklu gruplar üzerinde ilk kez kullanılan Lordif tekniği için elde edilen bu sonuç, testte hiçbir DMF'li madde bulunmaması durumunda diğer tekniklerin yanı sıra Lordif'in de tercih edilebilir olduğunu göstermiştir.

DMF ile ilgili yapılan iki grubun ele alındığı simülasyon çalışmalarında, istatistiksel güç oranları yorumlanırken, temel Tip I hata oranlarının göz önünde bulundurulması gerektiği çok sayıda araştırmacı tarafından belirtilmiştir (Cheng vd., 2013; Elosua ve Wells, 2013; Feinberg ve Rubright, 2016; Finch, 2005; Gao, 2019; Kim, 2003; Lopez Rivas, 2012; Morris vd., 2019; Rockoff, 2018; Socha vd., 2015). Bu noktada çoklu grupların dikkate alındığı Fidalgo ve Scalón (2010) tarafından yapılan araştırmada, Tip I hata oranları liberal aralığın dışında olan çoklu grup teknikleri için elde edilen istatistiksel güç oranlarının yorumlanmasının anlamsız olduğu vurgulanmıştır. Belirtilen araştırmalarda işaret edildiği üzere, temel Tip I hata oranlarının yüksek olması durumunda istatistiksel tekniklerin gerçekte DMF'li olmayan maddeleri de DMF'li olarak belirleyebilmesi nedeniyle bu teknikler için hesaplanan istatistiksel güç oranlarının sahte bir şekilde yüksek olabileceği açıklanmış ve istatistiksel güç oranlarındaki bu yükselişin beklenen bir durum olduğu belirtilmiştir (Battaüz, 2019). Başka bir ifadeyle temel Tip I hata oranı yüksek olan istatistiksel teknikler için hesaplanan istatistiksel güç oranlarının gerçeği yansıtmayacak şekilde yükselmesine bağlı olarak elde edilen sonuçların yanıltıcı yorumlara sebep olabileceği ifade edilmiştir. Bu açıklamalar doğrultusunda araştırma kapsamında hesaplanan temel Tip I hata oranlarının (.026 ile .059) kabul edilebilir aralıkta olması neticesinde, gerçekte DMF içermeyen maddelerin çoklu grup teknikleri tarafından DMF'li olarak belirlenme olasılığının düşük olduğu ortaya konulmuştur. Bu bulgulardan hareketle çoklu grup teknikleri için hesaplanan istatistiksel güç oranlarının sahte bir şekilde şişirilmemesi nedeniyle istatistiksel güç için elde edilen bulgulara güvenilebileceği sonucuna ulaşılmıştır.

Testteki hiçbir maddenin DMF içermediği durum için yapılan incelemelerde, GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerine ilişkin hesaplanan temel Tip I hata oranlarının referans grubun örneklem büyüklüğünden etkilenmemesi nedeniyle ulaşılan sonuçların hem ikili hem de çoklu grupların ele alındığı çok sayıda araştırma (Battaüz, 2019; Chun, 2014; Chun vd., 2016; Finch, 2016; Finch ve French, 2007; Kopf vd., 2015b; Li ve Zumbo, 2009; Penfield, 2001; Stark vd., 2006) tarafından desteklendiği belirlenmiştir. Böylece testteki maddelere DMF verilmediğinde hem KTK hem de MTK'ya dayalı eşleştirme kriterlerinde herhangi bir kirlenme söz konusu olmadığından grupların örneklem büyüklükleri artsa dahi istatistiksel tekniklere ilişkin hesaplanan

temel Tip 1 hata oranlarının değişmeyeceği sonucuna ulaşılmıştır (bkz. EK 25.10). Alanyazındaki ilgili araştırmalardan biri olan Battauz (2019) tarafından yapılan çalışmada testte hiçbir maddenin DMF'li olmadığı durum için hem 1PLM hem de 2PLM'ye dayalı veri üretildiği taktirde örneklem büyüklüğü artsa dahi GLK tekniğine ilişkin temel Tip 1 hata oranlarının farklılaşmadığı, buna karşın 3PLM'ye dayalı veri üretildiği durumda örneklem büyüklüğünün temel Tip I hata oranları üzerinde etkili olduğu açıklanmıştır. 2PLM'ye dayalı veri üretilen bu araştırmadan elde edilen bulguların, Battauz'un (2019) 1PLM ve 2PLM için elde ettiği sonuçlar ile tutarlı olmasına rağmen 3PLM için ortaya konulan sonuçlar ile farklılaştığı tespit edilmiştir. Bu bağlamda yapılan çalışmalardan yola çıkarak, verilerin üretildiği MTK modelinin teknikler için hesaplanan temel Tip I hata oranları üzerinde etkili olabileceği sonucuna varılmıştır.

Araştırma kapsamında odak grupların DMF içerdiği durum göz önüne alındığında üç DMF türü için de farklı senaryonun olduğu belirlenmiştir. İlk olarak TB DMF türü dikkate alınmış ve yalnızca güçlük parametresinin değiştiği koşullarda Lordif haricindeki teknikler için saflaştırma işleminin gerçekleştirilmemesi nedeniyle Tip I hata oranlarının temel Tip I hata oranlarından yüksek olduğu gözlemlenmiştir. Böylece analiz sürecinde saflaştırma işlemi gerçekleştirilmediği taktirde GLR-LRT, GLR-Wald ve GMH teknikleri için elde edilen değerlerin temel Tip I hata oranlarına göre yüksek olması nedeniyle bu araştırmadan elde edilen bulguların ilgili çalışmalarla (Fidalgo ve Scalón, 2010; Jin vd., 2015; Rollins III, 2018) tutarlı olduğu anlaşılmıştır. Nitekim Carroll'un (2015) çalışmasında da testteki DMF'li maddelerin istatistiksel tekniklerin Tip I hata oranlarını yükselten etkili bir faktör olduğu vurgulanmıştır. Bu anlamda testte birden fazla TB DMF'li madde bulunduğunda, KTK'ya dayalı teknikler ile yapılan analizlerde DMF açısından incelenen madde haricindeki bütün maddelerin DMF içermedikleri varsayıldığından (Magis vd., 2010), testte DMF'li madde yer aldığında hesaplanan Tip I hata oranlarının temel Tip I hata oranından nispeten daha yüksek olması beklenen bir sonuç olarak kabul edilmiştir (bkz. EK 25.10).

Yapılan bu çalışmada diğer koşullar sabitken TB DMF miktarı ve referans grubun örneklem büyüklüğü ayrı ayrı veya birlikte arttığında GLK, GLR-LRT, GLR-Wald ve GMH tekniklerine ilişkin hesaplanan Tip I hata oranlarının yükseldiği tespit edilmiştir. Lordif içinse saflaştırma işlemi otomatik olarak gerçekleştirildiğinden testte birden fazla DMF'li madde bulunmasına rağmen Tip I hata oranlarının yükselmeyerek (Schauberger ve Mair, 2020; Tunç, 2016) elde edilen değerlerin temel Tip I hata oranına yakın olduğu gözlemlenmiştir (bkz. EK 25.10). Böylece Lordif için elde edilen değerlerin ne TB DMF

miktarındaki ne de referans grubun örneklem büyüklüğündeki artıştan etkilenmediği ortaya konulmuştur. Öte yandan Rollins III (2018) testte DMF'li madde yer alması durumunda GLR ve GMH teknikleri için hesaplanan Tip I hata oranlarının yükselme eğiliminde olduğu ifade edilmiştir. Benzer şekilde Dell-Ross (2021), test DMF'li madde içerdiğinde GMH için hesaplanan Tip I hata oranlarının yükseldiğini açıklamıştır. Li ve Zumbo'nun (2009) çalışmasında ise testteki TB DMF'li madde sayısı arttığında Tip I hata oranının bu maddelerden olumsuz etkilenme ihtimalinin yükseleceği belirtilmiştir. İstatistiksel teknikler için hesaplanan daha yüksek Tip I hata oranları sonucunda bu tekniklerin daha düşük performans gösterdiği belirlendiğinden (Kim, 2003; Schauburger ve Tutz, 2016), TB DMF içeren maddelerden oluşan testlerde saflaştırmanın son derece gerekli bir işlem olduğu sonucuna varılmıştır.

DMF belirleme çalışmalarında eşleştirme kriteri olarak genellikle toplam test puanlarının veya yetenek parametrelerinin kullanılabildiği ve testteki herhangi bir madde DMF içerdiğinde eşleştirme kriterinde kirlenme oluşabildiği araştırmacılar tarafından belirtilmiştir. Bu bağlamda testteki herhangi bir maddenin DMF'li olup olmadığı incelenirken, diğer tüm maddelerin DMF içermediği kabul edilmiş (Magis vd., 2010) ve saflaştırma yapılmazsa eğer eşleştirme kriterinde oluşan ciddi miktardaki kirlenmenin DMF sonuçlarına yansıyabileceği ifade edilmiştir (Fikis ve Oshima, 2017; Guilera vd., 2013; Lee ve Geisinger, 2016; Rockoff, 2018). Bu bilgiler doğrultusunda yapılan bu çalışmada da Lordif için saflaştırma işleminin otomatik olarak gerçekleştirilmesine rağmen diğer teknikler için saflaştırmanın yapılmaması nedeniyle GLK, GLR-LRT, GLR-Wald ve GMH tekniklerine ilişkin performansların DMF miktarı ve referans grubun örneklem büyüklüğünden etkilenmesi beklenen bir sonuç olarak görülmüştür. Ayrıca bu tekniklerden KTK'ya dayalı olan GLR-LRT, GLR-Wald ve GMH'ın MTK'ya dayalı olan GLK'ya kıyasla saflaştırma yapılmamasından daha fazla olumsuz etkilenmeleri sonucunda, eşleştirme kriteri olarak kullanılan yetenek parametrelerinin, toplam puanlara göre daha güvenilir olduğu ortaya konulmuştur.

Alanyazında iki grubun dikkate alındığı Guilera ve diğerleri (2013) ile Lee ve Geisinger (2016) tarafından yapılan çalışmalarda MH tekniği için hesaplanan Tip I hata oranları üzerinde saflaştırma işleminin gerçekleştirilip gerçekleştirilmediğinin son derece ekili olduğu belirtilmiştir. Benzer şekilde başka araştırmalarda saflaştırma yapılmazsa eğer DMF'li madde sayısı arttıkça MH ve LR gibi KTK'ya dayalı tekniklerin performanslarının azaldığı açıklanmıştır (French ve Maller, 2007; Narayanan ve Swaminathan, 1996). Finch (2005) tarafından yapılan bir araştırmada ise iki grup dikkate

alınmış ve eşleştirme kriterinin gözlemlenen bir toplam puan olması sonucunda bu yaklaşımın doğasından kaynaklı olarak odak grubun örneklem büyüklüğü arttıkça Tip I hata oranlarının yükseldiği bildirilmiştir. Böylece saflaştırma yapılmayan çalışmalarda TB DMF miktarı ve grupların örneklem büyüklükleri ayrı ayrı veya birlikte arttığında KTK'ya dayalı teknikler için hesaplanan Tip I hata oranlarının artma eğiliminde olduğu sonucuna ulaşılmıştır (Hidalgo vd., 2014; Li ve Zumbo, 2009; Tay vd., 2013).

Alanyazında çoklu gruplarda TB DMF türünün ele alındığı araştırmalarda da saflaştırma yapılmadığı takdirde örneklem büyüklüğü (Carroll, 2015; Chun vd., 2016) ve DMF miktarı (Jin vd., 2015; Oshima vd., 2015; Tutz ve Schauburger, 2015) arttıkça GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için hesaplanan Tip I hata oranlarının yükseldiği açıklandığından elde edilen sonuçların benzer özellikteki çalışmalarla tutarlı olduğu tespit edilmiştir. Aynı zamanda TB DMF türünün ele alındığı birçok araştırmadaki gibi (Finch, 2016; Rollins III, 2018; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) GLR-LRT, GLR-Wald ve GMH tekniklerine ilişkin performansların birbirlerine son derece yakın olduğu belirlenmiştir. Öte yandan Lordif için saflaştırmanın otomatik olarak gerçekleştirilmesi neticesinde Tip I hata oranlarının örneklem büyüklüğü ile DMF miktarına bağlı olarak değişmemesi de önemli bir bulgu olarak kabul edilmiş ve elde edilen sonuçların iki grubun dikkate alındığı çalışmalarla (Schauburger ve Mair, 2020; Tunç, 2016) desteklendiği gözlemlenmiştir. Çok sayıda araştırmacı tarafından da belirtildiği gibi (Choi vd., 2011, 2016; Crane, Gibbons vd., 2006; Fischer ve Karl, 2019; Hahn vd., 2014; Opariuc-Dan vd., 2017) Lordif'in TB DMF türünde ortaya koyduğu bu performans, ilgili tekniğin gerçek uygulamalardan elde edilen verilerin analizinde kullanılmasının oldukça avantajlı olduğunu göstermiştir. Böylelikle yapılan bu simülasyon çalışması sonucunda, testin birden fazla TB DMF'li madde içerdiği durumda otomatik olarak gerçekleştirilen saflaştırma işlemi nedeniyle Lordif tekniğinin daha kullanışlı olduğunun ortaya konulmasının yanı sıra alanyazındaki çalışmalara benzer şekilde GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için saflaştırma işleminin gerçekleştirilmesinin son derece gerekli olduğu ortaya konulmuştur.

Yapılan incelemelerde TB DMF türünde referans grubun örneklem büyüklüğü ve DMF miktarı önemli olmaksızın KTK'ya dayalı olan GMH, GLR-LRT ve GLR-Wald teknikleri için hesaplanan Tip I hata oranlarının MTK'ya dayalı olan GLK'dan daha yüksek olduğu belirlenmiştir. Rollins III'in (2018) yaptığı çalışmada; gözlenen puanların sınava giren bireyin gizil yeteneğinin mükemmel bir temsilcisi olmaması, belirli alt grupların örneklem büyüklükleri azaldığında gözlenen puanların örnekleme hatalarını

içermesi ve testteki her bir madde puanlarının toplamından oluşan gözlenen puanların olası DMF'li maddelerden doğrudan etkilenebilmesi gibi sınırlılıklar nedeniyle gözlenen puan yaklaşımına dayalı tekniklerin gizil puana dayalı tekniklere kıyasla daha düşük performans göstermesinin beklendiği açıklanmıştır. Benzer şekilde model varsayımları karşılandığı sürece gizil puana dayalı tekniklerin daha doğru sonuçlar ortaya koydukları belirtildiğinden (Martinková vd., 2017; Sireci ve Rios, 2013), araştırma kapsamında yapılan analizler ışığında TB DMF türü için GLK'nın KTK'ya dayalı tekniklerden daha düşük Tip I hata oranlarına sahip olmasının kabul edilebilir olduğu sonucuna varılmıştır.

Araştırma kapsamında ikinci olarak TBO(a) DMF türü dikkate alınmış ve yalnızca ayırt edicilik parametresinin değiştiği koşullarda GLK, GLR-LRT, GLR-Wald ve Lordif teknikleri için hesaplanan Tip I hata oranlarının referans grubun örneklem büyüklüğü ve DMF miktarındaki artıştan etkilenmemesi nedeniyle elde edilen değerlerin temel Tip I hata oranlarına çok yakın olduğu belirlenmiştir. Çalışmada GLK, GLR-LRT ve GLR-Wald teknikleri için TBO(a) DMF miktarı ve referans grubun örneklem büyüklüğü ayrı ayrı veya birlikte artsa dahi belirtilen teknikler için hesaplanan Tip I hata oranlarında saflaştırma işlemi gerçekleştirilmemesine rağmen herhangi bir yükselişin söz konusu olmadığı tespit edilmiştir. Öte yandan saflaştırma işleminin otomatik olarak gerçekleştirildiği Lordif için de TBO(a) DMF miktarı ve referans grubun örneklem büyüklüğü ayrı ayrı veya birlikte arttığında Tip I hata oranlarının beklendiği şekilde bu artıştan etkilenmediği görülmüştür (bkz. EK 25.10). Alanyazındaki TBO(a) DMF türünün incelenmesi amacıyla yapılan hem ikili hem de çoklu grupların ele alındığı araştırmalarda da ayrı ayrı örneklem büyüklüğü (Battaüz, 2019; Chun vd., 2016; Stark vd., 2006) ve DMF miktarı (Jin vd. 2015; Oshima vd., 2015; Stark vd., 2006) artmasına rağmen dikkate alınan teknikler için Tip I hata oranlarının farklılaşmadığı açıklanmıştır.

Çoklu grupların ele alınarak saflaştırma işleminin gerçekleştirildiği araştırmaların birinde Battaüz (2019), grupların örneklem büyüklüğü artsa dahi GLK için hesaplanan Tip I hata oranlarının farklılaşmadığını bulmuştur. Öte yandan Jin ve diğerleri (2015) tarafından yapılan araştırmada, saflaştırma yapılmamasına rağmen odak grupların ayırt edicilik parametrelerine eklenen TBO(a) DMF miktarı arttığında GLR için hesaplanan Tip I hata oranlarının bu artıştan ciddi miktarda etkilenmediği tespit edilmiştir. Benzer şekilde Oshima ve diğerlerinin (2015) çalışmalarında, saflaştırma yapıp yapılmadığı hakkında bilgi verilmemiş ve TBO(a) DMF miktarı artsa dahi MG-NCDIF tekniği için hesaplanan Tip I hata oranlarında önemli bir yükselişin söz konusu olmadığı ifade edilmiştir. Bu bilgiler doğrultusunda saflaştırmanın gerçekleştirilip gerçekleştirilmediği

önemli olmaksızın, referans grubun örneklem büyüklüğü ve odak gruplara eklenen TBO(*a*) DMF miktarı arttığında GLK, GLR-LRT, GLR-Wald ve Lordif teknikleri için hesaplanan Tip I hata oranlarının bu artıştan önemli bir şekilde etkilenmediği ve elde edilen sonuçların benzer özellikteki araştırmalarla tutarlı olduğu ortaya konulmuştur.

Çoklu gruplar için testin TBO(*a*) DMF'li madde içerdiği koşullarda Lordif tekniğine ilişkin performansların Tip I hata açısından incelendiği ilk çalışma özelliği taşıyan bu araştırmada, GLK, GLR-LRT ve GLR-Wald teknikleri için saflaştırma işlemi gerçekleştirilmemesine rağmen TBO(*a*) DMF miktarı ile referans grubun örneklem büyüklüğü artsa dahi Tip I hata oranlarının farklılaşmadığı ve MTK ile KTK'ya dayalı tekniklerin benzer performans gösterdiği tespit edilmiştir. TB DMF türündeki bulguların aksine TBO(*a*) DMF içeren maddelerde GLK, GLR-LRT ve GLR-Wald teknikleri için saflaştırma işlemine ihtiyaç duyulmaksızın düşük Tip I hata oranlarının elde edilebilmesi ilginç bir sonuç olarak görülmüş ve böylece eşleştirme kriteri olarak kullanılan yetenek parametresi ile toplam puanın TBO(*a*) DMF'li maddelerin varlığından etkilenmediği ortaya konulmuştur. Böylelikle çoklu gruplar ve TBO(*a*) DMF türünün dikkate alındığı sınırlı sayıda araştırmadan biri olan bu simülasyon çalışmasından elde edilen bulgulara bağlı olarak, GLK, GLR-LRT ve GLR-Wald tekniklerinin performanslarının DMF'nin kaynağına göre farklılaşabildiği sonucuna ulaşılmıştır. Yalnızca ayırt edicilik parametresinden kaynaklı DMF'de belirli bir yetenek düzeyine kadar bir grubun bu yetenek düzeyinin üzerinde ise diğer grupların avantajlı olmasının eşleştirme kriterleri (hem yetenek parametreleri hem de toplam puanlar) üzerinde kirlenmeye yol açmaması sonucunda GLK, GLR-LRT ve GLR-Wald tekniklerinin Tip I hata oranlarının yükselmediği düşünülmüştür.

Araştırma kapsamında son olarak TBO(*ab*) DMF türü dikkate alınmış ve hem güçlük hem de ayırt edicilik parametrelerinin değiştiği koşullarda GLR-LRT ve GLR-Wald teknikleri için hesaplanan Tip I hata oranlarının referans grubun örneklem büyüklüğü ve DMF miktarındaki artıştan etkilenmemesi nedeniyle elde edilen değerlerin temel Tip I hata oranlarına çok yakın ve kabul edilebilir düzeyde olduğu belirlenmiştir. Buna karşın referans grubun örneklem büyüklüğü ve DMF miktarı ayrı ayrı veya birlikte arttığında GLK ve Lordif teknikleri için hesaplanan Tip I hata oranlarının yükseldiği tespit edilmiştir (bkz. EK 25.10). Saflaştırma işleminin otomatik olarak gerçekleştirildiği Lordif'in TB ve TBO(*a*) DMF türlerindeki performanslarının aksine TBO(*ab*) DMF türündeki Tip I hata oranlarının referans grubun örneklem büyüklüğü ile DMF miktarındaki artıştan etkilenmesi, bu tekniğin performansının ilginç bir şekilde güçlük ve

ayırt edicilik parametrelerinin etkileşimine bağlı olarak azaldığını göstermiştir. Böylece Lordif tekniğinin hem çoklu grup hem de TBO(*ab*) DMF türündeki performansının incelendiği ilk araştırma niteliğinde olan bu simülasyon çalışmasında; saflaştırma işleminin TBO(*ab*) DMF türü için hesaplanan Tip I hata oranları üzerinde olumlu yönde bir etkisinin olmadığı bulunmuş ve elde edilen bu çarpıcı sonucun alanyazına katkı sağlayacağı düşünülmüştür.

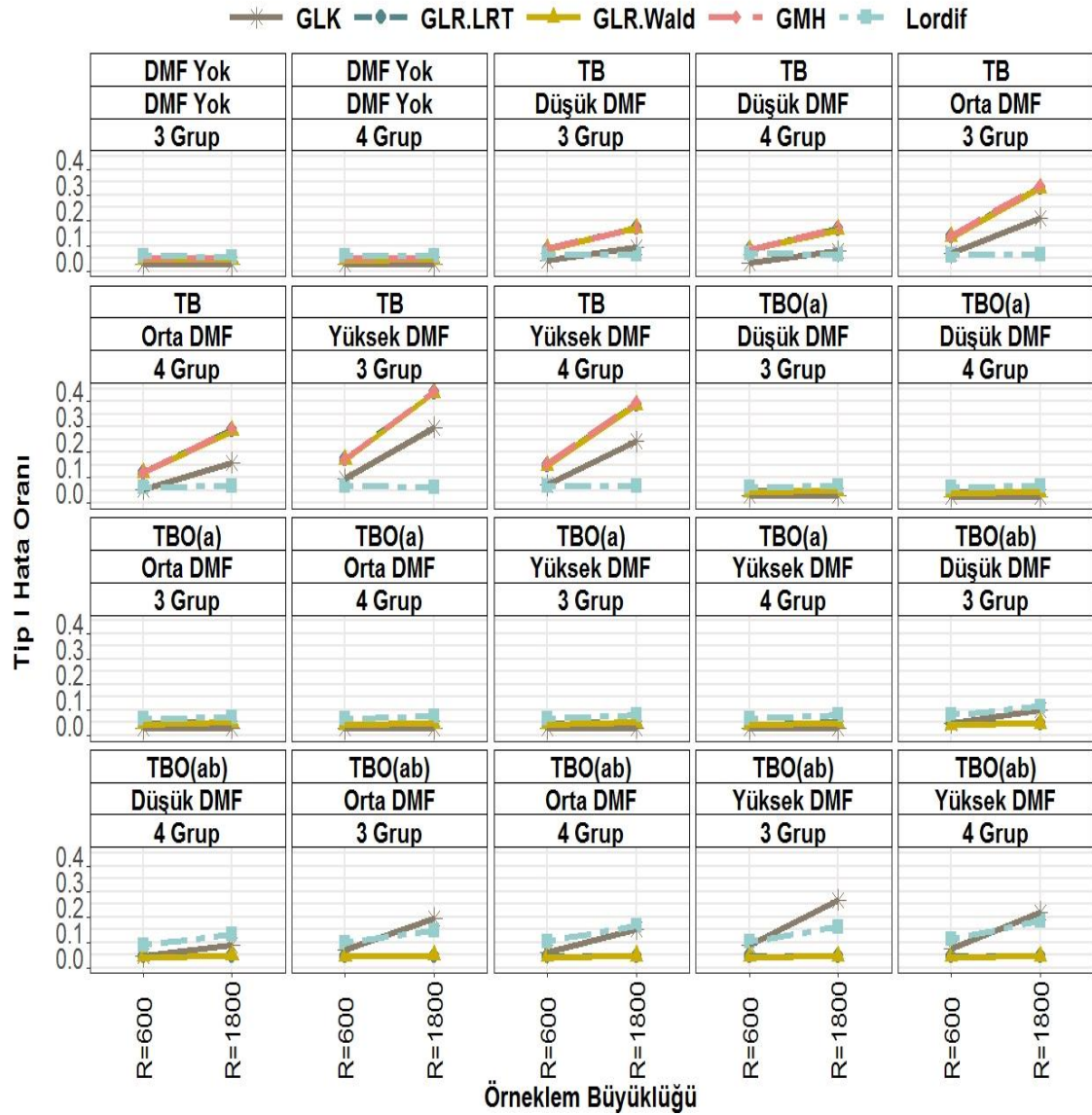
Analizlerin ek bir saflaştırma yapılmadan gerçekleştirildiği bu çalışmada; GLK'nın TBO(*ab*) DMF türünde göstermiş olduğu bu performansın TBO(*a*) DMF türünün aksine TB DMF türü ile benzer olması, bu teknik için hesaplanan Tip I hata oranlarının DMF miktarı ve referans grubun örneklem büyüklüğü ile birlikte güçlük parametresinden kaynaklı DMF'den etkilenerek yükseldiğini ortaya koymuştur. Başka bir ifadeyle GLK'nın TB ve TBO(*ab*) DMF türlerinde referans grubun örneklem büyüklüğü ile DMF miktarındaki artıştan etkilenmesi nedeniyle, bu tekniğin ayırt edicilik parametresinden ziyade güçlük parametresinden kaynaklı DMF'den olumsuz etkilendiği anlaşılmıştır. Alanyazındaki çalışmalarda da TBO(*ab*) DMF miktarı (Oshima vd., 2015; Stark vd., 2006) ve örneklem büyüklüğü (Oshima vd., 2015; Seybert ve Stark, 2012; Stark vd., 2006) arttıkça MTK'ya dayalı tekniklere (MG-NCDIF, LK, DFIT, MTKOO) ilişkin hesaplanan Tip I hata oranlarının yükseldiği açıklandığından, ulaşılan sonuçların desteklendiği görülmüştür. Buna karşın TBO(*ab*) DMF türünün incelendiği Seybert ve diğerlerinin (2014) çalışmalarında saflaştırma yapılması nedeniyle DMF miktarı artsa dahi MTK'ya dayalı olan DFIT ve LK teknikleri için hesaplanan Tip I hata oranlarının temel Tip I hata oranlarına yakın oldukları belirtilmiştir. Bu anlamda MTK kapsamında geliştirilmiş olan GLK için elde edilen bulguların hem ikili hem de çoklu grup tekniklerinin incelendiği araştırmalarla tutarlı olduğu tespit edilmiştir. Böylece benzer özellikteki araştırmaların bulgularına bağlı olarak; odak grupların güçlük parametreleri farklılaştırıldığı takdirde saflaştırma işlemi gerçekleştirilmezse eğer MTK'ya dayalı GLK tekniğine ilişkin elde edilen Tip I hata oranlarının hem DMF miktarı hem de referans grubun örneklem büyüklüğünden olumsuz etkilenebildiği sonucuna ulaşılmıştır.

Yapılan simülasyon çalışmasında TBO(*ab*) DMF türünde GLR-LRT ve GLR-Wald teknikleri için hesaplanan Tip I hata oranlarının hem birbirlerine çok yakın olduğu (Gao, 2019; Magis vd., 2011) hem de elde edilen değerlerin DMF miktarı ile referans grubun örneklem büyüklüğünden etkilenmediği belirlenmiştir. Buna karşın çoklu gruplar için saflaştırma işleminin gerçekleştirilmediği Jin ve diğerleri (2015) tarafından yapılan bir araştırmada GLR tekniğinin performansı TB ile TBO(*a*) DMF türü için dikkate

alınmış ve bu iki DMF türündeki sonuçlardan yola çıkılarak TBO(*ab*) DMF türünde de GLK için elde edilen Tip I hata oranlarının tahminen yükseleceği açıklanmıştır. Nitekim gelinen noktada simüle edilmiş veriler kullanılarak yapılan bu araştırma ile TBO(*ab*) DMF içeren maddelerde GLR-LRT ve GLR-Wald teknikleri için hesaplanan Tip I hata oranlarının TBO(*a*) DMF türünde elde edilen sonuçlara yakın olduğu belirlenmiş ve Jin ve diğerlerinin (2015) çalışmasında ortaya atılan iddianın hatalı olduğu kanıtlanmıştır.

Belirlenen dört maddenin yalnızca güçlük (TB DMF), yalnızca ayırt edicilik [TBO(*a*)] ve hem güçlük hem de ayırt edicilik [TBO(*ab*)] parametrelerinin değişimlendiği koşullarda çoklu grup tekniklerine ilişkin Tip I hata oranlarının incelendiği bu çalışma kapsamında elde edilen bulgulardan hareketle; TB, TBO(*a*) ve TBO(*ab*) DMF türleri için yapılan analizlerde ek bir saflaştırma işlemi gerçekleştirilmezse eğer çoklu grup tekniklerine ilişkin performansların odak gruplara eklenen DMF'nin kaynağı, DMF miktarı ve referans grubun örneklem büyüklüğünün etkileşimine bağlı olarak farklılaşabildiği bulunmuştur. Aynı zamanda her ne kadar kimi yazarlarca testteki DMF'li madde sayısının birden fazla olması durumunda, saflaştırma işleminin gerçekleştirilmesinin gerekli olduğu açıklanmış olsa da (Fikis ve Oshima, 2017; Guilera vd., 2013; Lee ve Geisinger, 2016; Rockoff, 2018) yapılan bu simülasyon çalışması ile saflaştırma işlemine olan gerekliliğin çoklu grup tekniklerinin kuramsal özelliklerinin yanı sıra DMF türüne de bağlı olarak değişebildiği sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında son olarak çoklu grup teknikleri, referans grubun örneklem büyüklüğü, DMF türü, DMF miktarı ve grup sayısının beşli etkileşimleri incelenmiştir. Dört faktöre grup sayısının da eklenmesi sonucunda oluşan beşli etkileşimin istatistiksel olarak manidar olmaması ($F_{(12,20272)} = 1.461, p > .05, \eta_p^2 = .001$) nedeniyle, grup sayısı farklılaşsa dahi GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarında önemli bir değişikliğin olmadığı sonucuna ulaşılmıştır (bkz. Tablo 25). Araştırma kapsamında istatistiksel olarak manidar olan dörtlü etkileşimin yanı sıra manidarlığın söz konusu olmadığı beşli etkileşim için de Tip I hata oranları hesaplanarak, çizilen grafikler Şekil 1'de verilmiştir.



Şekil 1. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü, DMF Türü, DMF Miktarı ve Grup Sayısı Etkileşimi İçin Tip I Hata Oranları

Şekil 1'deki sonuçlara göre; TBO(a) DMF türü için GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerine ilişkin hesaplanan Tip I hata oranlarının DMF miktarı, referans grubun örneklem büyüklüğü ve grup sayısından etkilenmediği belirlenmiştir. TB ve TBO(ab) DMF türlerinde ise bazı çoklu grup teknikleri için elde edilen Tip I hata oranlarının genellikle DMF miktarı ve referans grubun örneklem büyüklüğündeki artışa bağlı olarak yükselme eğiliminde olduğu belirlenmesine rağmen belirtilen tekniklere ilişkin Tip I hata oranlarının üç ve dört grupta son derece yakın olduğu gözlemlenmiştir. Benzer şekilde Carroll (2015) tarafından yapılan çalışmada da Tip I hata oranları üzerinde örneklem büyüklüğünün grup sayısından daha fazla etkili olduğu açıklandığından, elde edilen bulguların desteklendiği anlaşılmıştır.

Çoklu grupların dikkate alındığı çalışmalarda belirtildiği şekilde (Carroll, 2015; Fidalgo ve Scalón, 2010) testte DMF’li madde bulunmadığı durumlar için GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerine ilişkin hesaplanan temel Tip I hata oranlarının grup sayısından etkilenmediği belirlenmiş ve elde edilen bulguların alanyazındaki araştırmalarla tutarlı olduğu anlaşılmıştır. Araştırma kapsamında TB DMF türünde özellikle $N_R = 1800$ ve DMF miktarı orta veya yüksek iken grup sayısı arttıkça, GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için hesaplanan Tip I hata oranlarının bir miktar azaldığı görülmesine rağmen grup sayısı açısından oluşan bu farklılaşmanın istatistiksel olarak manidar olmadığı belirlenmiştir (bkz. Tablo 25). Bu bulguyu destekler nitelikteki çalışmalarda da TB DMF türü için GLK, GLR ve GMH tekniklerine ilişkin performansların grup sayısındaki artışa bağlı olarak farklılaştığına rastlanılmıştır. Bu araştırmaların birinde Penfield (2001) GMH tekniği için grup sayısı ikiden beşe yükseldiğinde Tip I hata oranlarında önemli bir artışın söz konusu olmadığını belirlemiştir. Başka bir araştırmada Fidalgo ve Scalón (2010) TB DMF türünde grup sayısı üçten dörde yükselmesine rağmen GMH tekniği için hesaplanan Tip I hata oranlarının ciddi bir şekilde farklılaşmadığını ortaya koymuşlardır. Svetina ve Rutkowski (2014) tarafından yapılan simülasyon çalışmasında grup sayısının 10 veya 20 olması durumunda GLR için hesaplanan Tip I hata oranlarının değişmediği açıklanmıştır. Schauburger ve Tutz’un (2016) yaptıkları çalışmalarında, TB DMF’de grup sayısı ikiden beşe yükseldiğinde, GLK için elde edilen değerlerin bir miktar azaldığı saptanmıştır. Benzer şekilde Battauz (2019) TB DMF’de grup sayısı ikiden üçe yükseldiğinde GLK için hesaplanan Tip I hata oranlarının nispeten azaldığını ifade etmiştir. Carroll (2015) ise yaptığı çalışmada testte TB DMF’li madde olması durumunda grup sayısının üçten altıya yükselmesinin, LGÇGW tekniği için hesaplanan Tip I hata oranlarında azalmaya yol açtığını belirtmiştir. İlgili araştırmaların bu sonuçlarından hareketle TB DMF türünde çoklu grup tekniklerine ilişkin performansların ya grup sayısından hiç etkilenmediği ya da oluşan etkinin ciddi miktarda olmadığı anlaşılmış ve elde edilen bulguların belirtilen araştırmalarla genel anlamda tutarlı olduğu sonucuna ulaşılmıştır.

Çoklu gruplar için testin TBO(a) DMF’li madde içerdiği koşullarda GLK, GLR-LRT, GLR-Wald ve Lordif teknikleri için hesaplanan Tip I hata oranlarının DMF miktarı, referans grubun örneklem büyüklüğü ve grup sayısındaki artıştan etkilenmediği ve elde edilen değerlerin genel olarak liberal kritere göre kabul edilebilir düzeyde olduğu tespit edilmiştir. Benzer şekilde Battauz (2019) tarafından TBO(a) DMF’nin dikkate alındığı bir araştırmada grup sayısı ikiden üçe yükselse dahi GLK için hesaplanan Tip I hata

oranlarının sabit kaldığı belirtilmiştir. Oshima ve diğerlerinin (2015) çalışmalarında, TBO(*a*) DMF türünde grup sayısı üçten beşe yükseldiğinde MG-NCDIF tekniğinin performansının ciddi bir şekilde etkilenmediği ifade edilmiştir. Dell-Ross (2021) ise araştırmasında TBO(*a*) DMF türü için grup sayısı üçten beşe yükselse dahi MG-NCDIF ve MIMIC tekniklerinin performanslarında bir farklılaşma olmadığını ortaya koymuştur. Böylelikle araştırma kapsamında dikkate alınan TBO(*a*) DMF türündeki GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerine ilişkin performansların DMF miktarı ve referans grubun örneklem büyüklüğünün yanı sıra grup sayısından da etkilenmediği ortaya konulduğundan, bu DMF türündeki bulguların alanyazınla tutarlı olduğu anlaşılmıştır.

Testin TBO(*ab*) DMF’li madde içerdiği koşullarda GLR-LRT ve GLR-Wald teknikleri için hesaplanan Tip I hata oranlarının DMF miktarı ve referans grubun örneklem büyüklüğünün yanı sıra grup sayısındaki artıştan da etkilenmediği ve elde edilen değerlerin genel olarak liberal kritere göre kabul edilebilir düzeyde olduğu saptanmıştır. Bununla birlikte araştırma kapsamında dikkate alınan TBO(*ab*) DMF türünde özellikle $N_R=1800$ ve DMF miktarı orta veya yüksek iken grup sayısı arttıkça, hesaplanan Tip I hata oranlarının Lordif için bir miktar yükseldiği, GLK için ise nispeten azaldığı tespit edilmesine rağmen grup sayısına göre oluşan bu farklılaşmanın istatistiksel olarak manidar olmadığı belirlenmiştir (bkz. Tablo 25). Benzer şekilde TBO(*ab*) DMF türünde grup sayısı üçten beşe yükseldiğinde Oshima ve diğerleri (2015) MG-NCDIF’nin, Dell-Ross (2021) ise MIMIC ile MG-NCDIF tekniklerinin performanslarında genel anlamda bir farklılaşma olmadığını açıkladıklarından, elde edilen bulguların alanyazındaki sınırlı sayıdaki araştırmayla desteklendiği tespit edilmiştir. Böylece üç DMF türü için de grup sayısının artması durumunda Tip I hata oranının mutlaka azalacağı yönünde kesin bir yargının söz konusu olmaması nedeniyle Carroll (2015) tarafından yapılan açıklamaların desteklendiği görülmüştür.

Yapılan bu simülasyon çalışmasında Tip I hata oranları açısından elde edilen bulgular özetlendiğinde; çoklu grup tekniklerine ilişkin performansların grup sayısına bağlı olarak ciddi bir değişiklik göstermediği tespit edilmesine rağmen DMF türü, DMF miktarı ve referans grubun örneklem büyüklüğü faktörlerinin Tip I hatayı etkileyen önemli faktörler oldukları ortaya konulmuştur. Elde edilen bu sonuçlara ilaveten TB DMF türünde KTK’ya dayalı (GLR-LRT, GLR-Wald, GMH), TBO(*ab*) DMF türünde ise MTK’ya dayalı (GLK, Lordif) teknikler için hesaplanan Tip I hata oranlarının daha yüksek olduğu bulunmuş ve bu nedenle tekniklere ilişkin performansların geliştirildikleri kurama bağlı olarak da farklılaşabildiği anlaşılmıştır. Böylece yapılan bu simülasyon

çalışması neticesinde elde edilen bulgular göz önüne alındığında, çoklu grup tekniklerine ilişkin kuramsal özelliklerin de Tip I hata oranlarını etkileyen önemli bir faktör olduğu sonucuna ulaşılmıştır (bkz. Tablo 26 ve Şekil 1).

4.2.2. Grup sayısı (3, 4), referans grubun örneklem büyüklüğü (600, 1800), DMF miktarı (Düşük DMF, Orta Düzey DMF, Yüksek DMF) ve DMF türü [TB, TBO(a), TBO(ab)] koşullarında, GLK, GLR-LRT, GLR-Wald, GMH ve Lordif için istatistiksel güç oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda çoklu grup DMF belirleme tekniklerinin performansları istatistiksel güç açısından incelenmiştir. İstatistiksel güç oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 27’de sunulmuştur.

Tablo 27

İstatistiksel Güç Oranları için Faktöriyel ANOVA Sonucu

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	235.748	4	18924.208	.000	.188	.796
Referans Örneklem Büyüklüğü (R)	533.649	1	171350.211	.000	.425	.899
Grup Sayısı (G)	0.057	1	18.408	.000	.000	.001
DMF Türü (K)	37.767	2	6063.411	.000	.030	.385
DMF Miktarı (M)	122.613	2	19685.031	.000	.098	.670
T x R	86.681	4	6958.172	.000	.069	.590
T x G	0.246	4	19.755	.000	.000	.004
T x K	67.843	6	3630.637	.000	.054	.530
T x M	15.362	8	616.590	.000	.012	.203
T x R x G	0.092	4	7.363	.000	.000	.002
T x R x K	32.227	6	1724.662	.000	.026	.348
T x R x M	11.075	8	444.509	.000	.009	.155
T x G x K	0.092	6	4.916	.000	.000	.002
T x G x M	0.016	8	0.657	.729	.000	.000
T x K x M	5.519	12	147.671	.000	.004	.084
T x R x G x K	0.077	6	4.114	.000	.000	.001
T x R x G x M	0.002	8	0.095	.999	.000	.000
T x R x K x M	2.625	12	70.228	.000	.002	.042
T x G x K x M	0.020	12	0.529	.897	.000	.000
T x R x G x K x M	0.004	12	0.106	.999	.000	.000
Hata	60.282	19356			.048	
Toplam	1255.645					

Tablo 27’de verilen istatistiksel güç sonuçlarına göre; çoklu grup teknikleri ($F_{(4,19356)} = 18924.208$, $p < .001$, $\eta^2 = .188$) ve referans grubun örneklem büyüklüğü ($F_{(1,19356)} = 171350.211$, $p < .001$, $\eta^2 = .425$) ana etkilerinin manidar ve yüksek düzey, DMF miktarı ($F_{(2,19356)} = 19685.031$, $p < .001$, $\eta^2 = .098$) ana etkisinin manidar ve orta düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan DMF türü ($F_{(2,19356)} =$

6063.411, $p < .001$, $\eta^2 = .030$) ana etkisinin manidar ve düşük düzey, grup sayısı ($F_{(1,19356)} = 18.408$, $p < .001$, $\eta^2 = .000$) ana etkisinin ise manidar, ancak çok düşük düzey bir etkiye sahip oldukları anlaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ($F_{(4,19356)} = 6958.172$, $p < .001$, $\eta_p^2 = .590$), istatistiksel teknik x DMF türü ($F_{(6,19356)} = 3630.637$, $p < .001$, $\eta_p^2 = .530$) ve istatistiksel teknik x DMF miktarı ($F_{(8,19356)} = 616.590$, $p < .001$, $\eta_p^2 = .203$) ikili etkileşimlerinin manidar ve yüksek düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan istatistiksel teknik x grup sayısı ($F_{(4,19356)} = 19.755$, $p < .001$, $\eta_p^2 = .004$) ikili etkileşiminin manidar, ancak düşük düzey bir etkiye sahip olduğu gözlemlenmektedir.

Faktörler arasındaki üçlü etkileşimlere göre; istatistiksel teknik x referans grubun örneklem büyüklüğü x DMF türü ($F_{(6,19356)} = 1724.662$, $p < .001$, $\eta_p^2 = .348$) ve istatistiksel teknik x referans grubun örneklem büyüklüğü x DMF miktarı ($F_{(8,19356)} = 444.509$, $p < .001$, $\eta_p^2 = .155$) etkileşimlerinin manidar ve yüksek düzey bir etkiye sahip oldukları anlaşılmaktadır. Ayrıca istatistiksel teknik x DMF türü x DMF miktarı ($F_{(12,19356)} = 147.671$, $p < .001$, $\eta_p^2 = .084$) etkileşiminin manidar ve orta düzey bir etkiye sahip olduğu görülmektedir. Öte yandan istatistiksel teknik x referans grubun örneklem büyüklüğü x grup sayısı ($F_{(4,19356)} = 7.363$, $p < .001$, $\eta_p^2 = .002$) ve istatistiksel teknik x grup sayısı x DMF türü ($F_{(6,19356)} = 4.916$, $p < .001$, $\eta_p^2 = .002$) etkileşimlerinin ise manidar, ancak çok düşük düzey bir etkiye sahip oldukları gözlemlenmektedir. Dörtlü etkileşim etkileri dikkate alındığında ise; istatistiksel teknik x referans grubun örneklem büyüklüğü x DMF türü x DMF miktarı ($F_{(12,19356)} = 70.228$, $p < .001$, $\eta_p^2 = .042$) etkileşiminin manidar ve düşük düzey bir etkiye sahip olduğu belirtilebilirken, istatistiksel teknik x referans grubun örneklem büyüklüğü x grup sayısı x DMF türü ($F_{(6,19356)} = 4.114$, $p < .001$, $\eta_p^2 = .001$) etkileşiminin manidar, ancak çok düşük düzey bir etkiye sahip olduğu ifade edilebilmektedir. Son olarak, istatistiksel teknik x referans grubun örneklem büyüklüğü x grup sayısı x DMF türü x DMF miktarı ($F_{(12,19356)} = 0.106$, $p > .05$, $\eta_p^2 = .000$) beşli etkileşiminin istatistiksel olarak manidar olmadığı sonucuna ulaşılmaktadır.

Araştırma kapsamında gerçekleştirilen ANOVA sonucuna göre çoklu grup teknikleri ile diğer dört faktörün ikili, üçlü, dörtlü ve beşli etkileşimleri için grafikler çizdirilmiş ve GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin istatistiksel

güç açısından performansları betimsel olarak sunulmuştur. Bu amaç doğrultusunda EK 26'da verilen grafikler incelenerek, hesaplanan değerlerin alanyazında önerilen (Lei ve Li, 2013; Scott vd., 2009) kriterlere göre kabul edilebilir olup olmadığı belirlenmiştir.

Çoklu grup teknikleri ile referans grubun örneklem büyüklüğünün ikili etkileşimi incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının, referans grubun örneklem büyüklüğüne bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .590$). Yapılan analizler ışığında, belirtilen teknikler için hesaplanan istatistiksel güç oranlarının referans grubun örneklem büyüklüğüne bağlı olarak yükseldiği ve örneklem büyüklüğü önemli olmaksızın GLR-LRT ve GLR-Wald teknikleri için elde edilen değerlerin diğer tekniklerden daha düşük olduğu anlaşılmıştır. Alanyazında hem ikili hem de çoklu grupların dikkate alındığı araştırmalarda (Battauz, 2019; Cao vd., 2017; Carroll, 2015; Finch, 2016; Guilera vd., 2013; Hidalgo vd., 2014; Lee, 2017; Oshima vd., 2015; Penfield, 2001; Seybert vd., 2014; Stark vd., 2006; Tunç, 2016) diğer koşullar sabitken örneklem büyüklüğü arttıkça istatistiksel tekniklere ilişkin hesaplanan istatistiksel güç oranlarının yükselme eğiliminde olduğu açıklandığından, elde edilen sonuçların ilgili araştırmalarla desteklendiği görülmüştür. Araştırma kapsamında $N_R = 600$ iken GLR-LRT ve GLR-Wald teknikleri için hesaplanan istatistiksel güç oranlarının kabul edilebilir ($.60 \leq \text{güç} < .80$), GLK, GMH ve Lordif teknikleri için hesaplanan değerlerin ise yüksek ($.80 \leq \text{güç}$) düzeyde oldukları tespit edilmiştir. Öte yandan $N_R = 1800$ iken beş çoklu grup tekniği için hesaplanan istatistiksel güç oranlarının da birbirlerine yakın ve .90'ın üzerinde oldukları belirlenmiştir (bkz. EK 26.1). Bu bilgilerden hareketle diğer koşullar sabitken GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerine ilişkin istatistiksel güç oranlarının referans grubun örneklem büyüklüğüne bağlı olarak farklılaştığı ortaya konulmuş ve tekniklerin bu performanslarının daha iyi anlaşılabilmesi amacıyla diğer faktörlerle olan üçlü etkileşimlerin incelenmesi gerektiği kanaatine varılmıştır.

Çoklu grup teknikleri ile DMF türünün ikili etkileşimi incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının, DMF türüne bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .530$). Testteki maddeler TB DMF içerdiğinde istatistiksel güç oranlarının GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için yüksek düzeyde ($.80 \leq \text{güç}$) olduğu belirlenmiş ve belirtilen teknikler arasından Lordif, GLR-LRT, GLR-Wald ve GMH'nin performanslarının, GLK'dan daha yüksek olduğu tespit edilmiştir (bkz.

EK 26.2). Alanyazında TB DMF'nin incelendiği benzer özellikteki araştırmalarda da GLR ve GMH tekniklerinin istatistiksel güç oranları açısından performanslarının GLK'dan yüksek olduğu belirlenmiş (Finch, 2016; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016) ve elde edilen sonuçların son derece tutarlı olduğu görülmüştür. Ayrıca TB DMF türünde analiz edilen veriler ışığında GLR-LRT, GLR-Wald ve GMH teknikleri için elde edilen değerlerin birbirlerine son derece yakın olması (Finch, 2016; Gao, 2019; Magis vd., 2011; Rollins III, 2018; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) ve saflaştırma işleminin otomatik olarak gerçekleştirildiği Lordif'in ise diğerlerinden yüksek performans göstermesi beklenen bir sonuç olarak kabul edilmiştir.

Testteki maddeler TBO(a) DMF içerdiğinde GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerine ilişkin istatistiksel güç oranlarının yüksek düzeyde ($.80 \leq \text{güç}$) olduğu görülmüş ve belirtilen teknikler arasından performansı en düşük olan tekniklerin GLR-LRT ile GLR-Wald, en yüksek olan tekniğin ise Lordif olduğu saptanmıştır (bkz. EK 26.2). Alanyazında TBO(a) DMF'nin ele alındığı Battauz (2019) tarafından yapılan çalışmada da GLK tekniğinin istatistiksel güç açısından performansının yüksek düzeyde olduğu tespit edildiğinden elde edilen değerlerin oldukça tutarlı olduğu belirlenmiştir. Aynı zamanda GLR-LRT ve GLR-Wald teknikleri için hesaplanan Tip I hata oranlarının birbirlerine çok yakın (Gao, 2019; Magis vd., 2011) olması nedeniyle sonuçların desteklendiği sonucuna ulaşılmıştır.

TBO(ab) DMF'nin dikkate alındığı koşullarda ise GLR-LRT ile GLR-Wald için hesaplanan istatistiksel güç oranlarının kabul edilebilir ($.60 \leq \text{güç} < .80$), GLK ile Lordif için elde edilen değerlerin ise yüksek ($.80 \leq \text{güç}$) olduğu anlaşılmış ve en yüksek performansı gösteren tekniğin GLK olduğu belirlenmiştir (bkz. EK 26.2). Lordif için saflaştırma işlemi otomatik olarak gerçekleştirilmesine rağmen, GLK'nın daha yüksek performans göstermesi ilginç bir sonuç olarak kabul edilmiştir. GLK'nın TBO(ab) DMF türündeki bu şaşırtıcı performansı nedeniyle bu teknik için hesaplanan istatistiksel güç oranının güçlük ve ayırt edicilik parametresinin etkileşimine bağlı olarak yükseldiği ortaya konulmuştur. Her ne kadar kimi yazarlarca saflaştırma işleminin gerçekleştirilmesinin gerekli olduğu açıklanmış olsa da (Fikis ve Oshima, 2017; Guilera vd., 2013; Lee ve Geisinger, 2016; Rockoff, 2018) yapılan bu simülasyon çalışması ile analiz edilen veriler ışığında, saflaştırma işlemine olan gerekliliğin çoklu grup tekniklerinin kuramsal özelliklerinin yanı sıra DMF türüne göre de değişebildiği tespit edilmiştir. Bu bilgiler doğrultusunda GLK, GLR-LRT, GLR-Wald ve Lordif için

hesaplanan istatistiksel güç oranlarının DMF türüne bağlı olarak farklılaşabildiği sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında çoklu grup teknikleri ve DMF miktarının ikili etkileşimi ele alındığında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının, DMF miktarına bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .203$). Araştırma kapsamında DMF miktarı düşük iken istatistiksel güç oranı en yüksek olan tekniğin GLK olduğu, DMF miktarı arttığında ise GLK, GMH ve Lordif tekniklerine ilişkin performansların yükselerek birbirlerine yaklaştığı tespit edilmiştir. DMF miktarı düşük olduğunda, hesaplanan ortalama istatistiksel güç oranlarının GLK, GMH ve Lordif için yüksek ($.80 \leq \text{güç}$) GLR-LRT ve GLR-Wald içinse kabul edilebilir ($.60 \leq \text{güç} < .80$) olduğu anlaşılmış ve belirtilen teknikler arasından GLK'nın en yüksek, GLR-LRT ve GLR-Wald'ın ise en düşük performansı gösterdiği belirlenmiştir. Benzer şekilde DMF miktarı orta olduğunda, hesaplanan ortalama istatistiksel güç oranlarının GLK, GMH ve Lordif için yüksek ($.80 \leq \text{güç}$), GLR-LRT ve GLR-Wald teknikleri içinse kabul edilebilir ($.60 \leq \text{güç} < .80$) olduğu görülmüş ve bu teknikler arasından GLK ve GMH'nin en yüksek, GLR-LRT ve GLR-Wald'ın ise en düşük performansı gösterdiği tespit edilmiştir. DMF miktarı yüksek olduğunda ise hesaplanan ortalama istatistiksel güç oranlarının GLK, GMH ve Lordif için birbirlerine yakın ve $.90$ 'ın üzerinde, GLR-LRT ve GLR-Wald içinse $.80$ ile $.90$ arasında olduğu gözlemlenmiştir (bkz. EK 26.3). Alanyazındaki hem ikili hem de çoklu grupların dikkate alındığı araştırmalarda hesaplanan istatistiksel güç oranlarına göre (Cao vd., 2017; Jin vd. 2015; Finch, 2016; Schauburger ve Mair, 2020; Schauburger ve Tutz, 2016; Seybert vd., 2014; Stark vd., 2006; Tay vd., 2013; Tunç, 2016; Tutz ve Berger, 2016) DMF miktarı arttıkça istatistiksel tekniklerin performanslarının genel olarak yükselme eğiliminde olduğu açıklandığından, elde edilen bulguların ilgili çalışmalarla tutarlı olduğu sonucuna ulaşılmıştır. Bu araştırmada GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin istatistiksel güç oranları açısından performansları farklı DMF türlerine göre incelenmesine rağmen yapılan analizlerde DMF türü dikkate alınmadığından, GMH haricindeki teknikler için bu sonuçlara dayalı olarak yapılacak yorumların yanıltıcı olabileceğini hatırlatmakta yarar vardır. Bu bilgilerden hareketle belirtilen tekniklerin performanslarının daha iyi anlaşılabilmesi amacıyla diğer faktörlerle olan üçlü etkileşimlerin incelenmesi önemli görülmektedir.

Çoklu grup teknikleri, referans grubun örneklem büyüklüğü ve DMF türünün üçlü etkileşimi ele alındığında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için

hesaplanan istatistiksel güç oranlarının referans grubun örneklem büyüklüğü ve DMF türünün etkileşimine bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .348$). Yalnızca güçlük parametresinin (TB DMF) değişimlendiği koşullarda referans grubun örneklem büyüklüğü arttıkça GLK, GLR-LRT, GLR-Wald, GMH ve Lordif için hesaplanan istatistiksel güç oranlarının yükseldiği belirlenmiştir. TB DMF türünde $N_R=600$ iken GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için elde edilen değerlerin yüksek ($.80 \leq \text{güç}$), $N_R = 1800$ olduğunda ise bu tekniklerin performanslarının birbirlerine ve 1.00'a çok yakın olduğu saptanmıştır (bkz. EK 26.4). Araştırmada $N_R = 600$ iken Lordif'in diğer tekniklerden daha yüksek performans göstermesi (Choi vd., 2011, 2016; Crane, Gibbons vd., 2006; Fischer ve Karl, 2019; Hahn vd., 2014; Opariuc-Dan vd., 2017) bu teknik için otomatik olarak gerçekleştirilen saflaştırma işleminin bir sonucu olarak görülmüştür. Aynı zamanda $N_R = 600$ iken en düşük performansı gösteren tekniğin GLK olması nedeniyle bu tekniğin yüksek performans gösterebilmesi için ya büyük örneklemelere ihtiyaç olduğu ya da saflaştırmanın gerçekleştirilmesi gerektiği sonucuna ulaşılmış ve elde edilen bu bulgunun Carroll'un (2015) çalışmasıyla tutarlı olduğu gözlemlenmiştir. Benzer şekilde TB DMF türü ve çoklu grupların dikkate alındığı çok sayıda araştırmada da (Battauz, 2019; Carroll, 2015; Finch, 2016; Penfield, 2001; Rollins III, 2018) diğer koşullar sabitken örneklem büyüklüğü arttıkça GLK, GLR ve GMH tekniklerine ilişkin hesaplanan istatistiksel güç oranlarının yükselme eğiliminde olduğu açıklandığından, ulaşılan sonuçların ilgili araştırmalarla desteklendiği anlaşılmıştır.

Ayır edicilik parametresinin [TBO(a) DMF] değişimlendiği koşullarda; GLK, GLR-LRT, GLR-Wald ve Lordif için hesaplanan istatistiksel güç oranlarının referans grubun örneklem büyüklüğüne bağlı olarak yükseldiği görülmüştür. Öyle ki, $N_R = 600$ olduğunda dört çoklu grup tekniği için elde edilen değerlerin kabul edilebilir ($.60 \leq \text{güç} < .80$) sınır değer aralığında yer aldığı, $N_R = 1800$ olduğunda ise belirtilen teknikler için hesaplanan istatistiksel güç oranlarının birbirlerine ve 1.00'a yaklaştığı belirlenmiş ve referans grubun örneklem büyüklüğü önemli olmaksızın Lordif'in diğer tekniklere kıyasla yüksek performans gösterdiği tespit edilmiştir (bkz. EK 26.4). Böylece $N_R = 600$ iken kabul edilebilir düzeyde performans gösteren GLK, GLR-LRT ve GLR-Wald'ın yüksek performans gösterebilmeleri için ya büyük örneklemelere ihtiyaç duyulduğu ya da saflaştırmanın gerçekleştirilmesi gerektiği sonucuna ulaşılmış ve elde edilen bu bulgunun Battauz'un (2019) çalışmasında açıklanan bilgilerle tutarlı olduğu anlaşılmıştır. Benzer şekilde TBO(a) DMF türünün incelendiği hem ikili hem de çoklu grup çalışmalarında

(Battaaz, 2019; Cao vd., 2017; Carroll, 2015; Chun vd., 2016; Finch ve French, 2007, 2008; Lee, 2017; Lopez Rivas, 2012; Oshima vd., 2015; Stark vd., 2006) örneklem büyüklüğü arttıkça istatistiksel teknikler için hesaplanan güç oranlarının yükseldiği açıklandığından elde edilen sonuçların desteklendiği ortaya konulmuştur.

Hem güçlük hem de ayırt edicilik parametrelerinin [TBO(*ab*) DMF] değişimlendiği koşullarda; referans grubun örneklem büyüklüğü arttıkça GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerine ilişkin elde edilen istatistiksel güç oranlarının yükseldiği belirlenmiştir. Araştırmada $N_R = 600$ iken, GLK ve Lordif için istatistiksel güç oranlarının yüksek ($.80 \leq \text{güç}$), GLR-LRT ve GLR-Wald için elde edilen değerlerin ise yaklaşık olarak kabul edilebilir ($.60 \leq \text{güç} < .80$) düzeyde olduğu tespit edilmiştir. Bununla birlikte $N_R = 1800$ olduğunda ise GLK, GLR-LRT, GLR-Wald ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının .90'ın üzerinde ve birbirlerine genel anlamda yakın olduğu gözlemlenmiştir (bkz. EK 26.4). Alanyazındaki çoklu grup çalışmalarında GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin TBO(*ab*) DMF'li maddelerdeki performanslarının incelendiği herhangi bir araştırma bulunmamasına rağmen, farklı tekniklerin dikkate alındığı hem ikili grup hem de çoklu grup araştırmalarında (Oshima vd., 2015; Seybert ve Stark, 2012; Stark vd., 2006) örneklem büyüklüğü arttıkça istatistiksel teknikler için hesaplanan güç oranlarının yükseldiği açıklandığından elde edilen sonuçların benzer özellikteki çalışmalarla desteklendiği görülmüştür. TB, TBO(*a*) ve TBO(*ab*) DMF türleri için elde edilen bu bulgulardan hareketle; DMF türü önemli olmaksızın GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının genellikle referans grubun örneklem büyüklüğüne bağlı olarak yükseldiği ve belirtilen tekniklere ilişkin performansların DMF türüne göre farklılaşabildiği sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında çoklu grup teknikleri, referans grubun örneklem büyüklüğü ve DMF miktarının üçlü etkileşimi incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının, referans grubun örneklem büyüklüğü ve DMF miktarının etkileşimine bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .155$). Bu bağlamda DMF miktarı sabit iken referans grubun örneklem büyüklüğü ve benzer şekilde referans grubun örneklem büyüklüğü sabit iken DMF miktarı arttığında GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerine ilişkin hesaplanan istatistiksel güç oranlarının yükselme eğiliminde olduğu görülmüştür (bkz. EK 26.5).

Araştırmada DMF miktarı düşük iken $N_R = 600$ olduğunda hesaplanan istatistiksel güç oranlarına göre; GLK'nın yüksek, GMH ve Lordif'in kabul edilebilir, GLR-LRT ve GLR-Wald'ın yetersiz, $N_R = 1800$ olduğunda ise GLK, GMH ve Lordif için elde edilen değerlerin .90'ın üzerinde ve birbirlerine yakın, GLR-LRT ve GLR-Wald'ın ise yüksek ve birbirlerine benzer performans ortaya koydukları anlaşılmıştır. DMF miktarının orta düzey olduğu durumda $N_R = 600$ iken hesaplanan istatistiksel güç oranlarına göre; GLK, GMH ve Lordif'in yüksek, GLR-LRT ve GLR-Wald'ın ise kabul edilebilir düzeyde performans sergilediği, $N_R = 1800$ olduğunda ise tekniklerin tamamı için elde edilen değerlerin .90'ın üzerinde ve birbirlerine yakın olduğu tespit edilmiştir. Son olarak DMF miktarı yüksek olduğunda $N_R = 600$ iken hesaplanan istatistiksel güç oranlarına göre; GLK, GMH ve Lordif'in yüksek, GLR-LRT ve GLR-Wald'ın ise kabul edilebilir düzeyde performans gösterdikleri, $N_R = 1800$ olduğunda ise GLK, GLR-LRT, GLR-Wald, GMH ve Lordif için elde edilen değerlerin .99'un üzerinde ve birbirlerine yakın oldukları belirlenmiştir. Böylece özellikle $N_R = 600$ olduğunda DMF miktarı arttıkça en iyi performansı gösteren tekniğin farklılaştığı, $N_R = 1800$ olduğunda ise beş çoklu grup tekniğine ilişkin performansın birbirine çok yakın olduğu ortaya konulmuştur (bkz. EK 26.5). Alanyazında hem ikili hem de çoklu grupların dikkate alındığı araştırmalarda da (Cao vd., 2017; Finch, 2016; Hidalgo vd., 2014; Oshima vd., 2015; Rockoff, 2018; Seybert vd., 2014; Stark vd., 2006) DMF miktarı ve örneklem büyüklüğü birlikte arttıkça istatistiksel güç oranlarının yükseldiği bulunmuş ve elde edilen sonuçların benzer araştırmalarla tutarlı olduğu gözlemlenmiştir. Ayrıca ilgili araştırmalarda belirtildiği şekilde (Gao, 2019; Magis vd., 2011) referans grubun örneklem büyüklüğü ve DMF miktarı önemli olmaksızın GLR-LRT ile GLR-Wald teknikleri için hesaplanan istatistiksel güç oranlarının birbirlerine son derece yakın olduğu saptanmıştır. Bu çalışmada GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin farklı DMF türlerindeki performanslarının incelenmesi amacıyla veri üretilmesine rağmen yapılan analizlerde DMF türü dikkate alınmadığından, GMH haricindeki teknikler için bu sonuçlara dayalı olarak yapılacak yorumların yanıltıcı olabileceğini hatırlatmakta yarar vardır. Böylece belirtilen tekniklerin performanslarının daha iyi anlaşılabilmesi için dördü ve beşli etkileşimlerin incelenmesi önemli görülmektedir.

Çoklu grup teknikleri, DMF türü ve DMF miktarının üçlü etkileşimi incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının, DMF türü ve DMF miktarının etkileşimine bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .084$). Yapılan incelemeler

sonucunda, referans grubun örneklem büyüklüğü dikkate alınmaksızın TB DMF türünde DMF miktarı arttıkça istatistiksel güç oranlarının yükseldiği belirlenmiştir. Araştırmada DMF miktarı düşük olduğunda hesaplanan istatistiksel güç oranlarına göre; GLR-LRT, GLR-Wald, GMH ve Lordif'in yüksek, GLK'nın ise kabul edilebilir düzeyde olduğu, DMF miktarı orta düzey ve yüksek olduğunda ise beş çoklu grup tekniği için elde edilen değerlerin .90'ın üzerinde ve birbirlerine yakın olduğu tespit edilmiştir. Aynı zamanda DMF türü TB iken en yüksek performansı gösteren tekniğin Lordif, en düşük performansı gösteren tekniğin ise GLK olduğu görülmüştür (bkz. EK 26.6). Alanyazında çoklu grupların dikkate alındığı araştırmalarda da TB DMF miktarı arttıkça bütün tekniklerin istatistiksel güç oranlarında önemli bir artışın söz konusu olduğu belirlenmiş ve elde edilen bu sonucun alanyazındaki araştırmalar (Battauz, 2019; Finch, 2016; Jin vd., 2015; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) ile desteklendiği anlaşılmıştır. Lordif tekniğinin çoklu gruplardaki performansının incelendiği herhangi bir araştırmaya rastlanılmamasına rağmen, ikili grupların ele alındığı simülasyon çalışmalarında (Schauburger ve Mair, 2020; Tunç, 2016) TB DMF miktarı arttığında Lordif için hesaplanan istatistiksel güç oranlarının yükseldiği belirtilmiştir. Bununla birlikte çoklu grup tekniklerinin benzer koşullardaki performanslarının incelendiği çalışmalarda (Finch, 2016; Rollins III, 2018; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) GLR ve GMH tekniklerinin son derece benzer ve GLK'dan daha yüksek performans gösterdikleri açıklandığından, ulaşılan sonuçların ilgili araştırmalarla son derece tutarlı olduğu sonucuna ulaşılmıştır.

Araştırmada referans grubun örneklem büyüklüğü dikkate alınmaksızın TBO(a) DMF türünün dikkate alındığı koşullarda istatistiksel güç hesaplandığında; DMF miktarı arttıkça GLK, GLR-LRT, GLR-Wald ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının yükseldiği belirlenmiştir. Araştırmada DMF miktarı düşük olduğunda hesaplanan istatistiksel güç oranlarına göre; GLK, GLR-LRT, GLR-Wald ve Lordif'in kabul edilebilir düzeyde, DMF miktarı orta düzey ve yüksek olduğunda ise belirtilen tekniklere ilişkin performansların yüksek düzeyde olduğu tespit edilmiştir. Aynı zamanda DMF türü TBO(a) iken en yüksek performansı gösteren tekniğin Lordif olduğu, GLK, GLR-LRT ve GLR-Wald tekniklerinin ise birbirlerine yakın ve Lordif'e göre düşük performansı gösterdikleri saptanmıştır (bkz. EK 26.6). Bu noktada yapılan birçok araştırmada saflaştırma işleminin otomatik olarak gerçekleştirilmesinin Lordif'in performansına olumlu yönde katkı sağladığı (Choi vd., 2011, 2016; Crane, Gibbons vd., 2006; Fischer ve Karl, 2019; Hahn vd., 2014; Opariuc-Dan vd., 2017) ve GLR-LRT ile

GLR-Wald tekniklerinin performanslarının benzer olduğu (Gao, 2019; Magis vd., 2011) vurgulandığından, GLR-LRT, GLR-Wald ve Lordif tekniklerinin göstermiş oldukları bu performansların ilgili çalışmalarda açıklanan bilgilerle tutarlı olduğu anlaşılmıştır. Ayrıca hem ikili hem de çoklu grupların dikkate alınarak TBO(a) DMF türünün incelendiği benzer özellikteki araştırmalarda da (Cao vd., 2017; Carroll, 2015; Finch ve French, 2008; Jin vd. 2015; Lee, 2017; Oshima vd., 2015; Stark vd., 2006; Svetina ve Rutkowski, 2014) DMF miktarı arttıkça teknikler için hesaplanan istatistiksel güç oranlarının yükseldiği belirtildiğinden, elde edilen sonuçların çok sayıda çalışma ile desteklendiği görülmüştür.

Araştırmada referans grubun örneklem büyüklüğü dikkate alınmaksızın TBO(ab) DMF türünün dikkate alındığı koşullarda istatistiksel güç hesaplandığında; DMF miktarı arttıkça GLR-LRT, GLR-Wald ve Lordif için elde edilen istatistiksel güç oranlarının yükseldiği, GLK'nın performansında ise ciddi bir farklılaşmanın söz konusu olmadığı tespit edilmiştir. DMF miktarı düşük olduğunda hesaplanan istatistiksel güç oranlarına göre; GLK ve Lordif'in yüksek, GLR-LRT ve GLR-Wald'ın yetersiz, DMF miktarı orta düzey olduğunda ise GLK ve Lordif'in yüksek, GLR-LRT ve GLR-Wald'ın ise kabul edilebilir performans gösterdiği saptanmıştır. Son olarak DMF miktarı yüksek iken GLK ve Lordif için elde edilen değerlerin .90'ın üzerinde ve birbirlerine yakın, GLR-LRT ve GLR-Wald'ın ise yüksek ve birbirlerine son derece benzer performans sergilediği belirlenmiştir. Aynı zamanda DMF türü TBO(ab) olduğunda en yüksek performansı gösteren tekniğin GLK olduğu ve bu tekniğin performansının DMF miktarındaki artıştan etkilenmediği ortaya konulmuştur (bkz. EK 26.6). Alanyazında farklı tekniklerin dikkate alındığı hem ikili grup hem de çoklu grup çalışmalarında (Oshima vd., 2015; Seybert ve Stark, 2012; Stark vd., 2006) DMF miktarı arttıkça istatistiksel teknikler için hesaplanan güç oranlarının yükseldiği açıklandığından elde edilen sonuçların benzer özellikteki çalışmalarla desteklendiği görülmüştür. Aynı zamanda ilgili araştırmalarda belirtildiği üzere yapılan bu araştırmada da GLR-LRT ve GLR-Wald teknikleri için elde edilen değerlerin birbirlerine çok yakın (Gao, 2019; Magis vd., 2011) ve diğer tekniklere kıyasla son derece düşük olduğu saptanmıştır. TB, TBO(a) ve TBO(ab) DMF türleri için elde edilen bu bulgulardan hareketle; DMF türü önemli olmaksızın GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının genellikle odak gruplara eklenen DMF'nin miktarı doğrultusunda yükseldiği ve belirtilen tekniklere ilişkin performansların DMF türüne göre farklılaşabildiği sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında çoklu grup teknikleri, referans grubun örneklem büyüklüğü, DMF türü ve DMF miktarının dörtlü etkileşimleri incelendiğinde;

GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının, referans grubun örneklem büyüklüğü, DMF türü ve DMF miktarının etkileşimine bağlı olarak manidar bir şekilde farklılaştığı belirlenmiştir ($\eta_p^2 = .042$). Bu nedenle istatistiksel olarak manidar olan dörtlü etkileşim için betimsel istatistikler hesaplanmış ve hesaplanan istatistiksel güç oranları Tablo 28’de verilmiştir.

Tablo 28

Çoklu Grup Teknikleri İçin Hesaplanan İstatistiksel Güç Oranlarına İlişkin Ortalama ve Standart Hatalar

Referans Örneklem Büyüklik	DMF Miktarı	DMF Türü	GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
			Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata
$N_R=600$	Düşük	b	.505	.007	.700	.006	.698	.006	.691	.006	.794	.006
		a	.415	.007	.392	.006	.379	.007			.520	.006
		ab	.924	.006	.333	.002	.320	.002			.676	.003
	Orta	b	.805	.006	.920	.003	.918	.003	.916	.003	.970	.002
		a	.653	.008	.630	.006	.612	.007			.771	.005
		ab	.956	.002	.552	.002	.529	.003			.837	.002
	Yüksek	b	.914	.003	.975	.001	.974	.001	.973	.001	.994	.001
		a	.806	.005	.774	.004	.764	.005			.886	.003
		ab	.973	.002	.701	.002	.686	.002			.914	.001
	Düşük	b	.948	.004	.986	.002	.986	.002	.984	.002	.997	.001
		a	.915	.004	.850	.006	.849	.006			.937	.004
		ab	.998	.000	.771	.002	.769	.002			.951	.001
$N_R=1800$	Orta	b	.998	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
		a	.990	.001	.976	.001	.975	.001			.995	.000
		ab	1.000	.000	.952	.001	.950	.001			.994	.000
	Yüksek	b	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
		a	.999	.000	.995	.000	.995	.001			.999	.000
		ab	1.000	.000	.987	.000	.987	.000			.999	.000

Araştırma kapsamında 20 maddelik testte dört maddenin DMF içerdiği durumlar için istatistiksel güç hesaplanmıştır. Tablo 28’de verilen istatistiksel güç oranları incelendiğinde; üç DMF türü için de tekniklere ilişkin istatistiksel güç oranlarının genellikle referans grubun örneklem büyüklüğü ve DMF miktarının birlikte artmasından etkilenecek yükseldiği belirlenmiştir. Simülasyon çalışmalarında .80’in üzerindeki istatistiksel güç oranlarının, tekniklerin performanslarının yüksek olduğunu göstermesi nedeniyle (Finch, 2016; Lei ve Li, 2013; Scott vd., 2009; Svetina ve Rutkowski, 2014) yapılan bu araştırma kapsamında DMF türü ve DMF miktarı önemli olmaksızın $N_R = 1800$ olan bütün koşullarda GLK, GMH ve Lordif tekniklerinin alanyazında belirtilen kriterlere göre yüksek düzeyde performans gösterdikleri ortaya konulmuştur. Benzer şekilde genel olarak GLR-LRT ile GLR-Wald tekniklerinin performanslarının da referans grubun örneklem büyüklüğüne bağlı olarak yükseldiği gözlemlenmesine rağmen DMF

türü TBO(*ab*), DMF miktarı ise düşük olduğunda $N_R = 1800$ olsa dahi bu teknikler için elde edilen istatistiksel güç oranlarının .80’den düşük olduğu tespit edilmiştir. Aynı zamanda üç farklı DMF türünde en yüksek performansı gösteren çoklu grup tekniği ile bu tekniğe ilişkin istatistiksel güç oranlarının farklılaştığı saptanmıştır. $N_R = 600$ ve DMF miktarının düşük olduğu durumlar için TB DMF türünde GLK, TBO(*a*) DMF türünde GLK, GLR-LRT, GLR-Wald ve Lordif, TBO(*ab*) DMF türünde ise GLR-LRT ve GLR-Wald tekniklerinin yetersiz performans ortaya koydukları gözlemlenmiştir (bkz. EK 26.7).

Araştırmada DMF türü TB ve $N_R = 600$ iken bütün DMF miktarlarında lordif’in diğer tekniklere (GLK, GLR-LRT, GLR-Wald ve GMH) kıyasla daha yüksek performans sergilediği görülmüştür. Lordif için saflaştırma işleminin otomatik olarak yapılması nedeniyle (Choi vd., 2011, 2016; Crane vd., 2006) bu teknik ile gerçekleştirilen analiz sürecinde testte yer alan birden fazla TB DMF’li maddenin eşleştirme kriteri üzerindeki etkisi arındırılarak, tekniğin performansının yükselmesine katkı sağlanmıştır (Schauburger ve Mair, 2020; Tunç, 2016). Buna karşın ek bir saflaştırma işleminin gerçekleştirilmediği bu araştırmada benzer koşullar için, MTK’ya dayalı olan GLK’nın KTK’ya dayalı olan GLR-LRT, GLR-Wald ve GMH tekniklerine kıyasla açık bir şekilde daha düşük ve bazı durumlarda yetersiz düzeyde istatistiksel güç oranlarına sahip olduğu belirlenmiştir. Alanyazında TB DMF türünde çoklu grupların incelendiği birçok araştırmada da (Finch, 2016; Rollins III, 2018; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) GLK’nın GLR ve GMH tekniklerinden daha düşük performans gösterdiğinin açıklanması nedeniyle, ulaşılan sonuçların ilgili araştırmalarla tutarlı olduğu anlaşılmıştır. Böylelikle gerçek uygulamalardaki DMF belirleme çalışmalarında, daha doğru kararlar alınabilmesi için ya saflaştırma yapılarak eşleştirme kriterinde oluşabilecek kirlenmeye engel olunması ya da ek bir saflaştırma işlemine ihtiyaç duyulmayan Lordif’in kullanılması gerektiği sonucuna ulaşılmıştır.

Yapılan bu araştırma kapsamında DMF türü TBO(*a*) ve $N_R = 600$ iken bütün DMF miktarlarında lordif’in diğer tekniklere (GLK, GLR-LRT ve GLR-Wald) kıyasla bir miktar daha yüksek performans sergilediği belirlenmesine rağmen, DMF miktarı düşük ve orta düzey olduğunda tekniklerin tamamına ilişkin performansların yetersiz veya kabul edilebilir olduğu tespit edilmiştir. Lordif için saflaştırma işlemi otomatik olarak gerçekleştirilmesine rağmen (Choi vd., 2011, 2016; Crane vd., 2006) TB DMF türünde yüksek performans gösteren bu tekniğin TBO(*a*) DMF türündeki özellikle düşük ve orta düzey DMF miktarı koşullarında yüksek performans göstermemesi ilginç bir sonuç olarak

görülmüştür. Bu bilgiler doğrultusunda TBO(*a*) DMF'li madde içeren gerçek uygulamalardaki DMF belirleme çalışmalarında, daha doğru kararlar alınabilmesi için ya DMF miktarının yüksek ya da referans grubun örneklem büyüklüğünün 1800 olması gerektiği belirlenmiştir. Aynı zamanda alanyazında yapılan açıklamaların (Desjardins ve Bulut, 2018; Martinková vd., 2017; Sireci ve Rios, 2013) aksine dikkate alınan koşullar için, MTK'ya dayalı olan GLK ile KTK'ya dayalı olan GLR-LRT ve GLR-Wald tekniklerinin benzer performans gösterdikleri tespit edilmiştir. GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin en az ikisinin TBO(*a*) DMF türündeki performansının incelendiği herhangi bir araştırmaya rastlanılmaması nedeniyle bu DMF türünde elde edilen sonuçların alanyazına katkı sağlayacağı kanaatine varılmıştır.

Yapılan bu araştırma kapsamında DMF türü TBO(*ab*) ve $N_R = 600$ iken bütün DMF miktarlarında GLK'nın diğer tekniklere (GLR-LRT, GLR-Wald ve Lordif) kıyasla oldukça yüksek ve bütün koşullarda .90 ve üzerinde performans sergilediği gözlemlenmiştir. Lordif için saflaştırma işleminin otomatik olarak gerçekleştirilmesine rağmen (Choi vd., 2011, 2016; Crane vd., 2006) TBO(*ab*) DMF türünde GLK'nın ciddi bir farkla daha yüksek performans göstermesi çarpıcı bir sonuç olarak kabul edilmiştir. Böylece GLK tekniği için TBO(*ab*) DMF türünde hesaplanan istatistiksel güç oranlarının DMF miktarından ziyade güçlük ve ayırt edicilik parametrelerinin etkileşimine bağlı olarak yükseldiği ortaya konulmuştur. Aynı zamanda bütün koşullarda benzer performans ortaya koyan GLR-LRT ve GLR-Wald teknikleri için $N_R = 600$ olduğu taktirde DMF miktarı artsa dahi istatistiksel güç oranlarının yüksek düzeyde olmadığı belirlenmiştir. Bu bilgiler doğrultusunda TBO(*ab*) DMF'li madde içeren gerçek uygulamalardaki DMF belirleme çalışmalarında, doğru kararlar alınabilmesi için GLK tekniğinin kullanılmasının son derece uygun olacağı sonucuna ulaşılmıştır. Bununla birlikte TBO(*ab*) DMF türünde GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin dikkate alındığı herhangi bir araştırmaya rastlanılmaması nedeniyle bu DMF türünde elde edilen sonuçların alanyazına katkı sağlayacağı düşünülmüştür.

Yapılan analizler doğrultusunda DMF türü TBO(*ab*) ve DMF miktarı ise düşük olduğunda oluşan istisna bir durum haricinde DMF türü ile DMF miktarı önemli olmaksızın $N_R = 1800$ olduğunda çoklu grup tekniklerinin istatistiksel güç oranları arasındaki farklılıkların son derece azaldığı belirlenmiştir. Bu bağlamda $N_R = 600$ olduğu durumda tekniklerin istatistiksel güç oranları arasındaki fark genellikle .10 ile .60 arasında değişirken $N_R = 1800$ olduğu durumda istatistiksel güç oranları arasındaki farkın .10'dan büyük olmadığı tespit edilmiştir. Alanyazındaki hem ikili hem de çoklu grupların

dikkate alındığı çok sayıda araştırmada da (Battauz, 2019; Finch, 2016; Penfield, 2001; Seybert vd., 2014; Stark vd., 2006; Tunç, 2016) örneklem büyüklüğü arttıkça DMF belirleme teknikleri için hesaplanan istatistiksel güç oranlarının yükselerek 1.00'a yaklaştığı ve teknikler arasındaki farkların azaldığı belirtildiğinden, elde edilen bu sonucun şaşırtıcı olmadığı anlaşılmıştır. Elde edilen bu sonuç; büyük örneklemelerin dikkate alındığı gerçek test uygulamalarında GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinden herhangi birinin tercih edilebileceğini, küçük örneklemelerin varlığında ise tekniklerin performanslarının diğer koşullara bağlı olarak değişebileceğini göstermiştir. Nitekim MTK'ya dayalı tekniklerin büyük örneklem gerektirdiği birçok araştırmacı tarafından da belirtildiğinden (Camilli, 2006; Clauser ve Mazor, 1998; Fidalgo ve Scalón, 2010; Jin vd., 2018), yapılan bu araştırmada TBO(*ab*) DMF türü haricinde GLK'nın yüksek performans gösterebilmesi için geniş örneklem büyüklüğüne ihtiyaç duyulmasının beklenen bir sonuç olarak kabul edilebileceği sonucuna ulaşılmıştır.

Alanyazındaki araştırmacılar tarafından model varsayımlarının karşılanması durumunda MTK'ya dayalı tekniklerin KTK'ya kıyasla daha fazla güce sahip olabileceği belirtilmesine rağmen (Desjardins ve Bulut, 2018; Martinková vd., 2017; Sireci ve Rios, 2013), yapılan bu simülasyon çalışmasında TB ve TBO(*a*) DMF türleri için GLK'nın KTK'ya dayalı tekniklerden düşük veya onlara benzer performans gösterdiği belirlenmiştir. Böylelikle bu araştırmadan elde edilen bulgular, çok sayıda araştırmacı tarafından yapılan açıklamaların özellikle küçük örneklemelerin dikkate alındığı TB ile TBO(*a*) DMF türleri için gerçeği yansıtmadığını ortaya koymuştur. Çoklu grupların incelendiği Kim ve diğerlerinin (1995) çalışmalarında da GLK'nın küçük örneklem ve düşük DMF miktarlarındaki performansının araştırılması gerektiği vurgulandığından, bu simülasyon çalışmasından elde edilen bulguların yol gösterici olduğu kabul edilmiştir.

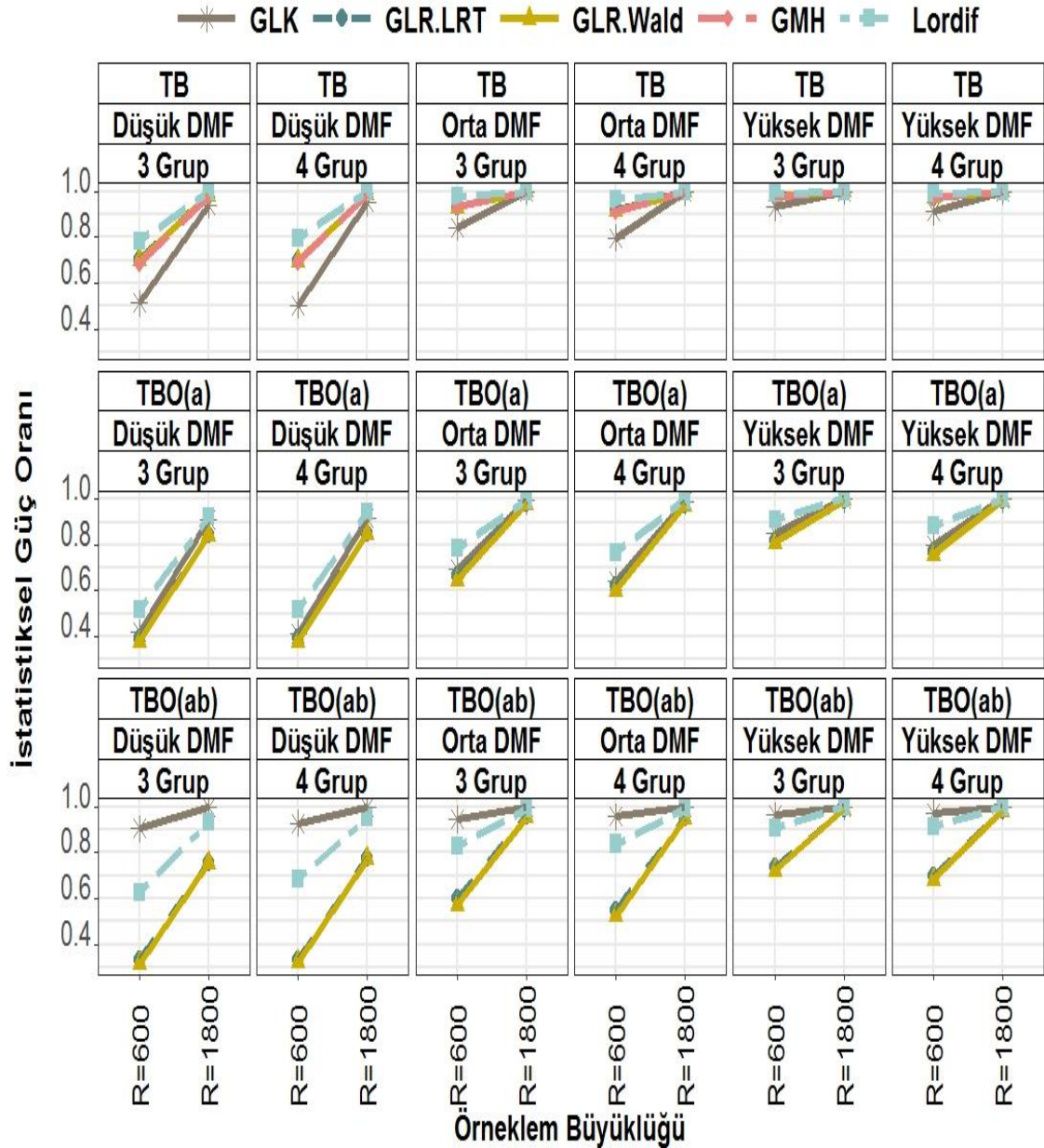
Yapılan bu araştırmada TB, TBO(*a*) ve TBO(*ab*) DMF türleri için referans grubun örneklem büyüklüğü ile DMF miktarı ayrı ayrı veya birlikte arttıkça istatistiksel güç oranlarının ciddi bir şekilde yükseldiği tespit edilmiştir. Alanyazında hem ikili hem de çoklu grupların dikkate alındığı çok sayıda araştırmada da (Cao vd., 2017; Carroll, 2015; Fidalgo ve Cobian, 2018; Finch, 2016; Hidalgo vd., 2014; Jin vd., 2015; Oshima vd., 2015; Rockoff, 2018; Rollins III, 2018; Schauburger ve Tutz, 2016; Seybert vd., 2014; Stark vd., 2006; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) benzer bulgulara rastlanıldığından elde edilen sonuçların desteklendiği görülmüştür. Rockoff (2018) tarafından iki grup için 20 maddelik testteki dört maddesinin TB DMF'li olarak belirlendiği simülasyon çalışmasında, DMF miktarı ve grupların örneklem büyüklüğü

arttıkça MH ve LR teknikleri için hesaplanan istatistiksel güç oranlarının arttığı ifade edilmiştir. Benzer şekilde çoklu grupların dikkate alındığı başka araştırmalarda ise DMF türü önemli olmaksızın genellikle örneklem büyüklüğü (Carroll, 2015; Finch, 2016; Rollins III, 2018) ve DMF miktarı (Carroll, 2015; Finch, 2016; Jin vd., 2015; Schauburger ve Tutz, 2016; Tutz ve Berger, 2016) arttıkça GLK, GLR ve GMH tekniklerinin istatistiksel güç oranları açısından performanslarında ciddi bir yükselişin söz konusu olduğu belirtilmiştir. Böylelikle benzer bulguların raporlanmış olması nedeniyle ilgili çalışmalarda ulaşılan sonuçların son derece tutarlı olduğu anlaşılmıştır. Bu bulgulardan hareketle GLK, GLR-LRT, GLR-Wald GMH ve Lordif teknikleri için istenilen güç oranını (0.80) elde etmek amacıyla gerekli olan en küçük örneklem büyüklüğü ve en düşük DMF miktarı belirlendiğinden, elde edilen bulguların önemli olduğu ve bu simülasyon çalışmasının alanyazına katkı sağlayacağı kabul edilmiştir.

DMF miktarı ve örneklem büyüklüğünün istatistiksel güç oranı ile doğrudan ilişkili olduğu bilinen bir gerçektir ve alanyazındaki çok sayıda araştırmada bu durum açıklanmıştır (Cohen, 1988; Cohen vd., 2018; Cuevas ve Cervantes, 2012; Kim, 2010; Socha vd., 2015; Wood, 2011; Wu vd., 2016; Yu vd., 2008). Bu çalışmalarda genellikle, istatistiksel gücün örneklem büyüklüğüne bağlı olduğu ve örneklem büyüklüğü arttıkça DMF’li maddelerin doğru tespit edilebilme olasılığının yükselerek DMF belirleme tekniklerinin daha doğru sonuçlar ortaya koydukları belirtilmiştir (Cohen, 1988; Cohen vd., 2018; Wood, 2011; Wu vd., 2016). Kim (2010) tarafından yapılan çalışmada diğer koşullar sabit kalmak şartıyla küçük örneklemelerde gerçekten DMF’li olan maddelerin doğru bir şekilde belirlenemediği, büyük örneklemelerde ise DMF’li olan maddelerin doğru tespit edilmesinin yanı sıra DMF miktarı düşük olan maddelerin dahi DMF’li olarak işaretlenebildiği ifade edilmiştir. Benzer şekilde Wood’un (2011) çalışmasında çok büyük DMF miktarı mevcut olmadıkça, küçük örneklemelerde yüksek güç oranları (yani, %80 üzerindeki güç oranları) elde etmenin çok zor olduğuna işaret edilmiştir. Bu bilgiler doğrultusunda DMF miktarı ve örneklem büyüklüğü daha büyük olduğunda yüksek istatistiksel güç oranlarının üretildiği ortaya konulmuştur (Yu vd., 2008). Böylece gerçek uygulamalarda örneklem büyüklüğü ile DMF miktarının yeterince artması durumunda neredeyse tüm DMF’li maddelerin doğru belirlenebildiği anlaşılmıştır.

Yapılan bu simülasyon çalışmasında son olarak çoklu grup teknikleri, referans grubun örneklem büyüklüğü, DMF türü, DMF miktarı ve grup sayısının beşli etkileşimleri incelenmiştir. Dört faktöre grup sayısının da eklenmesi sonucunda oluşan beşli etkileşimin istatistiksel olarak manidar olmaması ($F_{(12,19356)} = 0.106, p > .05, \eta_p^2 =$

.000) nedeniyle, grup sayısı farklılaşsa dahi GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarında önemli bir değişikliğin olmadığı sonucuna ulaşılmıştır (bkz. Tablo 27). Araştırma kapsamında istatistiksel olarak manidar olan dörtlü etkileşimin yanı sıra manidarlığın söz konusu olmadığı beşli etkileşim için de istatistiksel güç oranları hesaplanarak, çizilen grafikler Şekil 2’de verilmiştir.



Şekil 2. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü, DMF Türü, DMF Miktarı ve Grup Sayısı Etkileşimi İçin İstatistiksel Güç Oranları

Şekil 2'deki sonuçlara göre; TB, TBO(a) ve TBO(ab) DMF türlerinde GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının genellikle DMF miktarı ve referans grubun örneklem büyüklüğündeki artışa bağlı olarak yükselme eğiliminde olduğu belirlenmesine rağmen, belirtilen tekniklere ilişkin performansların üç ve dört grup için nispeten benzer olduğu tespit edilmiştir. Carroll (2015) tarafından yapılan çalışmada da istatistiksel güç oranları üzerinde örneklem büyüklüğünün grup sayısından daha fazla etkili olduğu açıklandığından, araştırma kapsamında elde edilen bulguların desteklendiği anlaşılmıştır.

Araştırma kapsamında yapılan analizler sonucunda TB DMF türünde GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerine ilişkin performansların hiçbir koşulda grup sayısından etkilenmediği gözlemlenmiştir. Bununla birlikte TB DMF türünde özellikle $N_R = 600$ ve DMF miktarı orta veya yüksek iken grup sayısı arttıkça, GLK için hesaplanan istatistiksel güç oranlarının bir miktar azaldığı görülmeye rağmen grup sayısına göre oluşan bu farklılaşmanın istatistiksel olarak manidar olmadığı belirlenmiştir (bkz. Tablo 27 ve Şekil 2). Alanyazında TB DMF türünün ele alındığı çoklu grup çalışmaları incelendiğinde de grup sayısındaki artışa bağlı olarak GLK, GLR ve GMH tekniklerine ilişkin performansların farklılaştığını belirten araştırmaların var olduğu görülmüştür.

İlgili araştırmalardan birinde Penfield (2001) grupların yetenek dağılımlarının eşit olduğu durum için GMH tekniğinde grup sayısı ikiden beşe yükseldikçe istatistiksel güç oranlarında önemli bir değişikliğin söz konusu olmadığını açıklamıştır. Svetina ve Rutkowski'nin (2014) çalışmalarında DMF türü TB iken grup sayısının 10 veya 20 olması durumunda GLR için hesaplanan istatistiksel güç oranlarının değişmediği raporlanmıştır. Tutz ve Schauburger (2015) TB DMF türü için diğer koşullar sabit iken yalnızca grup sayısı arttığında GLK, GMH ve GLR teknikleri için hesaplanan istatistiksel güç oranlarında bir farklılaşmanın söz konusu olmadığını ortaya koymuşlardır. Dell-Ross (2021) TB DMF türünde diğer koşullar sabit tutulurken grup sayısı üçten beşe yükseldiğinde GMH için hesaplanan istatistiksel güç oranlarında ciddi bir farklılaşma olmadığını belirtmişlerdir. Öte yandan Fidalgo ve Scalón'un (2010) çalışmalarında GMH tekniği için hesaplanan istatistiksel güç oranlarının grup sayısı üçten dörde yükseldiğinde bir miktar azaldığı ortaya konulmuştur. Benzer şekilde Finch (2016) tarafından yapılan araştırmada grupların örneklem büyüklükleri eşit iken grup sayısı ikiden altıya yükseldiğinde GLK ve GMH teknikleri için hesaplanan istatistiksel güç oranlarının önemli miktarda azaldığı, GLR için elde edilen değerlerin ise farklılaşmadığı belirlenmiştir. Buna karşın grupların örneklem büyüklükleri eşit olmadığında grup sayısı

ikiden altıya yükseldikçe GLK, GLR ve GMH teknikleri için hesaplanan istatistiksel güç oranlarının bir miktar arttığı ortaya konulmuştur. Tutz ve Berger (2016) çalışmalarında TB DMF türünde DMF miktarı düşük iken grup sayısı ikiden beşe yükseldiğinde GLK, GLR ve GMH teknikleri için hesaplanan istatistiksel güç oranlarının azaldığını bildirmişlerdir. Aksine aynı araştırmada DMF miktarı orta veya yüksek iken grup sayısındaki farklılaşmanın istatistiksel güç oranlarını ciddi bir şekilde etkilemediği tespit edilmiştir. TB DMF türü için alanyazındaki sonuçlar incelendiğinde; çoklu gruplara ilişkin performansların grup sayısından ya hiç etkilenmediği ya da istatistiksel güç oranlarının azaldığı anlaşılmış ve bu simülasyon çalışmasından elde edilen bulguların belirtilen ilgili araştırmalarla genel anlamda tutarlı olduğu sonucuna ulaşılmıştır.

Çoklu gruplar için testin TBO(a) DMF’li madde içerdiği koşullarda; özellikle $N_R = 600$ ve DMF miktarı orta veya yüksek iken grup sayısı arttıkça, GLK, GLR-LRT, GLR-Wald ve Lordif teknikleri için hesaplanan istatistiksel güç oranlarının bir miktar azaldığı görülmesine rağmen, grup sayısına göre oluşan bu farklılaşmanın istatistiksel olarak manidar olmadığı belirlenmiştir (bkz. Tablo 27 ve Şekil 2). Alanyazında Lordif tekniğinin TBO(a) DMF türündeki performansının incelendiği herhangi bir simülasyon çalışması bulunmazken, GLR (Svetina ve Rutkowski, 2014) ile GLK (Battauz, 2019) tekniklerine ilişkin performansların araştırıldığı görülmüştür. Svetina ve Rutkowski’nin (2014) çalışmalarında DMF türü TBO(a) iken grup sayısının 10 veya 20 olması durumunda GLR için hesaplanan istatistiksel güç oranlarının değişmediği açıklanmıştır. TBO(a) DMF’nin dikkate alındığı başka bir çalışmada (Battauz, 2019) grup sayısının ikiden üçe yükselmesi durumunda GLK için hesaplanan istatistiksel güç oranlarının da bir miktar yükseldiği ifade edilmiştir. Oshima ve diğerlerinin (2015) çalışmalarında ise TBO(a) DMF türünde diğer koşullar sabitken grup sayısı üçten beşe yükseldiğinde MG-NCDIF tekniğinin istatistiksel güç açısından performansının azaldığı belirtilmiştir. Böylelikle TBO(a) DMF türü için alanyazındaki sonuçlar incelendiğinde; çoklu gruplara ilişkin performansların grup sayısından ya hiç etkilenmediği ya da istatistiksel güç oranlarının azaldığı gözlemlenmiş ve bu simülasyon çalışmasından elde edilen bulguların belirtilen araştırmalarla genel anlamda tutarlı olduğu sonucuna ulaşılmıştır.

TBO(ab) DMF türünün dikkate alındığı koşullarda özellikle $N_R = 600$ ve DMF miktarı düşük iken grup sayısı arttığında, Lordif tekniği için hesaplanan istatistiksel güç oranlarının bir miktar yükseldiği, GLK, GLR-LRT ve GLR-Wald tekniklerinin performanslarının ise değişmediği tespit edilmiştir. Aynı zamanda $N_R = 600$ ve DMF miktarının orta ve yüksek olduğu durum için grup sayısı arttıkça, GLR-LRT ve GLR-

Wald tekniklerine ilişkin performansların bir miktar azaldığı, GLK ve Lordif teknikleri için elde edilen değerlerin ise farklılaşmadığı belirlenmiştir. Böylelikle yapılan araştırmada GLR-LRT, GLR-Wald ve Lordif tekniklerinin bazı koşullardaki performanslarının grup sayısına göre farklılaştığı saptanmasına rağmen, elde edilen bu sonuçların istatistiksel olarak manidar olmadığı anlaşılmıştır (bkz. Tablo 27 ve Şekil 2). Alanyazında GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin TBO(*ab*) DMF türündeki performanslarının incelendiği herhangi bir çalışma bulunmazken, TBO(*ab*) DMF türünün incelendiği bir araştırmada (Oshima vd., 2015) MTK'ya dayalı olan MG-NCDIF tekniği için istatistiksel güç oranları hesaplanmış ve diğer koşullar sabitken grup sayısı üçten beşe yükseldiğinde MG-NCDIF tekniğinin istatistiksel güç açısından performansının azaldığı belirtilmiştir. Benzer şekilde Dell-Ross'un (2021) çalışmasında, grup sayısı üçten beşe yükseldiğinde MG-NCDIF tekniğinin istatistiksel güç açısından performansının bir miktar azaldığı, MIMIC tekniğinde bir farklılaşmanın söz konusu olmadığı raporlanmıştır. Böylece MTK'ya dayalı geliştirilen GLK ile MG-NCDIF tekniklerinin TBO(*ab*) DMF türündeki performanslarının farklı koşullar altında değişebildiği anlaşılmıştır. Bu bulgulardan hareketle çoklu grup tekniklerinin TBO(*ab*) DMF türündeki performansları hakkında daha sağlıklı bilgilere ulaşabilmek için çeşitli koşulların dikkate alındığı simülasyon çalışmalarının yapılmasına ihtiyaç olduğu kanaatine varılmıştır.

Yapılan bu simülasyon çalışmasında istatistiksel güç oranları açısından elde edilen bulgular özetlendiğinde; çoklu grup tekniklerine ilişkin performansların grup sayısına bağlı olarak ciddi bir değişiklik göstermediği tespit edilmesine rağmen DMF türü, DMF miktarı ve referans grubun örneklem büyüklüğü faktörlerinin istatistiksel gücü etkileyen önemli faktörler oldukları ortaya konulmuştur. Elde edilen bu sonuçlara ilaveten TBO(*ab*) DMF türünde MTK'ya dayalı teknikler (GLK, Lordif) için hesaplanan istatistiksel güç oranlarının KTK'ya dayalı tekniklerden (GLR-LRT ve GLR-Wald) daha yüksek oldukları, TBO(*a*) DMF türünde ise KTK ve MTK'ya dayalı tekniklerin benzer performans gösterdikleri bulunmuştur. Öte yandan TB DMF türünde KTK'ya dayalı tekniklerin benzer performans gösterdikleri belirlenmişken, MTK'ya dayalı tekniklerin performanslarının farklılaşabildiği anlaşılmıştır. Böylece yapılan bu simülasyon çalışması neticesinde elde edilen bulgular göz önüne alındığında, çoklu grup tekniklerine ilişkin kuramsal özelliklerin de istatistiksel güç oranlarını etkileyen önemli bir faktör olduğu sonucuna ulaşılmıştır (bkz. Tablo 28 ve Şekil 2).

4.2.3. Grup sayısı (3, 4), referans grubun örneklem büyüklüğü (600, 1800), DMF miktarı (Düşük DMF, Orta Düzey DMF, Yüksek DMF) ve DMF türü [TB, TBO(a), TBO(ab)] koşullarında, GLK, GLR-LRT, GLR-Wald, GMH ve Lordif için Tip III Hata oranları nasıl değişmektedir? araştırma sorusuna ilişkin elde edilen bulgular aşağıda verilmiştir.

Bu araştırma sorusunda çoklu grup DMF belirleme tekniklerinin performansları Tip III Hata açısından incelenmiştir. Tip III Hata oranlarına ilişkin Faktöriyel ANOVA sonuçları Tablo 29’da sunulmuştur.

Tablo 29

Tip III Hata Oranları için Faktöriyel ANOVA Sonucu

	Kareler Toplamı	sd	F	p	η^2	η_p^2
Teknik (T)	8.127	4	4120.358	.000	.006	.417
Referans Örneklem Büyüklüğü (R)	43.951	1	89134.139	.000	.030	.794
Grup Sayısı (G)	0.070	1	142.677	.000	.000	.006
DMF Türü (K)	891.994	2	904506.063	.000	.616	.987
DMF Miktarı (M)	91.602	2	92886.874	.000	.063	.889
T x R	2.785	4	1412.186	.000	.002	.197
T x G	0.093	4	47.261	.000	.000	.008
T x K	274.449	7	79513.900	.000	.190	.960
T x M	4.819	8	1221.723	.000	.003	.297
T x R x G	0.011	4	5.541	.000	.000	.001
T x R x K	4.762	7	1379.616	.000	.003	.295
T x R x M	0.293	8	74.151	.000	.000	.025
T x G x K	0.084	7	24.366	.000	.000	.007
T x G x M	0.003	8	0.795	.607	.000	.000
T x K x M	3.655	14	529.405	.000	.003	.243
T x R x G x K	0.032	7	9.172	.000	.000	.003
T x R x G x M	0.005	8	1.323	.227	.000	.000
T x R x K x M	2.436	14	352.912	.000	.002	.176
T x G x K x M	0.026	14	3.821	.000	.000	.002
T x R x G x K x M	0.006	14	0.867	.595	.000	.001
Hata	11.384	23088				
Toplam	1448.024					

Tablo 29’da verilen Tip III Hata sonuçlarına göre; DMF türü ($F_{(2,23088)} = 904506.063$, $p < .001$, $\eta^2 = .616$) ana etkisinin manidar ve yüksek düzey, DMF miktarı ($F_{(2,23088)} = 92886.874$, $p < .001$, $\eta^2 = .063$) ana etkisinin manidar ve orta düzey, referans grubun örneklem büyüklüğü ($F_{(1,23088)} = 89134.139$, $p < .001$, $\eta^2 = .030$) ana etkisinin ise manidar ve düşük düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan çoklu grup teknikleri ($F_{(4,23088)} = 4120.358$, $p < .001$, $\eta^2 = .006$) ve grup sayısı ($F_{(4,23088)} = 142.677$, $p < .001$, $\eta^2 = .000$) ana etkilerinin manidar ancak çok düşük düzey bir etkiye sahip oldukları anlaşılmaktadır.

Çoklu grup teknikleri ile ikili etkileşimler incelendiğinde; istatistiksel teknik x referans grubun örneklem büyüklüğü ($F_{(4,23088)} = 1412.186, p < .001, \eta_p^2 = .197$), istatistiksel teknik x DMF türü ($F_{(7,23088)} = 79513.900, p < .001, \eta_p^2 = .960$) ve istatistiksel teknik x DMF miktarı ($F_{(8,23088)} = 1221.723, p < .001, \eta_p^2 = .297$) ikili etkileşimlerinin manidar ve yüksek düzey bir etkiye sahip oldukları görülmektedir. Diğer taraftan istatistiksel teknik x grup sayısı ($F_{(4,23088)} = 47.261, p < .001, \eta_p^2 = .008$) ikili etkileşiminin manidar, ancak düşük düzey bir etkiye sahip olduğu gözlemlenmektedir.

Faktörler arasındaki üçlü etkileşimlere göre; istatistiksel teknik x referans grubun örneklem büyüklüğü x DMF türü ($F_{(7,23088)} = 1379.616, p < .001, \eta_p^2 = .295$) ve istatistiksel teknik x DMF türü x DMF miktarı ($F_{(14,23088)} = 529.405, p < .001, \eta_p^2 = .243$) etkileşimlerinin manidar ve yüksek düzey bir etkiye sahip oldukları anlaşılmaktadır. Ayrıca istatistiksel teknik x referans grubun örneklem büyüklüğü x DMF miktarı ($F_{(8,23088)} = 74.151, p < .001, \eta_p^2 = .025$) etkileşiminin manidar ve düşük düzey bir etkiye sahip olduğu belirlenmektedir. Öte yandan istatistiksel teknik x referans grubun örneklem büyüklüğü x grup sayısı ($F_{(4,23088)} = 5.541, p < .001, \eta_p^2 = .001$) ve istatistiksel teknik x grup sayısı x DMF türü ($F_{(7,23088)} = 24.366, p < .001, \eta_p^2 = .007$) etkileşimlerinin manidar, ancak çok düşük düzey bir etkiye sahip oldukları görülmektedir.

Dörtlü etkileşim etkileri dikkate alındığında ise; istatistiksel teknik x referans grubun örneklem büyüklüğü x DMF türü x DMF miktarı ($F_{(14,23088)} = 352.912, p < .001, \eta_p^2 = .176$) etkileşiminin manidar ve yüksek düzey bir etkiye sahip olduğu belirtilebilirken, istatistiksel teknik x referans grubun örneklem büyüklüğü x grup sayısı x DMF türü ($F_{(7,23088)} = 9.172, p < .001, \eta_p^2 = .003$) ve istatistiksel teknik x grup sayısı x DMF türü x DMF miktarı ($F_{(14,23088)} = 3.821, p < .001, \eta_p^2 = .002$) etkileşimlerinin manidar, ancak çok düşük düzey bir etkiye sahip oldukları ifade edilebilmektedir. Son olarak, istatistiksel teknik x referans grubun örneklem büyüklüğü x grup sayısı x DMF türü x DMF miktarı ($F_{(14,23088)} = 0.867, p > .05, \eta_p^2 = .001$) beşli etkileşiminin istatistiksel olarak manidar olmadığı sonucuna ulaşılmaktadır.

Araştırma kapsamında gerçekleştirilen ANOVA sonucuna göre GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri ile diğer dört faktörün ikili, üçlü, dörtlü ve beşli etkileşimleri için grafikler çizdirilmiş ve çoklu grup DMF belirleme tekniklerinin Tip III hata açısından performansları betimsel olarak sunulmuştur. Bu amaç doğrultusunda EK

27’de verilen grafikler incelenerek, hesaplanan deęerlerin Bradley (1978) tarafından önerilen liberal kritere göre kabul edilebilir olup olmadığı belirlenmiştir.

Çoklu grup teknikleri ve referans grubun örneklem büyüklüğünün ikili etkileşimi incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının, referans grubun örneklem büyüklüğüne baęlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .197$). Yapılan analizler sonucunda, referans grubun örneklem büyüklüğü arttıkça belirtilen teknikler için hesaplanan Tip III hata oranlarının yükseldiğı tespit edilmiştir. Alanyazında ikili grupların dikkate alındığı bir araştırmada (Finch ve French, 2008) dięer koşullar sabitken örneklem büyüklüğü arttıkça istatistiksel teknikler için hesaplanan Tip III hata oranlarının yükselme eğiliminde olduğı açıklandığından, elde edilen sonucun desteklendiğı görülmüştür. Araştırmada referans grubun örneklem büyüklüğü önemli olmaksızın en yüksek Tip III hata oranının GMH, en düşük Tip III hata oranının ise GLR-LRT ve GLR-Wald teknikleri için elde edildiğı ve ilgilenilen teknikler için hesaplanan deęerlerin liberal kritere göre kabul edilebilir olmadıkları belirlenmiştir (bkz. EK 27.1). Bu bilgiler doğrultusunda dięer koşullar sabitken çoklu grup tekniklerine ilişkin performansların genellikle referans grubun örneklem büyüklüğüne baęlı olarak farklılaştığı ortaya konulmuş ve tekniklerin bu performanslarının daha iyi anlaşılabilmesi amacıyla dięer faktörlerle olan üçlü etkileşimlerin incelenmesi gerektiğı kanaatine varılmıştır.

Yapılan bu simülasyon çalışmasında çoklu grup teknikleri ve DMF türünün ikili etkileşimi ele alındığında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının, DMF türüne baęlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .960$). Testteki maddeler TB DMF içerdğinde, GLK ve Lordif teknikleri için hesaplanan Tip III hata oranlarının liberal kritere göre kabul edilebilir sınır deęer aralığından son derece yüksek, GLR-LRT ve GLR-Wald teknikleri için elde edilen deęerlerin ise kabul edilebilir olduğı tespit edilmiştir (bkz. EK 27.2). Böylece gerçekte TB DMF içeren maddelere ilişkin DMF türünün GLR-LRT ve GLR-Wald teknikleri ile hatalı belirlenebilme olasılığının son derece düşük ve kabul edilebilir düzeyde olduğı, GLK tekniğıyle ise DMF türünün doğru tespit edilemediğı ortaya konulmuştur. Bu araştırmanın amacına yakın çalışmaların birinde (Lopez Rivas, 2012) TBO DMF’yi belirlemek amacıyla geliştirilen MTKOO tekniğı ile gerçekte TB DMF içeren maddelerdeki Tip III hata oranının son derece yüksek olduğı tespit edildiğinden, GLK’nın performansının MTK’ya dayalı başka bir teknikle desteklendiğı anlaşılmıştır.

Finch ve French (2008) tarafından yapılan iki gruplu başka bir çalışmada, TBO DMF'li maddeleri belirlemek için kullanılan SIBTEST ve MTKOO tekniklerinin gerçekte güçlük parametresinden kaynaklı DMF içeren maddeleri yine DMF'li olarak belirleme oranlarının liberal kritere göre oldukça yüksek olduğu, buna karşın LR tekniği için elde edilen değerin ise kabul edilebilir aralıkta yer aldığı belirtilmiştir. Bu bulgulardan hareketle alanyazındaki çalışmalarda benzer sonuçların elde edildiği gözlemlenmiştir.

Testteki maddeler TBO(*a*) DMF içerdiğinde GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının liberal kritere göre kabul edilebilir sınır değer aralığından son derece yüksek ve bu teknikler arasından GLK'ya ilişkin elde edilen değerlerin nispeten daha düşük olduğu gözlemlenmiştir (bkz. EK 27.2). Böylece gerçekte TBO(*a*) DMF içeren maddelere ilişkin DMF türünün belirtilen tekniklerle genel anlamda doğru tespit edilemediği ortaya konulmuştur. Benzer şekilde alanyazındaki ilgili araştırmalarda da (Finch ve French, 2008; Lopez Rivas, 2012) TB DMF'li maddeleri belirlemek amacıyla geliştirilen SIBTEST, MH, LR ve MTKOO tekniklerinin gerçekte TBO(*a*) DMF içeren maddeleri yine DMF'li olarak belirleme oranının yüksek olduğu belirlenmiş ve bu simülasyon çalışmasından elde edilen bulguların belirtilen araştırmalarla genel anlamda tutarlı olduğu ortaya konulmuştur.

Testteki maddelerin TBO(*ab*) DMF içermesi durumunda GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerine ilişkin Tip III hata oranlarının birbirlerine ve 1.00'a son derece yakın olmasına bağlı olarak, beş çoklu grup tekniği için elde edilen değerlerin liberal kritere göre kabul edilebilir sınır değer aralığından yüksek olduğu ortaya konulmuştur (bkz. EK 27.2). Böylece gerçekte TBO(*ab*) DMF içeren maddelerin TB DMF'yi belirlemek amacıyla geliştirilen tekniklerle analiz edilmesi sonucunda, maddelere ilişkin DMF türünün çok yüksek oranda hatalı belirlendiği görülmüştür. Lopez Rivas (2012) tarafından yapılan araştırmada da gerçekte hem güçlük hem de ayırt edicilik parametresinden kaynaklı DMF'li maddelerin TB DMF'yi belirlemek amacıyla geliştirilen tekniklerle yüksek oranda DMF'li olarak belirlendiği açıklandığından, TBO(*ab*) DMF türünde elde edilen sonuçların son derece tutarlı olduğu anlaşılmıştır. Bu bulgulardan hareketle Tip III hata oranları açısından en düşük veya en yüksek performansı gösteren tekniklerin DMF türüne bağlı olarak farklılaştığı ve yalnızca TB DMF türü için GLR-LRT ve GLR-Wald teknikleri için elde edilen değerlerin kabul edilebilir aralıkta yer aldığı belirlenmiştir. TBO(*a*) ve TBO(*ab*) DMF türlerinde ise çoklu grup tekniklerinden hiçbirinin .075'den düşük Tip III hata oranına sahip olmamasına ek olarak özellikle TBO(*ab*) DMF türünde bütün tekniklerin 1.00'e yakın veya eşit

performans göstermesi çarpıcı bir sonuç olarak kabul edilmiştir. Sonuç olarak Lordif için saflaştırma işlemi otomatik olarak gerçekleştirilmesine rağmen saflaştırmanın yapılıp yapılmaması önemli olmaksızın istatistiksel teknikler için hesaplanan Tip III hata oranlarının genel anlamda yüksek olduğu ortaya konulmuştur.

Yapılan bu simülasyon çalışmasında çoklu grup teknikleri ve DMF miktarının ikili etkileşimi ele alındığında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının, DMF miktarına bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .297$). Yapılan incelemelerde, belirtilen teknikler için hesaplanan Tip III hata oranlarının kabul edilebilir sınır değerin oldukça üzerinde ve birbirlerine yakın olmalarının yanı sıra DMF miktarı arttıkça elde edilen değerlerin yükselme eğiliminde olduğu belirlenmiştir (bkz. EK 27.3). Finch ve French (2008) tarafından yapılan araştırma kapsamında da DMF miktarı arttıkça genel anlamda SIBTEST, MTKOO ve MH teknikleri için hesaplanan Tip III hata oranlarının yükselme eğiliminde olduğu açıklandığından, elde edilen bu sonucun desteklendiği anlaşılmıştır. Aynı zamanda araştırmada bütün DMF miktarları için GLR-LRT ile GLR-Wald tekniklerinin en düşük Tip III hata oranına sahip olduğu ve DMF miktarı düşük iken Lordif, DMF miktarı orta düzey veya yüksek olduğunda ise GMH için elde edilen değerlerin diğer tekniklerden daha yüksek hesaplandığı görülmüştür. Böylece bütün tekniklerin .60'ın üzerinde Tip III hata oranlarına sahip olması nedeniyle, maddelerin DMF türünün çoklu grup teknikleri tarafından yüksek oranda hatalı belirlenebildiği tespit edilmiştir. Buna karşın yapılan bu çalışmada GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin farklı DMF türlerindeki performanslarının incelenmesi amacıyla veri üretilmesine rağmen analizlerde DMF türü dikkate alınmadığından, GMH haricindeki teknikler için yapılacak yorumların yanıltıcı olabileceğini hatırlatmakta yarar vardır. Bu bağlamda tekniklerin performanslarının daha iyi anlaşılabilmesi için diğer faktörlerle olan üçlü etkileşimlerin incelenmesi önemli görülmektedir.

Çoklu grup teknikleri, referans grubun örneklem büyüklüğü ve DMF türünün üçlü etkileşimi ele alındığında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının referans grubun örneklem büyüklüğü ve DMF türünün etkileşimine bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .295$). Araştırma kapsamında yalnızca güçlük parametresinin (TB DMF) değişimlendiği koşullarda; referans grubun örneklem büyüklüğü arttıkça GLK için Tip III hata oranlarının yükseldiği, GLR-LRT, GLR-Wald ve Lordif için elde edilen değerlerin ise

referans grubun örneklem büyüklüğünden etkilenmediği belirlenmiştir. Aynı zamanda araştırmada GLR-LRT ve GLR-Wald teknikleri için hesaplanan Tip III hata oranlarının her iki örneklem büyüklüğü için de liberal kritere göre kabul edilebilir olduğu tespit edilmiştir. Buna karşın GLK ve Lordif teknikleri için hesaplanan Tip III hata oranlarının kabul edilebilir aralıktan yüksek oldukları görülmesine rağmen Lordif'in performansının nispeten daha iyi olduğu saptanmıştır (bkz. EK 27.4). Böylece .70'in üzerinde Tip III hata oranlarına sahip olması nedeniyle, TB DMF'li maddelere ilişkin DMF türünün GLK tarafından yüksek oranda hatalı belirlenebildiği ortaya konulmuştur. Finch ve French (2008) tarafından incelenen TB DMF koşullarında, referans grubun örneklem büyüklüğü arttıkça SIBTEST ve MTKOO teknikleri için hesaplanan Tip III hata oranlarının yükselme eğiliminde oldukları, LR tekniği içinse elde edilen değerlerin referans grubun örneklem büyüklüğünden etkilenmediği açıklanmıştır. Bu bağlamda yapılan araştırmalarda MTK'ya dayalı hem ikili (MTKOO) hem de çoklu grup (GLK) teknikleri için hesaplanan Tip III hata oranlarının örneklem büyüklüğüne bağlı olarak yükseldikleri, KTK'ya dayalı LR ve GLR-LRT ile GLR-Wald tekniklerine ilişkin performansların ise örneklem büyüklüğündeki artıştan etkilenmedikleri gözlemlenmiştir.

Yalnızca ayırt edicilik parametresinin $[TBO(a) \text{ DMF}]$ değişimlendiği koşullarda; referans grubun örneklem büyüklüğü arttıkça GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının yükseldiği belirlenmiştir. Aynı zamanda araştırmada belirtilen teknikler için hesaplanan Tip III hata oranlarının her iki örneklem büyüklüğü için de kabul edilebilir aralığın son derece üzerinde yer aldıkları görülmesine rağmen GLK'nın performansının nispeten daha iyi olduğu saptanmıştır (bkz. EK 27.4). Böylece .50'nin üzerinde Tip III hata oranlarına sahip olması nedeniyle, $TBO(a) \text{ DMF}$ 'li maddelere ilişkin DMF türünün TB DMF'yi belirlemek amacıyla geliştirilen GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri tarafından yüksek oranda hatalı belirlenebildiği tespit edilmiştir. Finch ve French (2008) tarafından $TBO(a) \text{ DMF}$ türünün incelendiği koşullarda, referans grubun örneklem büyüklüğü arttıkça SIBTEST, MH, LR ve MTKOO teknikleri için hesaplanan Tip III hata oranlarının yükselme eğiliminde oldukları açıklanmıştır. Öte yandan MH ile LR tekniklerinin oldukça benzer, MTKOO tekniğinin ise bu iki teknikten biraz daha düşük Tip III hata oranına sahip olduğu belirtilmiştir. Bu bağlamda KTK ve MTK'ya dayalı tekniklerin farklı örneklem büyüklüklerindeki Tip III hata oranları açısından performanslarının ikili ve çoklu gruplarda son derece tutarlı olduğu sonucuna ulaşılmıştır.

Hem güçlük hem de ayırt edicilik parametrelerinin [TBO(*ab*) DMF] değişimlendiği koşullarda; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının kabul edilebilir aralığın oldukça üzerinde yer aldıkları ve elde edilen değerlerin referans grubun örneklem büyüklüğünden ciddi miktarda etkilenmedikleri tespit edilmiştir. Aynı zamanda belirtilen tekniklere ilişkin elde edilen değerlerin birbirlerine ve 1.00'a yakın olması nedeniyle TBO(*ab*) DMF'li maddelerin DMF türünün belirtilen beş çoklu grup tekniği tarafından çok yüksek oranda hatalı belirlenebildiği saptanmıştır (bkz. EK 27.4). Böylece TBO(*ab*) DMF türünde .80'in üzerinde Tip III hata oranlarına sahip olması nedeniyle, bu maddelere ilişkin DMF türünün TB DMF'yi belirlemek amacıyla geliştirilen GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri tarafından yüksek oranda hatalı belirlenebildiği ortaya konulmuştur. Alanyazında ikili gruplarda TBO(*ab*) DMF türü için Tip III hatanın hesaplandığı başka bir simülasyon çalışmasında da (Lopez Rivas, 2012) TB DMF'yi tespit etmek için kullanılabilen tekniklerin testteki TBO(*ab*) DMF'li maddeleri yanlışlıkla DMF'li olarak belirleme oranlarının yüksek olduğu belirlenmiş ve böylelikle bu araştırma kapsamında elde edilen bulguların desteklendiği sonucuna ulaşılmıştır.

Yapılan bu çalışmada çoklu grup teknikleri, referans grubun örneklem büyüklüğü ve DMF miktarının üçlü etkileşimi incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının, referans grubun örneklem büyüklüğü ve DMF miktarının etkileşimine bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .025$). Araştırmada DMF miktarı düşük olduğunda Lordif, DMF miktarı orta veya yüksek olduğunda ise GMH tekniklerinin en yüksek Tip III hata oranlarına sahip oldukları belirlenmiştir. Aynı zamanda DMF miktarı düşük, orta düzey veya yüksek olduğunda referans grubun örneklem büyüklüğü arttıkça beş çoklu grup tekniğine ilişkin elde edilen değerlerin yükseldiği ve her iki örneklem büyüklüğü için hesaplanan Tip III hata oranlarının liberal kritere göre kabul edilebilir aralığın oldukça üzerinde yer aldığı tespit edilmiştir (bkz. EK 27.5). Bu bağlamda DMF miktarı ve referans grubun örneklem büyüklüğü arttıkça GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerine ilişkin elde edilen Tip III hata oranlarının yükselme eğiliminde olduğu ve DMF miktarı orta veya yüksek iken beş çoklu grup tekniğinin .80'in üzerinde Tip III hata oranlarına sahip olması nedeniyle bu teknikler ile DMF türünün hatalı belirlenebildiği saptanmıştır. Finch ve French (2008) tarafından yapılan çalışmada da referans grubun örneklem büyüklüğü ve DMF miktarı birlikte arttıkça SIBTEST, MH, LR ve MTKOO teknikleri için hesaplanan Tip III hata oranlarının

yükselme eğiliminde oldukları açıklandığından, aynı amaca hizmet eden çalışmaların benzer sonuçlar ürettikleri ortaya konulmuştur. Yapılan bu simülasyon çalışmasında GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin farklı DMF türlerindeki performanslarının incelenmesi amacıyla veri üretilmesine rağmen yapılan analizlerde DMF türü dikkate alınmadığından, GMH haricindeki teknikler için bu sonuçlara dayalı olarak yapılacak yorumların yanıltıcı olabileceğini hatırlatmakta yarar vardır. Bu bilgilerden hareketle belirtilen tekniklerin performanslarının daha iyi anlaşılabilmesi amacıyla diğer faktörlerle olan dörtlü ve beşli etkileşimlerin incelenmesi gerekmektedir.

Araştırma kapsamında çoklu grup teknikleri, DMF türü ve DMF miktarının üçlü etkileşimi incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının, DMF türü ve DMF miktarının etkileşimine bağlı olarak manidar bir şekilde farklılaştığı sonucuna ulaşılmıştır ($\eta_p^2 = .243$). Yapılan analizler sonucunda, TB DMF türünde GLK için hesaplanan Tip III hata oranlarının DMF miktarlarına bağlı olarak yükseldiği görülmüş, GLR-LRT, GLR-Wald ve Lordif teknikleri için elde edilen değerlerin ise DMF miktarlarındaki artıştan etkilenmedikleri belirlenmiştir. Aynı zamanda GLR-LRT ve GLR-Wald teknikleri için bütün TB DMF miktarlarında hesaplanan Tip III hata oranlarının liberal kritere göre kabul edilebilir oldukları gözlemlenmiş, Lordif için elde edilen değerlerin ise kabul edilebilir aralıkta yer almadığı anlaşılmıştır. Buna karşın DMF miktarı düşük olsa dahi GLK için hesaplanan Tip III hata oranlarının liberal kriterden son derece yüksek olduğu tespit edilmiştir (bkz. EK 27.6). Elde edilen bu bulguları destekleyen Finch ve French (2008) tarafından yapılan iki gruplu çalışmada TB DMF miktarı arttıkça TBO(a) DMF'yi belirlemek amacıyla geliştirilen MTKOO için Tip III hata oranlarının yükseldiği, LR'nin performansının ise DMF miktarından etkilenmediği ortaya konulmuştur. Benzer şekilde Lopez Rivas'ın (2012) iki gruplu simülasyon çalışmasında da TB DMF miktarı artsa dahi TBO(a) DMF'yi belirlemek amacıyla geliştirilen KTK'ya dayalı teknikler için hesaplanan Tip III hata oranlarının ciddi miktarda farklılaşmadığı açıklanmıştır. Böylece gerçekte TB DMF'li maddeler için yapılan analizler sonucunda, GLR-LRT ve GLR-Wald teknikleri tarafından tespit edilen DMF türünün DMF miktarından bağımsız olarak doğru tanımlanabildiği sonucuna ulaşılmış ve bu sonucun alanyazındaki sınırlı sayıdaki ilgili araştırmalarla son derece tutarlı olduğu gözlemlenmiştir.

Yalnızca ayırt edicilik parametresinin [TBO(a) DMF] değiştiği koşullarda; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için DMF miktarı arttıkça belirtilen teknikler için hesaplanan Tip III hata oranlarının yükseldiği tespit edilmiştir.

Ayrıca GLK için elde edilen değerlerin diğer tekniklere kıyasla daha düşük olduğu görülmeye rağmen, bu DMF türünde beş çoklu grup tekniği için hesaplanan Tip III hata oranlarının kabul edilebilir sınır değerden oldukça yüksek olması nedeniyle DMF miktarı önemli olmaksızın DMF türünün hatalı belirlenebilme olasılığının oldukça yüksek olduğu saptanmıştır (bkz. EK 27.6). Finch ve French (2008) tarafından yapılan iki gruplu çalışmada da TBO(a) DMF miktarı arttıkça TB DMF'yi belirlemek amacıyla kullanılan SIBTEST, LR, MH ve MTKOO teknikleri için hesaplanan Tip III hata oranlarının yükseldiği belirtildiğinden, yapılan benzer çalışmalar sonucunda hem ikili hem de çoklu grup tekniklerinin birbirlerine paralel performans gösterdikleri belirlenmiştir.

Hem güçlük hem de ayırt edicilik parametrelerinin [TBO(ab) DMF] değiştiği koşullarda; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının birbirlerine yakın ve kabul edilebilir aralığın oldukça üzerinde yer aldıkları tespit edilmiştir. Ayrıca düşük DMF miktarında beş çoklu grup tekniği için Tip III hata oranlarının .75'in üzerinde hesaplandığı belirlenmişken DMF miktarı orta düzey veya yüksek olduğunda değerlerin 1.00'a yaklaştığı gözlemlenmiştir (bkz. EK 27.6). İkili grupta yapılan bir simülasyon çalışmasında da (Lopez Rivas 2012) TB DMF türü için geliştirilen tekniklerin testteki TBO(ab) DMF'li maddelerdeki Tip III hata oranlarının oldukça yüksek olduğu açıklanmıştır. Bu bulgulardan hareketle DMF miktarı arttıkça GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerine ilişkin elde edilen Tip III hata oranlarının yükselme eğiliminde olduğu ve DMF miktarı orta veya yüksek iken TB DMF'yi tespit etmek amacıyla geliştirilen çoklu grup teknikleri ile gerçekte TBO(ab) DMF'li maddelerin DMF türünün neredeyse tamamen (%100 olasılıkla) hatalı belirlendiği sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında çoklu grup teknikleri, referans grubun örneklem büyüklüğü, DMF türü ve DMF miktarının dörtlü etkileşimleri incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının, referans grubun örneklem büyüklüğü, DMF türü ve DMF miktarının etkileşimine bağlı olarak manidar bir şekilde farklılaştığı belirlenmiştir ($\eta_p^2 = .176$). Bu nedenle istatistiksel olarak manidar olan dörtlü etkileşim için betimsel istatistikler hesaplanmış ve hesaplanan Tip III hata oranları Tablo 30'da verilmiştir.

Tablo 30

Çoklu Grup Teknikleri İçin Hesaplanan Tip III Hata Oranlarına İlişkin Ortalama ve Standart Hatalar

Referans Örneklem Büyüklik	DMF Miktarı	DMF Türü	GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
			Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata
N _R =600	Düşük	b	.368	.007	.045	.002	.042	.002			.224	.004
		a	.099	.005	.336	.006	.332	.006	.342	.006	.356	.006
		ab	.667	.002	.678	.002	.676	.002	.668	.002	.823	.002
	Orta	b	.706	.007	.046	.001	.042	.001			.230	.003
		a	.176	.005	.523	.006	.519	.006	.533	.006	.533	.006
		ab	.919	.001	.947	.001	.947	.001	.943	.001	.991	.000
	Yüksek	b	.852	.004	.046	.001	.043	.001			.232	.002
		a	.235	.006	.647	.005	.644	.005	.662	.005	.646	.005
		ab	.970	.000	.985	.000	.985	.000	.983	.000	.999	.000
	Düşük	b	.915	.005	.048	.002	.048	.002			.227	.004
		a	.344	.010	.769	.006	.769	.006	.787	.006	.747	.006
		ab	.887	.002	.861	.002	.861	.002	.847	.003	.981	.001
N _R =1800	Orta	b	.996	.000	.048	.001	.047	.001			.228	.003
		a	.519	.007	.930	.003	.930	.003	.940	.002	.905	.003
		ab	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	Yüksek	b	1.000	.000	.047	.001	.046	.001			.231	.002
		a	.630	.006	.977	.001	.977	.001	.982	.001	.961	.001
		ab	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000

Tablo 30’da verilen Tip III hata oranları incelendiğinde; TBO(a) ve TBO(ab) DMF türlerinde çoklu grup teknikleri için hesaplanan Tip III hata oranlarının genellikle Bradley (1978) tarafından önerilen liberal kriterden (.075) yüksek olduğu saptanmış ve elde edilen değerlerin referans grubun örneklem büyüklüğü ile DMF miktarının birlikte artmasına bağlı olarak yükseldiği tespit edilmiştir. Öte yandan araştırma kapsamında dikkate alınan TB DMF türünde referans grubun örneklem büyüklüğü ve DMF miktarı birlikte arttıkça GLK için hesaplanan Tip III hata oranlarının yükseldiği, GLR-LRT, GLR-Wald ve Lordif tekniklerine ilişkin performansların ise ne referans grubun örneklem büyüklüğü ne de DMF miktarındaki artıştan etkilenmedikleri belirlenmiştir (bkz. EK 27.7). Yapılan analizlerde Lordif için saflaştırmanın otomatik olarak gerçekleştirilmesi nedeniyle (Choi vd., 2011, 2016; Crane, Gibbons vd., 2006; Fischer ve Karl, 2019; Hahn vd., 2014; Opariuc-Dan vd., 2017) bu teknik için elde edilen değerlerin referans grubun örneklem büyüklüğünden etkilenmemesinin kabul edilebilir olabileceği düşünülmüştür.

Çalışma kapsamında TB DMF türündeki bütün koşullar altında GLR-LRT ve GLR-Wald teknikleri için elde edilen değerlerin liberal kritere göre iyi kontrol edildiği belirlenmişken Lordif için benzer durumun söz konusu olmadığı gözlemlenmiştir (bkz. EK 27.7). Böylece çoklu grup teknikleri için hesaplanan Tip III hata oranlarının DMF

türü, DMF miktarı ve referans grubun örneklem büyüklüğüne bağlı olarak farklılaşabildiği ortaya konulmuştur. İki gruplu çalışmalarda da (Finch ve French, 2008; Lopez Rivas 2012) DMF türü, örneklem büyüklüğü ve DMF türünün önemli birer faktör oldukları açıklandığından, çoklu gruplar için elde edilen bulguların ilgili araştırmaların sonuçlarıyla genel anlamda örtüştüğü sonucuna ulaşılmıştır.

Araştırmada dikkate alınan DMF türü TBO(ab) iken DMF miktarının düşük olduğu durumda Lordif için hesaplanan Tip III hata oranlarının diğer tekniklerden daha yüksek olduğu, DMF miktarı orta veya yüksek olduğunda ise ele alınan çoklu grup tekniklerinin tamamının birbirlerine oldukça yakın performans ortaya koydukları tespit edilmiştir. Chun ve diğerlerinin (2016) çalışmalarında da Tip III hata oranlarının bazı koşullar altında son derece yüksek (.90 ve üzeri) olabildiği ifade edildiğinden, yapılan bu çalışmada çoklu grup teknikleri için 1.00'a yakın hesaplanan Tip III hata oranlarının Chun ve diğerlerinin (2016) elde ettikleri bulgularla desteklendiği anlaşılmıştır. Böylece TBO(ab) DMF türü için hem MTK'ya hem de KTK'ya dayalı olarak geliştirilen çoklu grup tekniklerinin (GLK, GLR-LRT, GLR-Wald, GMH ve Lordif) gerçekte var olmayan DMF türünü yüksek oranlarda hatalı belirleyebilmeleri nedeniyle uygulayıcıların bu durumu dikkate almalarının gerekli olduğu sonucuna ulaşılmıştır.

Araştırmada DMF türünün TB olduğu koşullarda, TBO DMF'yi belirlemek amacıyla geliştirilen GLR-LRT ve GLR-Wald'ın; DMF türü TBO(a) olduğunda ise TB DMF'yi belirlemek amacıyla geliştirilen GLR-LRT, GLR-Wald, GMH ve Lordif'in Tip III hata oranları açısından birbirlerine son derece yakın performans gösterdikleri belirlenmiştir. Alanyazındaki bazı çalışmalarda da GLR ve GMH tekniklerine ilişkin performansların (Tip I hata ve istatistiksel güç) birbirlerine çok benzer oldukları açıklandığından (Gao, 2019; Finch, 2016; Magis vd., 2011; Rollins III, 2018; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) yapılan bu araştırma kapsamında GLR-LRT, GLR-Wald ve GMH teknikleri için hesaplanan Tip III hata oranlarının da birbirlerine yakın olmasının son derece doğal olduğu kabul edilmiştir. Öte yandan GLK ve Lordif tekniklerinin MTK'ya dayalı geliştirildikleri belirtilmesine rağmen (Choi vd., 2011; Crane, Gibbons vd., 2006; Kim vd., 1995) bu tekniklere ilişkin Tip III hata oranlarının TB ve TBO(a) DMF türlerinde birbirlerinden farklılaştığı ve Lordif'in GLK'dan ziyade KTK'ya dayalı tekniklerle benzer performans gösterdiği belirlenmiştir. LR temelinde geliştirilen Lordif tekniğinde eşleştirme kriteri olarak MTK'ya dayalı yetenek kestirimlerinin kullanıldığı bilinmesine rağmen, bu tekniğin genel olarak KTK'ya dayalı (GLR-LRT, GLR-Wald ve GMH) teknikler ile paralel performans göstermesi ilginç bir

bulgu olarak kabul edilmiştir. Aynı zamanda bu simülasyon çalışmasında saflaştırma işleminin otomatik olarak gerçekleştirildiği Lordif'in $TBO(a)$ ve $TBO(ab)$ DMF türlerindeki Tip III hata oranlarının diğer tekniklerden daha yüksek olduğunun belirlenmesi nedeniyle saflaştırma işleminin bu iki DMF türünde Lordif'in performansına olumlu yönde bir katkısının söz konusu olmadığı anlaşılmıştır.

Araştırma kapsamında yapılan analizler ışığında elde edilen bulgulara göre; gerçekte DMF'li olan maddelerin DMF türünün bazı teknikler tarafından hatalı belirlenme olasılığının (Tip III hata) genel olarak yüksek olduğu belirlenmiştir. Hesaplanan yüksek Tip III hata oranları; gerçekte DMF'li olan maddeler için tespit edilen DMF türünün analizin gerçekleştirildiği istatistiksel teknikler tarafından hatalı tanımlandığını gösterdiğinden (Chun vd., 2016; Finch ve French, 2008; Lopez Rivas, 2012) yapılan bu simülasyon çalışmasında $TBO(a)$ ve $TBO(ab)$ DMF türlerinde beş çoklu grup tekniğinin, TB DMF türünde ise yalnızca GLK'nın, incelenen maddelerin DMF türünü belirlemede ciddi sorun yaşadığı ortaya konulmuştur. MTK'ya dayalı modellerden 2PLM'de güçlük ve ayırt edicilik parametrelerinin dikkate alınması nedeniyle, bu modelin güçlük parametresi açısından 1PLM'yi kapsadığı ve 2PLM'ye dayalı GLK tekniğinin [$TBO(a)$], 1PLM'ye dayalı tekniğe (TB) kıyasla daha yüksek Tip III hata oranları ortaya koymasının doğal olduğu kabul edilmiştir. Dolayısıyla $TBO(a)$ DMF'yi belirlemek amacıyla geliştirilen GLK'nın (2PLM), TB DMF'yi belirlemek amacıyla geliştirilen maddeleri yüksek oranda DMF'li belirlemesinin anlaşılabilir olduğu sonucuna ulaşılmıştır. Böylece Tip III hata açısından elde edilen bu bulgular; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerine ilişkin performansların DMF'nin hangi madde parametresinden kaynaklı olduğuna bağlı olarak değişebileceğini göstermiştir.

Alanyazında DMF'nin kaynaklandığı madde parametresine göre genel anlamda TB ve TBO olmak üzere iki farklı DMF türü tanımlanmış ve analiz sonuçlarının doğru yorumlanabilmesi için DMF türü hakkında da bilgi sahibi olunması gerektiği açıklanmıştır (Mellenbergh, 1982; Penfield ve Camilli, 2007). Kuramsal olarak bir DMF belirleme tekniği eğer TB DMF'li maddeleri belirlemek amacıyla geliştirildiyse, bu tekniğin TBO DMF'ye duyarsız olduğu kabul edilmiş ve TBO DMF'li maddeleri DMF'li olarak işaretleyememesi gerektiği belirtilmiştir. Benzer şekilde TBO DMF'li maddeler için geliştirilen tekniklerin ise TB DMF'li maddeleri DMF'li olarak belirleyememesinin son derece önemli olduğu ifade edilmiştir (Camilli, 2006).

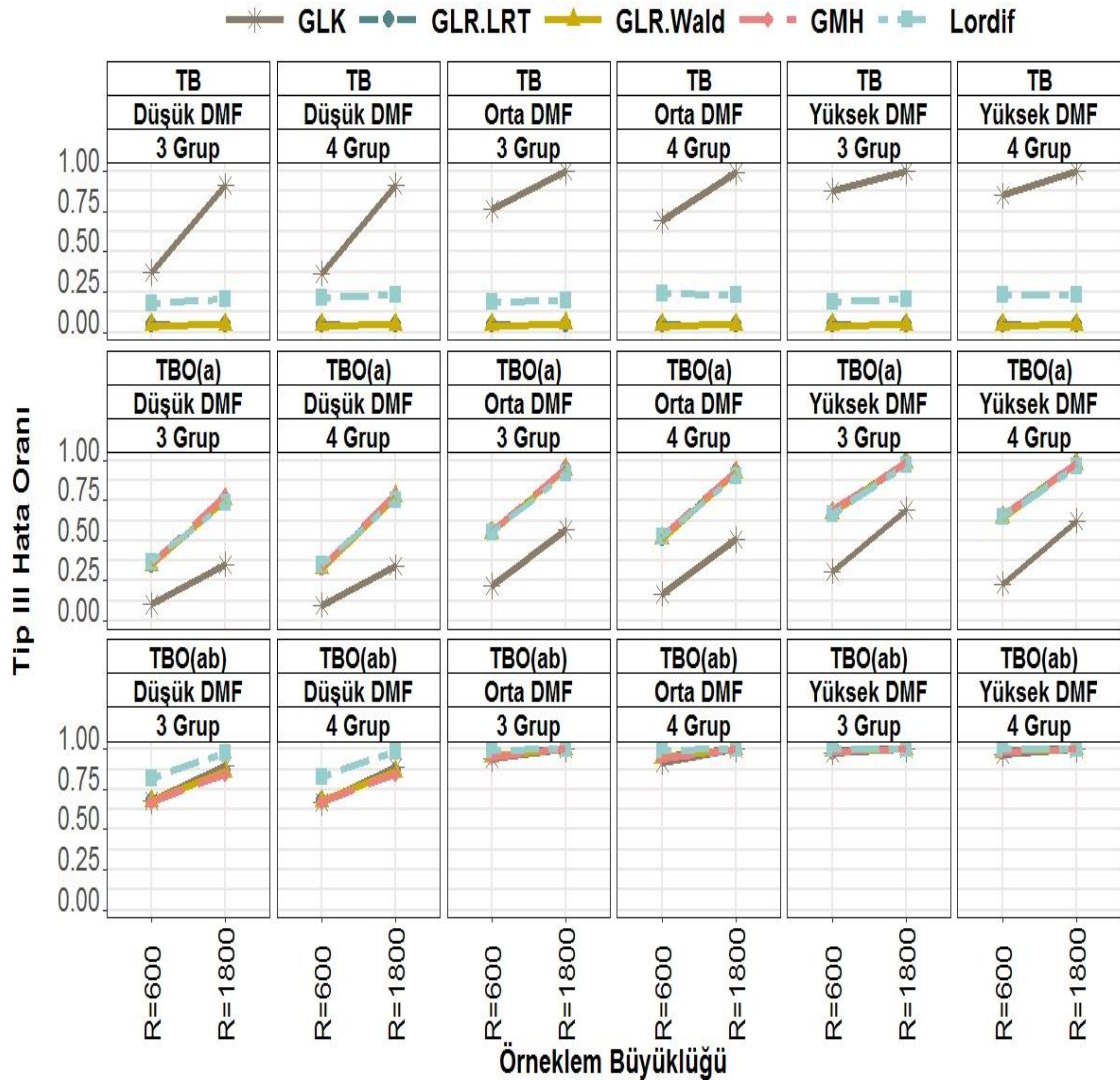
Alanyazında GMH tekniğinin TB DMF'yi belirlemek amacıyla geliştirildiği açıklanmasına rağmen (Finch vd., 2016; Holland ve Thayer, 1988; Swaminathan ve

Rogers, 1990), Dell-Ross (2021) yaptığı çalışmasında yalnızca ayırt edicilik ve hem güçlük hem de ayırt edicilik parametrelerinden kaynaklı DMF'li maddeleri GMH tekniği ile analiz etmiştir. Araştırmacı GMH'in TBO(a) ve TBO(ab) DMF türlerindeki performanslarını istatistiksel güç olarak yorumlasa da, aslında ulaşılan sonuçların Tip III hataya karşılık geldiği ifade edilebilmektedir. Nitekim GMH tekniğine ilişkin performansların TBO(a) DMF türünde daha fazla 0.00'a yakın, TBO(ab) DMF türünde ise 1.00'a yakın olması, Dell-Ross'un (2021) çalışmasında raporlanan istatistiksel güç sonuçları ile yapılan bu araştırmadaki Tip III hata sonuçlarının tutarlı olduğunu göstermiştir. GMH tekniği için her iki araştırmadan elde edilen bu sonuçlar; istatistiksel tekniklerin hangi DMF türünü belirlemek amacıyla geliştirildiyse, belirtilen DMF türü için kullanılmalarının uygun olduğunu, aksi taktirde elde edilen sonuçların hatalı yorumlara yol açabileceğini ortaya koymuştur.

Yapılan kuramsal çalışmalar ile TB ve TBO DMF türünü tespit edebilen çok sayıda istatistiksel teknik geliştirilmesine rağmen DMF türünün doğru belirlenip belirlenmediği konusu alanyazında ilk defa Finch ve French (2008) tarafından yapılan araştırmada gündeme getirilmiş ve gerçekte var olmayan DMF türünün tespit edilmesi durumunda yapılacak yorumlar için dikkatli olunması gerektiği vurgulanmıştır. Chun ve diğerlerinin (2016) çalışmalarında da çok yüksek Tip III hata oranlarına sahip olan tekniklerin güvenilir sonuçlar ortaya koyamadıkları bildirilerek, DMF'nin gerçek kaynağının doğru bir şekilde tanımlanıp tanımlanmadığının incelenmesi gerektiğine işaret edilmiştir.

Yapılan bu simülasyon çalışmasında DMF'li maddelere ilişkin DMF türünün hatalı bir şekilde işaretlenme olasılığının liberal kritere göre oldukça yüksek olması nedeniyle çoklu grup tekniklerine ilişkin performansların dikkatli yorumlanması gerektiği anlaşılmıştır. Böylece bu araştırmadan elde edilen bulgular sonucunda gerçek uygulamalarda hem TB hem de TBO DMF'yi belirlemek amacıyla geliştirilen çoklu grup tekniklerinin birlikte kullanılması gerektiği ve yapılan analizlerden sonra maddelere ilişkin DMF türünün doğru belirlenebilmesi amacıyla MKE'lere başvurulmasının önemli olduğu belirlenmiştir. Finch ve French (2008) ile Lopez Rivas (2012) tarafından yapılan çalışmalarda da gerçek test uygulamalarında DMF belirlendiği taktirde DMF'nin hangi parametreden kaynaklandığının tespit edilebilmesi için grafiksel gösterimlerden yararlanılabileceği ifade edildiğinden, belirtilen araştırmalarda yapılan açıklamalar ile elde edilen bulguların bu simülasyon çalışması ile desteklendiği sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında son olarak çoklu grup teknikleri, referans grubun örneklem büyüklüğü, DMF türü, DMF miktarı ve grup sayısının beşli etkileşimleri incelenmiştir. Dört faktöre grup sayısının da eklenmesi sonucunda oluşan beşli etkileşimin istatistiksel olarak manidar olmaması ($F_{(14,23088)} = 0.867, p > .05, \eta_p^2 = .001$) nedeniyle, grup sayısı farklılaşsa dahi GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarında önemli bir değişikliğin olmadığı sonucuna ulaşılmıştır (bkz. Tablo 29). Araştırma kapsamında istatistiksel olarak manidar olan dörtlü etkileşimin yanı sıra manidarlığın söz konusu olmadığı beşli etkileşim için de Tip III hata oranları hesaplanarak, çizilen grafikler Şekil 3'te verilmiştir.



Şekil 3. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü, DMF Türü, DMF Miktarı ve Grup Sayısı Etkileşimi İçin Tip III Hata Oranları

Şekil 3'teki sonuçlara göre; TB DMF türünde GLR-LRT ve GLR-Wald teknikleri için hesaplanan Tip III hata oranlarının DMF miktarı ile referans grubun örneklem büyüklüğündeki artıştan etkilenmediği ve belirtilen tekniklere ilişkin performansların üç ve dört grup için son derece yakın olduğu tespit edilmiştir. Aynı zamanda araştırmada Lordif için elde edilen değerlerin de DMF miktarı ile referans grubun örneklem büyüklüğünün farklılaşmasından etkilenmemesine karşın bu tekniğe ilişkin Tip III hata oranlarının grup sayısı arttıkça nispeten yükseldiği görülmüştür. Son olarak GLK tekniği için ise özellikle $N_R = 600$ ve DMF miktarı orta veya yüksek iken elde edilen Tip III hata oranlarının, grup sayısındaki artışa bağlı olarak bir miktar azaldığı, ancak hem GLK hem de Lordif'in performanslarındaki bu farklılaşmanın istatistiksel olarak manidar olmadığı ($p > .05$, $\eta_p^2 = .001$) anlaşılmıştır (bkz. Tablo 29 ve Şekil 3).

Araştırma kapsamında dikkate alınan TBO(a) DMF türünde GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının genellikle DMF miktarı ve referans grubun örneklem büyüklüğündeki artışa bağlı olarak yükseldiği belirlenmesine rağmen, belirtilen tekniklere ilişkin performansların üç ve dört grup için son derece benzer olduğu tespit edilmiştir. Öte yandan DMF miktarının orta veya yüksek olduğu koşullarda grup sayısı arttıkça, her iki örneklem büyüklüğü için de GLK tekniğine ilişkin Tip III hata oranlarının bir miktar azaldığı görülmüş ve grup sayısından kaynaklı oluşan bu farklılaşmanın istatistiksel olarak manidar olmadığı anlaşılmıştır. TBO(ab) DMF türünde ise GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının genel olarak birbirlerine ve 1.00'a yakın olmasının yanı sıra GLK haricindeki tekniklere ilişkin performansların üç ve dört grup için son derece benzer olduğu gözlemlenmiştir. GLK tekniği için $N_R = 600$ ve DMF miktarının orta düzey olduğu koşullarda grup sayısı arttıkça, Tip III hata oranlarının bir miktar azaldığı belirlenmiş, ancak grup sayısına bağlı oluşan bu farklılaşmanın istatistiksel olarak manidar olmadığı ($p > .05$, $\eta_p^2 = .001$) görülmüştür (bkz. Tablo 29 ve Şekil 3). Böylece araştırma kapsamında yapılan analizler ışığında; DMF türü, DMF miktarı ve referans grubun örneklem büyüklüğü faktörleri sabit iken grup sayısı farklılaşsa dahi özellikle GLR-LRT, GLR-Wald ve GMH teknikleri için hesaplanan Tip III hata oranlarında hiçbir değişikliğin söz konusu olmadığı belirlenmiştir. Buna karşın dikkate alınan üç DMF türünde de GLK ve Lordif teknikleri için elde edilen değerlerin grup sayısına bağlı olarak bir miktar farklılaştığı gözlemlenmiş, ancak bu farklılaşmanın önemli olmadığı sonucuna ulaşılmıştır.

Yapılan bu simülasyon çalışmasında Tip III hata oranları açısından elde edilen bulgular özetlendiğinde; çoklu grup tekniklerine ilişkin performansların grup sayısına bağlı olarak ciddi bir değişiklik göstermediği tespit edilmesine rağmen DMF türü, DMF miktarı ve referans grubun örneklem büyüklüğü faktörlerinin Tip III hatayı etkileyen önemli faktörler oldukları ortaya konulmuştur. Elde edilen bu sonuçlara ilaveten TB DMF türünde MTK'ya dayalı teknikler (GLK, Lordif) için hesaplanan Tip III hata oranlarının KTK'ya dayalı tekniklerden (GLR-LRT ve GLR-Wald) daha yüksek oldukları, TBO(*ab*) DMF türünde ise KTK ve MTK'ya dayalı tekniklerin nispeten benzer performans gösterdikleri bulunmuştur. Öte yandan TBO(*a*) DMF türünde KTK'ya dayalı tekniklerin birbirlerine yakın performans gösterdikleri belirlenmişken, kuramsal temelleri KTK'ya dayalı olan ancak yetenek kestirimlerinin MTK'ya dayalı yapılması nedeniyle MTK çerçevesinde dikkate alınan Lordif'in performansının ise GLR-LRT, GLR-Wald ve GMH teknikleri ile benzer olduğu gözlemlenmiştir. Böylece yapılan bu simülasyon çalışması neticesinde elde edilen bulgular göz önüne alındığında, çoklu grup tekniklerine ilişkin kuramsal özelliklerin de Tip III hata oranlarını etkileyen önemli bir faktör olduğu sonucuna ulaşılmıştır (bkz. Tablo 30 ve Şekil 3).

Çoklu grup tekniklerinin performanslarının karşılaştırıldığı bu simülasyon çalışmasında; öncelikle GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin farklı koşullardaki performansları Tip I hata, istatistiksel güç ve Tip III hata açısından ayrı ayrı incelenerek raporlanmıştır. DMF belirleme çalışmalarında tekniklere ilişkin istatistiksel gücün (1-Tip II hata) yüksek, Tip I ve Tip III hatanın ise düşük olması istenilen durumdur (Finch, 2016; Lopez Rivas, 2012; Schauburger ve Tutz, 2016; Svetina ve Rutkowski, 2014). Başka bir ifadeyle yüksek Tip I, Tip II ve Tip III hata oranları istatistiksel tekniklerin düşük performans gösterdikleri anlamına geldiğinden karşılaştırılan teknikler arasından istatistiksel gücü yüksek (dolayısıyla Tip II hatası düşük), Tip I ve Tip III hatası ise düşük olan tekniğin tercih edilmesi gerekmektedir. Bu nedenle araştırma kapsamında çoklu grup teknikleri hakkında genel bir yargıya varılabilmesi amacıyla Tip I hata, istatistiksel güç ve Tip III hata oranları için elde edilen değerler birlikte raporlanarak sonuçlar Tablo 31'de sunulmuştur. Böylece belirtilen tekniklerin üç değerlendirme ölçütündeki performansları incelenip optimum performans gösteren tekniğin belirlenmesi ve gerçek test uygulamalarındaki hangi koşullarda hangi tekniğin kullanılması gerektiğine yönelik önerilerde bulunulması amaçlanmıştır.

Tablo 31

Çoklu Grup Teknikleri İçin Hesaplanan Tip I Hata, İstatistiksel Güç ve Tip III Hata Oranlarına İlişkin Ortalamalar

Referans Örneklem Büyüklik	DMF Miktarı	Teknik	Tip I Hata			İstatistiksel Güç			Tip III Hata			
			DMF Yok	TB	TBO (a)	TBO (ab)	TB	TBO (a)	TBO (ab)	TB	TBO (a)	TBO (ab)
N _R =600	DMF Yok	GLK	.026									
		LRT	.046									
		Wald	.045									
		GMH	.049									
		Lordif	.059									
	Düşük DMF	GLK		.035	.027	.045	.505	.415	.924	.368	.099	.667
		LRT		.085	.045	.045	.700	.392	.333	.045	.336	.678
		Wald		.084	.042	.042	.698	.379	.320	.042	.332	.676
		GMH		.085			.691				.342	.668
		Lordif		.067	.064	.085	.794	.520	.676	.224	.356	.823
	Orta DMF	GLK		.057	.027	.060	.805	.653	.956	.706	.176	.919
		LRT		.123	.045	.045	.920	.630	.552	.046	.523	.947
		Wald		.121	.042	.042	.918	.612	.529	.042	.519	.947
		GMH		.123			.916				.533	.943
		Lordif		.065	.065	.100	.970	.771	.837	.230	.533	.991
	Yüksek DMF	GLK		.074	.026	.074	.914	.806	.973	.852	.235	.970
		LRT		.152	.045	.045	.975	.774	.701	.046	.647	.985
		Wald		.150	.042	.042	.974	.764	.686	.043	.644	.985
		GMH		.153			.973				.662	.983
		Lordif		.066	.067	.113	.994	.886	.914	.232	.646	.999
N _R =1800	DMF Yok	GLK	.026									
		LRT	.046									
		Wald	.045									
		GMH	.049									
		Lordif	.059									
	Düşük DMF	GLK		.081	.027	.089	.948	.915	.998	.915	.344	.887
		LRT		.164	.046	.047	.986	.850	.771	.048	.769	.861
		Wald		.164	.045	.047	.986	.849	.769	.048	.769	.861
		GMH		.169			.984				.787	.847
		Lordif		.065	.067	.128	.997	.937	.951	.227	.747	.981
	Orta DMF	GLK		.169	.027	.156	.998	.990	1.000	.996	.519	.999
		LRT		.293	.049	.047	1.000	.976	.952	.048	.930	1.000
		Wald		.292	.048	.046	1.000	.975	.950	.047	.930	1.000
		GMH		.301			1.000				.940	1.000
		Lordif		.067	.073	.162	1.000	.995	.994	.228	.905	1.000
	Yüksek DMF	GLK		.247	.027	.219	1.000	.999	1.000	1.000	.630	1.000
		LRT		.390	.050	.046	1.000	.995	.987	.047	.977	1.000
		Wald		.389	.050	.045	1.000	.995	.987	.046	.977	1.000
		GMH		.400			1.000				.982	1.000
		Lordif		.068	.079	.182	1.000	.999	.999	.231	.961	1.000

Tablo 31’de referans grubun örneklem büyüklüğü ($N_R = 600$, $N_R = 1800$), DMF miktarı (DMF Yok, Düşük DMF, Orta DMF, Yüksek DMF) ve DMF türü [DMF Yok, TB, TBO(a), TBO(ab)] koşullarında, çoklu grup teknikleri (GLK, GLR-LRT, GLR-Wald, GMH ve Lordif) için hesaplanan Tip I hata, istatistiksel güç ve Tip III hata oranları birlikte sunulmuştur. Wang ve Yeh (2003) tarafından da önerildiği üzere, Tip I hata hem hiçbir maddenin DMF içermediği durumda hem de testte DMF’li madde yer aldığı

durumda hesaplanmıştır. DMF içermeyen durumdaki Tip I hata oranı, tekniklerin etkililiğinin değerlendirilmesine yardımcı olan temel durumdur ve simülasyon çalışmalarında öncelikle temel Tip I hata oranlarının hesaplanması gerektiği araştırmacılar tarafından belirtilmiştir (Hidalgo vd., 2014; Kim, 2010). Maddelerin DMF (DMF miktarı düşük, orta ve yüksek) içerdiği koşullarda hesaplanan Tip I hata oranları ise saflaştırma yapılmadığı durumda GLK, GLR-LRT, GLR-Wald ve GMH tekniklerinin performanslarının incelenmesine olanak sağlamıştır. Öte yandan istatistiksel güç ve Tip III hatanın ise amaçları gereği yalnızca testte DMF’li madde bulunması durumunda hesaplanması gerektiği ifade edildiğinden (Lopez Rivas, 2012) bu çalışma kapsamında DMF içeren maddeler dikkate alınarak istatistiksel güç ve Tip III hata oranları hesaplanmıştır.

Alanyazındaki araştırmalarda gerçekte DMF’li olmayan maddelerin yanlışlıkla DMF’li olarak tespit edilebilmesinin mümkün olması nedeniyle Tip I hata oranı yüksek olan DMF belirleme tekniklerinin kullanılması durumunda hatalı yorumların söz konusu olabileceği ifade edilmiştir (Chun, 2014; Finch, 2016; Kim, 2010). Bu bağlamda yapılan bu simülasyon çalışmasında da öncelikle Tip I hata oranları hesaplanmış ve Tip I hatanın; testte DMF’li madde bulunup bulunmamasından, DMF türünden, DMF miktarından ve referans grubun örneklem büyüklüğünden etkilenebildiği belirlenmiştir. Elde edilen sonuçlara göre testte DMF’li madde bulunmadığı durumda GLK tekniğinin diğerlerinden daha tutucu (conservative) Tip I hata oranına sahip olduğu tespit edilmiş (Langer, 2008), diğer tekniklere ilişkin Tip I hata oranlarının ise nominal değere yakın olduğu (Penfield, 2001; Rollins III, 2018) görülmüştür. Ayrıca temel Tip I hata oranlarının referans grubun örneklem büyüklüğünden etkilenmediği (Battauz, 2019; Chun, 2014; Chun vd., 2016; Finch, 2016; Finch ve French, 2007; Kopf vd., 2015b; Li ve Zumbo, 2009; Penfield, 2001; Stark vd., 2006) de gözlemlenmiştir. Böylece gerçek test uygulamalarındaki testlerde gerçekte DMF’li madde yok ise GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri kullanılarak yapılacak değerlendirmelerin geçerli olacağı ortaya konulmuştur.

Testteki maddeler DMF içerdiğinde, toplam puan veya kestirilen yeteneğin DMF’li maddelerin etkisine bağlı olarak hatalı belirlenebildiği ifade edilmiştir (Kopf vd., 2015; Magis vd., 2010; Woods, 2009). Alanyazıdaki çok sayıda araştırmada da test TB DMF’li madde içerdiğinde hesaplanan Tip I hata oranlarının temel Tip I hata oranlarından daha yüksek olduğu belirtildiğinden (Camilli ve Shepard, 1994; Carroll, 2015; Meade ve Wright, 2012; Wang, 2004; Woods, 2009) dört maddenin TB DMF içerdiği bu çalışma kapsamında da eşleştirme kriterindeki kirlenme nedeniyle temel Tip I hata oranlarından

yüksek Tip I hata oranları hesaplanmasının doğal olduğu anlaşılmıştır. Benzer bulguya Rollins III'in (2018) çalışmasında da ulaşılmış ve testin TB DMF'li madde içermesi durumunda GLR ve GMH teknikleri için hesaplanan Tip I hata oranlarının ciddi miktarda yükseldiği belirtilmiştir. TB DMF'nin dikkate alındığı başka simülasyon çalışmalarında da KTK'ya dayalı tekniklerin Tip I hata oranlarının testteki DMF'li maddelerden etkilendiğinin açıklanması nedeniyle (Finch ve French, 2008; Kim, 2010; Magis ve Boeck, 2012; Scott, 2014) testin TB DMF'li madde içermesi durumunda saflaştırma yapılmazsa eğer Tip I hatanın yükselebileceği tespit edilmiştir. Öte yandan saflaştırma işleminin gerçekleştirildiği Tutz ve Berger (2016) tarafından yapılan araştırmada GLK, GLR ve GMH teknikleri için hesaplanan temel Tip I hata oranları ile Tip I hata oranlarının birbirlerine son derece yakın oldukları bulunmuştur. Bu bulgulardan hareketle TB DMF türündeki maddelerin yer aldığı gerçek test uygulamalarında saflaştırma yapılmadığı takdirde çoklu grup tekniklerinin Tip I hata yapma olasılığının beklendiği şekilde yüksek olması nedeniyle DMF belirleme çalışmalarında saflaştırma yapılmasının önemli olduğu sonucuna ulaşılmıştır.

Araştırmacılar tarafından DMF belirleme çalışmalarında istatistiksel tekniklerin performanslarını etkileyen önemli faktörlerden birinin de DMF türü olduğu ifade edilmiştir (Seybert ve Stark, 2012). İdeal olarak, bir tekniğin farklı DMF türlerine karşı olabildiğince sağlam olması (Gamerman vd., 2018) ve hem TB hem de TBO DMF türünü belirlemek amacıyla geliştirilen versiyonlarının benzer performans göstermesi gerekmektedir (Gao, 2019). Buna rağmen Tip I hata oranlarının testteki maddelerin DMF türünden etkilenebildiğinin açıklanması nedeniyle yapılan bu simülasyon çalışmasında DMF türü açısından da incelemeler yapılmış ve GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin Tip I hata açısından performanslarının TB, TBO(*a*) ve TBO(*ab*) DMF türlerinde farklılaştığı belirlenmiştir. Jin ve diğerlerinin (2015) çalışmalarında da test DMF'li madde içerdiğinde GLR tekniği için TB DMF türünde hesaplanan Tip I hata oranının TBO DMF türünden yüksek olduğu açıklandığından, bu araştırmadan elde edilen sonuçların desteklendiği anlaşılmıştır. Elde edilen bulgular birlikte değerlendirildiğinde; bazı tekniklerin farklı DMF türlerine daha duyarlı olduğu ve Tip I hata oranlarının hesaplandığı bütün koşullarda diğer tekniklerden daha iyi performans gösteren bir çoklu grup tekniğinin var olmadığı saptanmıştır. Buna karşın saflaştırmanın otomatik olarak gerçekleştirildiği Lordif için Tip I hata oranlarının diğer tekniklerden nispeten düşük olması nedeniyle gerçek test uygulamalarında ek bir saflaştırma işlemi yapılmayacaksa, Lordif'in kullanılmasının daha avantajlı olacağı sonucuna ulaşılmıştır.

Araştırma kapsamında Tip I hatanın yanı sıra DMF içeren dört madde dikkate alınarak istatistiksel güç hesaplanmıştır. Alanyazındaki çok sayıda araştırmada DMF türü önemli olmaksızın DMF miktarı ve örneklem büyüklüğünün istatistiksel gücü etkileyen önemli faktörler oldukları (Carroll, 2015; Cuevas ve Cervantes, 2012; French ve Maller, 2007; Kim, 2010; Lopez Rivas, 2012; Penfield, 2001; Socha vd., 2015; Wood, 2011; Yu vd., 2008) ve DMF miktarı ile örneklem büyüklüğü arttıkça istatistiksel gücün yükselmesinin beklendiği ifade edilmiştir. Bu araştırma kapsamında da dikkate alınan üç DMF türü için referans grubun örneklem büyüklüğü ile DMF miktarı arttıkça çoklu grup teknikleri için hesaplanan istatistiksel güç oranlarının yükseldiği belirlendiğinden, elde edilen sonuçların alanyazınla örtüştüğü görülmüştür. Böylece gerçek test uygulamalarında örneklemin büyük ve/veya DMF miktarının yüksek olması durumunda, çoklu grup tekniklerinin gerçekte DMF’li olan maddeleri daha yüksek oranda DMF’li olarak işaretleyebildikleri sonucuna ulaşılmıştır.

DMF ile ilgili simülasyon çalışmalarında istatistiksel tekniklerin farklı örneklem büyüklüklerindeki performanslarının tespit edilip, en uygun istatistiksel tekniğin seçilmesinin son derece önemli olması nedeniyle örneklem büyüklüğüne ayrı bir önem verilmesi gerektiği belirtilmiştir. Çünkü gerçek test uygulamalarında maddelerin DMF’li olup olmadığı ve madde DMF’li ise bu maddenin DMF türü ile DMF miktarı analizler gerçekleştirilmeden önce belirlenememesine rağmen, veri setindeki örneklem büyüklüğünün analizlerin öncesinde bilinebilmesi nedeniyle örneklem büyüklüğüne uygun istatistiksel tekniğin kullanılabilmesi mümkün olmaktadır (Muñiz vd., 2001). Örneklem büyüklüğünün istatistiksel tekniklerin performansları üzerindeki etkisinin incelendiği çalışmalarda; DMF belirleme tekniklerinin örneklem büyüklüğünden büyük ölçüde etkilendiği belirtilmiş ve küçük örneklemelerde gerçekten DMF’li olan maddelerin doğru bir şekilde belirlenemediği açıklanırken, büyük örneklemelerde DMF miktarı düşük olan maddelerin dahi DMF’li olarak işaretlenebildiği vurgulanmıştır (Cuevas ve Cervantes, 2012; Kim, 2010; Rockoff, 2018; Rogers ve Swaminathan, 1993; Wood, 2011; Wu vd., 2016). Magis ve diğerlerinin (2011) çalışmalarında büyük örneklemelerde parametre kestirimlerinin iyileşmesi sonucunda istatistiksel tekniklerin DMF’li maddeleri doğru belirleme olasılığının artacağı belirtilmiştir. Bazı araştırmalarda ise örneklemin büyük olduğu koşullarda DMF belirleme tekniklerinin performansları arasındaki farklılıkların çok belirgin olmaması nedeniyle DMF belirleme tekniklerinin kararlı sonuçlar ortaya koyabilmesi için gerekli olan örneklem büyüklüğünün belirlenmesinin önemine dikkat çekilmiştir (Rollins III, 2018; Zieky, 1993). Yapılan bu simülasyon

çalışmasının sonuçları da DMF miktarı önemli olmaksızın büyük örneklemelerde ($N_R=1800$) bütün teknikler için hesaplanan istatistiksel güç oranlarının yüksek (yani, %80 ve üzerindeki güç oranları) olduğunu ortaya koymuştur. Buna karşın $N_R = 600$ iken, yani örneklem küçük olsa dahi, DMF miktarının yüksek olması durumunda bazı teknikler için yüksek güç oranları elde edilmesine rağmen, TBO(*ab*) DMF türünde GLR-LRT ve GLR-Wald'ın performanslarının umut verici olmadığı gözlemlenmiştir. Bu bilgiler doğrultusunda testte yer alan DMF'li maddelerin bazı çoklu grup teknikleri tarafından doğru belirlenebilmesi için geniş örneklem büyüklüğünün gerekli olduğu anlaşıldığından, DMF analizleri öncesinde kullanılacak tekniğe karar verilirken veri setindeki örneklem büyüklüğünün dikkate alınması gerektiği sonucuna ulaşılmıştır.

Çoklu grup tekniklerinin farklı DMF türlerindeki performansları incelendiğinde; GLK, GLR-LRT, GLR-Wald ve Lordif için hesaplanan istatistiksel güç oranlarının TB, TBO(*a*) ve TBO(*ab*) DMF türlerinde farklılaştığı görülmüştür. Oshima ve diğerlerinin (2015) yapmış oldukları çalışmada MTK'ya dayalı olan MG-NCDIF tekniği için TBO(*ab*) DMF türünde hesaplanan istatistiksel güç oranlarının hem TB hem de TBO(*a*) DMF türlerine göre daha yüksek olduğu tespit edilmiştir. Bu araştırma kapsamında da GLK tekniğinin TBO(*ab*) DMF türündeki performansının diğer iki DMF türünden daha yüksek olması nedeniyle MTK'ya dayalı tekniklerin birbirlerine paralel sonuçlar ortaya koydukları tespit edilmiştir. Bununla birlikte Lordif'in farklı DMF türlerindeki performanslarının değişiklik göstermesi, yapılan saflaştırmanın etkisinin DMF'nin kaynaklandığı madde parametresine bağlı olarak farklılaştığını göstermiştir. Elde edilen bulgular birlikte değerlendirildiğinde; istatistiksel güç oranlarının hesaplandığı bütün koşullarda diğer tekniklerden daha iyi performans gösteren bir çoklu grup tekniğinin var olmadığı ve bazı tekniklerin farklı DMF türlerine daha duyarlı olduğu saptanmıştır. Buna karşın Lordif için hesaplanan istatistiksel güç oranlarının diğer tekniklerden nispeten daha yüksek olması nedeniyle, gerçek test uygulamalarında ek bir saflaştırma işlemi yapılmayacaksa, Lordif'in kullanılmasının daha avantajlı olacağı belirlenmiştir.

İstatistiksel güç oranları gerçekte DMF'li olan maddelerin çoklu grup teknikleri tarafından DMF'li olarak işaretlenmesi ile birlikte DMF kaynağının da doğru belirlendiği anlamına gelmektedir. Başka bir ifadeyle bir maddenin DMF'li olarak işaretlenmesinin yanı sıra bu madde için belirlenen DMF türünün de doğru olması, istatistiksel tekniğin güçlü olduğunu göstermektedir. DMF'li olarak belirlenen maddelerin gerçek DMF türü ile analizler sonucunda tespit edilen DMF türünün uyuşmaması durumunda ise Tip III hatanın söz konusu olduğu ifade edilmektedir. Bu açıklamalardan hareketle gerçek test

uygulamalarındaki DMF belirleme sürecinin iki aşamadan oluşması gerektiği ve testteki DMF'li maddelerin tespit edilmesinden sonra DMF türünün de belirlenmesinin önemli olduğu anlaşılmaktadır. Trzpiot (2015) ile Chun ve diğerleri (2016) tarafından yapılan çalışmalarda Tip III hata kontrol edilemediğinde istatistiksel gücün yanıltıcı olabileceği belirtilmiş ve araştırmacıların yüksek Tip III hata oranlarına karşı dikkatli olmaları gerektiği açıklanmıştır.

Yapılan bu simülasyon çalışması kapsamında DMF içeren dört madde üzerinde istatistiksel gücün yanı sıra Tip III hata da hesaplanarak, DMF türünün çoklu grup teknikleri tarafından doğru belirlenip belirlenemediği incelenmiştir. Finch ve French (2008) ile Lopez Rivas'ın (2012) çalışmalarında maddelerin gerçekte güçlük veya ayırt edicilik parametresinden kaynaklı DMF içermesi durumunda sırasıyla TBO ve TB DMF'yi belirlemek amacıyla geliştirilen tekniklerin, bu maddeleri DMF'li olarak belirleme olasılığının yüksek olduğu belirlenmiştir. Benzer şekilde gerçek veri kullanılarak yapılan bir çalışmada (Magis vd., 2011) GLK'nın aynı maddeleri hem TB hem de TBO DMF'li olarak tespit edebildiği açıklanmıştır. Aynı DMF belirleme tekniğinin her iki DMF türündeki [TB ve TBO(*a* ve *ab*)] maddelerde etkili olması, gerçek test uygulamalarında maddelerin hangi gruba avantaj sağladığının kesin olarak bilinmemesine neden olmaktadır. Özellikle ölçek geliştirme ve uyarlama çalışmalarında tespit edilen DMF'li maddelere ilişkin DMF türünün yanlış tanımlanması, bu maddelerin doğru bir şekilde revize edilebilmesine engel olabilmektedir. DMF'li maddelerin hem TB hem de TBO DMF için geliştirilen teknikler kullanılarak işaretlenmesi durumunda, Finch ve French (2008) tarafından DMF'nin kaynağının belirlenebilmesi için ek bir işleme ihtiyaç olduğu belirtilmekte ve DMF türünün kesin olarak tespit edilebilmesi için gruplara ilişkin MKE'lerin çizdirilmesi gerektiği vurgulanmaktadır. Böylece Tip III hata oranları yüksek olan teknikler için MKE'ler çizdirildiğinde, DMF türünün doğru belirlenebilmesi nedeniyle, sınava girenlerin puanlarının DMF'li maddelerden nasıl etkilendiğine dair farkındalığın artacağı ifade edilebilmektedir.

Çoklu grup tekniklerinin Tip III hata açısından performanslarını inceleyen herhangi bir simülasyon çalışması bulunmamasına rağmen ikili gruplarda Tip III hatayı dikkate alan sınırlı sayıda araştırmaya rastlanılmış (Chun vd., 2016; Finch ve French, 2008; Lopez Rivas, 2012) ve DMF miktarı ile örneklem büyüklüğünün Tip III hatayı etkileyen önemli faktörler oldukları belirtilmiştir (Finch ve French, 2008). Bu çalışmada da dikkate alınan TBO(*a*) ve TBO(*ab*) DMF türleri için referans grubun örneklem büyüklüğü ile DMF miktarı arttıkça çoklu grup teknikleri için hesaplanan Tip III hata

oranlarının genel anlamda yükseldiğinin belirlenmesi nedeniyle elde edilen sonuçların ilgili çalışmalarla örtüştüğü görülmüştür.

Çoklu grup tekniklerinin farklı DMF türlerindeki performansları incelendiğinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif için hesaplanan Tip III hata oranlarının TB, TBO(*a*) ve TBO(*ab*) DMF türlerinde farklılaştığı tespit edilmiştir. Özellikle bu teknikler arasından Lordif'in farklı DMF türlerindeki performanslarının değişiklik göstermesi sonucunda, yapılan saflaştırmanın etkisinin DMF'nin hangi madde parametresinden kaynaklandığına bağlı olarak farklılaştığı ortaya konulmuştur. Aynı zamanda araştırma kapsamında yalnızca TBO DMF'yi belirlemek amacıyla geliştirilen GLR-LRT ile GLR-Wald'ın TB DMF türündeki performansının Tip III hata açısından iyi kontrol edildiği, diğer teknikler için elde edilen değerlerin ise son derece yüksek olduğu belirlenmiştir. Ulaşılan bulgular birlikte değerlendirildiğinde; Tip III hata oranlarının hesaplandığı bütün koşullarda diğer tekniklerden daha iyi performans gösteren bir çoklu grup tekniğinin var olmadığı saptanmıştır. Böylece çoklu gruplarda Tip III hata oranlarının ilk defa araştırıldığı bu simülasyon çalışmasında gerçek uygulamalarda DMF'li olarak işaretlenen maddelerin DMF türünün belirlenebilmesi için MKE'lere başvurulmasının gerekli olduğu sonucuna ulaşılmış ve elde edilen bulguların alanyazına katkı sağlayacağı düşünülmüştür.

Tip I hata, istatistiksel güç ve Tip III hata sonuçları birlikte incelendiğinde; Tip I ve Tip III hata oranlarının .075'den düşük veya bu değere yakın, istatistiksel gücün ise .80 ve üzerinde olup olmadığı değerlendirilerek optimum performans gösteren tekniğin belirlenmesi amaçlanmıştır. Testteki dört maddenin TB DMF içerdiği koşullarda; bu DMF türü için en kullanışlı tekniğin Lordif olduğu tespit edilmiştir. Aynı zamanda Tutz ve Berger (2016) tarafından yapılan çalışmada da belirtildiği gibi yapılan analizlere ek olarak saflaştırma işlemi de gerçekleştirildiğinde GLR-LRT, GLR-Wald ve GMH tekniklerinin yeterli performans göstererek Lordif'e alternatif olabileceği düşünülmüştür. Nitekim Finch (2016) tarafından yapılan çalışmada da Tip I hata ve istatistiksel gücün optimum kombinasyonunu sağlayan tekniğin GMH olduğunun açıklanması nedeniyle bu araştırma sonucunda elde edilen bulguların iki değerlendirme kriteri için de alanyazınla örtüştüğü saptanmıştır. Öte yandan yapılan bu araştırma kapsamında çoklu gruplar için ilk geliştirilen teknik olan GLK'nın TB DMF türü için istenen düzeyde performans göstermediği tespit edilmiştir. Böylece TB DMF türü için bazı koşullarda GLK'nın DMF'li maddeleri doğru bir şekilde belirleyebildiği, yani istatistiksel gücün yüksek olduğu, gözlemlenmesine rağmen hesaplanan Tip III hata oranları sonucunda bu teknik

ile DMF'nin kaynağı hakkında doğru kararlar verebilmenin kolay olmadığı açıkça ortaya konulmuştur. Langer (2008) tarafından da MTK'ya dayalı DMF belirleme tekniklerinin güçlü teknikler oldukları belirtilmesine rağmen, LK ve dolayısıyla GLK tekniklerinin bazı sınırlılıklarının söz konusu olduğundan bahsedilmesi nedeniyle yapılan bu simülasyon çalışmasından elde edilen bulguların son derece doğal olduğu anlaşılmıştır. Böylece GLK tekniğinin gerçek test uygulamalarında kullanılabilmesi için kesinlikle MKE'lerin incelenmesine ihtiyaç olduğu sonucuna ulaşılmıştır.

Testteki dört maddenin TBO(α) DMF içerdiği koşullarda, bu DMF türü için istatistiksel güç açısından Lordif, Tip III hata açısından da GLK'nın kullanışlı teknikler oldukları belirlenmiştir. Bununla birlikte yapılan analizlere ek olarak saflaştırma işleminin gerçekleştirilmesi durumunda, en yüksek performansı gösteren tekniğin GLK olabileceği düşünülmüştür. Son olarak dört maddenin TBO($\alpha\beta$) DMF içerdiği koşullarda da istatistiksel güç açısından GLK, Tip I hata açısından GLR-LRT, GLR-Wald'ın kullanışlı teknikler oldukları, Tip III hata açısından da teknikler arasında önemli bir fark olmadığı tespit edilmiştir. Lordif için ise saflaştırma otomatik olarak yapılmasına rağmen istenen düzeyde bir performans gözlenememesi nedeniyle TBO($\alpha\beta$) DMF türündeki maddelerde bu teknik ile doğru sonuçlara varmanın zor olduğu anlaşılmıştır. Güler ve Penfield (2009) tarafından yapılan çalışmada da DMF belirlemek amacıyla geliştirilen teknikler arasından hem TB hem de TBO DMF'yi tespit etmek için tek bir istatistiksel tekniğin etkili olmadığı ifade edildiğinden, bu açıklamaların yapılan simülasyon çalışması ile desteklendiği sonucuna ulaşılmıştır.

Alanyazında çoklu grup DMF belirleme tekniklerine ilişkin bazı üstünlük ve sınırlılıkların söz konusu olduğu belirtilmiştir (Carroll, 2015; Langer, 2008). Özellikle MTK'ya dayalı teknikler kullanılarak yapılan analizlerin büyük örneklem gerektirmesi ve hesaplama süresinin uzun olması gibi bazı kısıtlayıcı durumlar nedeniyle gerçek uygulamalarda bu tekniklerin kullanılmasının KTK'ya dayalı tekniklere kıyasla daha zor olduğu ifade edilmiştir (Finch, 2016; Jin vd., 2015). MTK dayalı olduğu belirtilen Lordif tekniğinin çoklu gruplardaki performansının ilk kez incelendiği bu çalışmada, belirtilen tekniğin genel anlamda diğer tekniklerden daha iyi veya benzer performansa sahip olduğu belirlenmiştir. Vaughn-Coaxum ve diğerlerinin (2016) çalışmalarında da Lordif tekniği gerçek veride iki grup arasında DMF belirlemek amacıyla kullanmış ve bireyleri eşleştirmek için MTK'ya dayalı kestirilen yetenek parametrelerinin kullanılmasının daha doğru sonuçlar elde edilmesine katkı sağladığı açıklanmıştır. Benzer şekilde Allahyari (2020) tarafından yapılan çalışmada Lordif'in DMF belirlemek amacıyla kullanılabilecek

en iyi tekniklerden biri olduđu vurgulanmıřtır. Bu arařtırma kapsamında da Tip I hata, istatistiksel g ve Tip III hata oranları birlikte incelendiđinde; beklendiđi gibi eřleřtirme kriteri olarak yetenek parametrelerinin dikkate alındıđı MTK (GLK ve Lordif) ve gzlenen puanların dikkate alındıđı KTK (GLR-LRT, GLR-Wald ve GMH) temelli tekniklerin performanslarının farklılařtıđı anlařılmıřtır.

Bu simlasyon alıřmasında GLR-LRT, GLR-Wald ve GMH tekniklerinin Tip I hata, istatistiksel g ve Tip III hata oranlarına iliřkin performansların btn kořullarda birbirlerine olduka yakın oldukları tespit edilmiřtir. Alanyazında yapılan ok sayıda ilgili arařtırmada da (Schauberger ve Tutz, 2016; Rollins III, 2018; Tutz ve Berger, 2016; Tutz ve Schauburger, 2015) hem istatistiksel g hem de Tip I hata oranları aısından GLR ve GMH tekniklerinin birbirleri ile son derece benzer, GLK'dan ise farklı performans ortaya koydukları gzlemlenmiřtir. Uyar ve Kaya Uyanık (2016) ve Morales-Ortiz ve Peđa-Cardenas (2019) tarafından gerek veriler kullanılarak yapılan arařtırmalarda, GMH ve GLR tekniklerinin genellikle benzer performans gsterdikleri aıklandıđından, elde edilen bulguların ilgili alıřmalarla rtřtđ belirlenmiřtir. Ayrıca alanyazındaki alıřmalarda birbirlerine benzer performans ortaya koydukları belirtilen GLR-LRT ve GLR-Wald teknikleri (Gao, 2019; Magis vd., 2011) iin hesaplanan Tip I hata, istatistiksel g ve Tip III hata oranlarının btn kořullarda son derece tutarlı oldukları saptanmıř ve bu sonu simlasyonun beklendiđi gibi gerekleřtiđini gstermiřtir.

BÖLÜM 5

SONUÇLAR VE ÖNERİLER

Çalışmanın bu bölümünde araştırma sorularına yanıt aramak amacıyla yapılan analizlerden yola çıkılarak bulgulara dayalı ulaşılan sonuçlara ve bu sonuçlar doğrultusunda yapılan bazı önerilere yer verilmiştir.

Sonuçlar

Bu simülasyon çalışmasında GLK, GLR-LRT, GLR-Wald, GMH ve Lordif'in grup sayısı, referans grubun örneklem büyüklüğü, odak grubun örneklem büyüklüğü oranı, DMF miktarı ve DMF türü faktörleri altında Tip I hata, istatistiksel güç ve Tip III hata oranlarının nasıl etkilendiği incelenmiştir. Araştırmada DMF'nin varlığı hakkında hiçbir şeyin bilinmediği gerçek test uygulamalarına uygun koşulların incelenmesi amaçlanmıştır. Çalışmanın amacına uygun olması için simülasyon sürecinde ek bir saflaştırma işlemi gerçekleştirilmeden çoklu grup tekniklerinin birbirlerine göre üstün ve zayıf yönleri tespit edilmiştir. Öte yandan simülasyonun istatistiksel geçerliğini kanıtlamak amacıyla mutlak yanlılık ve RMSE hesaplanarak, güçlük ve ayırt edicilik parametrelerine ilişkin kestirimlerin doğruluğu araştırılmıştır. Araştırma kapsamında yapılan analizler sonucunda istatistiksel geçerlik, Tip I hata, istatistiksel güç ve Tip III hata oranları için ulaşılan sonuçlar aşağıda sunulmuştur. Sonrasında ise dikkate alınan çoklu grup tekniklerine ilişkin genel sonuçlar raporlanmıştır.

İstatistiksel Geçerliğe İlişkin Sonuçlar

Simülasyonun geçerli olup olmadığının belirlenmesi amacıyla 2PLM'ye dayalı yapılan parametre kestirimlerinin doğruluğu, mutlak yanlılık ve RMSE hesaplanarak değerlendirilmiştir. Araştırmanın bütün koşullarında hesaplanan RMSE değerlerinin mutlak yanlılık değerlerinden nispeten daha yüksek olduğu belirlenmiştir. Öte yandan ayırt edicilik parametresi için hesaplanan mutlak yanlılık ile RMSE'lerin güçlük parametresine kıyasla daha yüksek olduğu gözlemlenmiş ve hem güçlük hem de ayırt

edicilik parametresi için toplam örneklem büyüklüğü arttıkça mutlak yanlışlık ile RMSE değerlerinin sıfıra yaklaşma eğiliminde oldukları sonucuna ulaşılmıştır. Madde parametreleri için hesaplanan sıfıra yakın mutlak yanlışlık ile RMSE değerleri, kestirilen ve gerçek parametreler arasındaki farkın az, yani yapılan kestirimlerin kabul edilebilir olduğunu göstermiştir.

Tip I Hata Oranlarına İlişkin Sonuçlar

Tıpkı herhangi bir hipotez testinde olduğu gibi DMF çalışmalarında da hesaplanan Tip I hata oranlarının beklendiği gibi olup olmadığı konusunda endişe duyulduğu belirtilmiştir (Lan, 2014). Bu nedenle yapılan bu simülasyon çalışmasında da hem testte DMF’li maddelerin bulunmadığı hem de testteki dört maddenin DMF’li olduğu koşullarda Tip I hata hesaplanmıştır. Madde parametrelerinde herhangi bir değişimleme yapılmadığında; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip I hata oranlarının nominal değere yakın olmaları nedeniyle testte hiçbir DMF’li madde bulunmazsa eğer, çoklu grup tekniklerinin maddeleri yanlışlıkla DMF’li olarak işaretleme olasılığının kabul edilebilir düzeyde olduğu sonucuna ulaşılmıştır. GLK tekniğinin diğerlerinden daha düşük temel Tip I hata oranına sahip olduğu araştırmada, referans grubun örneklem büyüklüğü veya grup sayısı artsa dahi çoklu grup tekniklerine ilişkin performansların farklılaşmadığı görülmüştür.

Testteki dört maddenin TB DMF’li olması durumunda; Lordif haricindeki teknikler için hesaplanan Tip I hata oranlarının temel Tip I hata oranlarından nispeten daha yüksek olduğu belirlenmiştir. GLR-LRT, GLR-Wald ve GMH tekniklerine ilişkin performansların bütün koşullarda liberal kritere göre kabul edilebilir olmadığı, Lordif içinse elde edilen değerlerin liberal kritere göre iyi kontrol edildiği tespit edilmiştir. Bu DMF türünde Lordif için hesaplanan temel Tip I hata ile Tip I hata oranlarının son derece benzer olması, saflaştırma işleminin Lordif’in performansına olumlu yönde katkı sağladığını göstermiştir. Aynı zamanda TB DMF miktarı ile referans grubun örneklem büyüklüğü arttığında GLK, GLR-LRT, GLR-Wald ve GMH teknikleri için hesaplanan Tip I hata oranlarının yükseldiği, buna karşın belirtilen tekniklerin performanslarının grup sayısından etkilenmediği gözlemlenmiştir.

Testteki dört maddenin TBO(a) DMF’li olması durumunda; hesaplanan Tip I hata oranlarının temel Tip I hata oranlarına son derece yakın olduğu ve referans grubun örneklem büyüklüğü, TBO(a) DMF miktarı veya grup sayısı artsa dahi çoklu grup

tekniklerine ilişkin performansların değişmediği gözlemlenmiştir. TBO(*a*) DMF türünde GLK, GLR-LRT, GLR-Wald ve Lordif tekniklerinin benzer performans göstermesi, bu DMF türündeki maddeleri içeren testlerde saflaştırma işleminin Lordif'in performansında iyileşmeye yol açmadığını ortaya koymuştur.

Testteki dört maddenin TBO(*ab*) DMF'li olması durumunda ise GLR-LRT ve GLR-Wald haricindeki teknikler için hesaplanan Tip I hata oranlarının temel Tip I hata oranlarından nispeten daha yüksek olduğu belirlenmiştir. TBO(*ab*) DMF türünde Lordif için hesaplanan Tip I hata oranlarının temel Tip I hata oranlarından yüksek olması, saflaştırmanın Lordif'in performansına olumlu yönde katkı sağlamadığını göstermiştir. Aynı zamanda TBO(*ab*) DMF miktarı ile referans grubun örneklem büyüklüğü arttığında GLK ve Lordif teknikleri için hesaplanan Tip I hata oranlarının yükseldiği, buna karşın belirtilen tekniklerin performanslarının grup sayısından etkilenmediği tespit edilmiştir.

GLK tekniğinin TBO(*a*) DMF türünde düşük, TB ve TBO(*ab*) DMF türlerinde yüksek, GLR-LRT ve GLR-Wald tekniklerinin ise TB DMF türünde yüksek, TBO(*a*) ve TBO(*ab*) DMF türlerinde düşük Tip I hata oranlarına sahip olmaları, bu tekniklerin performanslarının güçlük parametresinden olumsuz etkilendiğini göstermiştir. Lordif tekniği için yapılan analizler sonucunda ise TBO(*ab*) DMF türünde yüksek, TB ve TBO(*a*) DMF türlerinde düşük Tip I hata oranları elde edilmiş ve bu tekniğin performansının güçlük ile ayırt edicilik parametrelerinin etkileşiminden olumsuz yönde etkilendiği belirlenmiştir.

İstatistiksel Güç Oranlarına İlişkin Sonuçlar

Testteki dört maddenin TB DMF'li olması durumunda; referans grubun örneklem büyüklüğü ve TB DMF miktarı arttıkça istatistiksel güç oranlarının yükseldiği, buna karşın istatistiksel güç oranlarının grup sayısından manidar bir şekilde etkilenmediği görülmüştür. Bu DMF türünde $N_R = 600$ ve DMF miktarı düşük iken GLK'nın yetersiz, GLR-LRT ve GLR-Wald'ın kabul edilebilir, Lordif'in ise yüksek düzeyde performans gösterdikleri belirlenmiştir. DMF miktarı orta ve yüksek düzey olduğunda ise bütün koşullarda GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin yüksek istatistiksel güç oranlarına sahip oldukları tespit edilmiştir. Ayrıca TB DMF miktarı önemli olmaksızın $N_R = 1800$ iken dikkate alınan çoklu grup tekniklerine ilişkin istatistiksel güç oranlarının yükselmesi sonucunda tekniklerin performansları arasındaki farkların azaldığı gözlemlenmiştir. Bu DMF türünde Lordif için hesaplanan istatistiksel

güç oranlarının diğer tekniklerden daha yüksek olması, saflaştırmanın Lordif'in performansına olumlu yönde katkı sağladığını göstermiştir.

Testteki dört maddenin TBO(*a*) DMF'li olması durumunda; referans grubun örneklem büyüklüğü ve TBO(*a*) DMF miktarı arttıkça istatistiksel güç oranlarının yükseldiği, buna karşın istatistiksel güç oranlarının grup sayısından manidar bir şekilde etkilenmediği görülmüştür. Bu DMF türünde hem $N_R = 600$ hem de DMF miktarı düşük iken GLK, GLR-LRT, GLR-Wald ve Lordif'in yetersiz düzeyde performans gösterdikleri belirlenmiştir. Bununla birlikte $N_R = 600$ ve DMF miktarı orta düzey olduğunda GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin kabul edilebilir istatistiksel güç oranlarına sahip oldukları tespit edilmiştir. Son olarak $N_R = 600$ ve DMF miktarı yüksek olduğunda GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerine ilişkin performansların yükseldiği saptanmıştır. Ayrıca TBO(*a*) DMF miktarı önemli olmaksızın $N_R = 1800$ iken dikkate alınan çoklu grup tekniklerine ilişkin istatistiksel güç oranlarının yükselmesi sonucunda tekniklerin performansları arasındaki farkların azaldığı gözlemlenmiştir. TBO(*a*) DMF türünde Lordif için hesaplanan istatistiksel güç oranlarının diğer tekniklerden daha yüksek olması, saflaştırmanın Lordif'in performansına olumlu yönde katkı yaptığını göstermiştir.

Testteki dört maddenin TBO(*ab*) DMF'li olması durumunda; referans grubun örneklem büyüklüğü ve TBO(*ab*) DMF miktarı arttıkça istatistiksel güç oranlarının yükseldiği, buna karşın istatistiksel güç oranlarının grup sayısından manidar bir şekilde etkilenmediği görülmüştür. Bu DMF türünde hem $N_R = 600$ hem de DMF miktarı düşük iken GLK'nın yüksek, Lordif'in kabul edilebilir ve GLR-LRT ile GLR-Wald'ın yetersiz düzeyde performans gösterdikleri belirlenmiştir. Bununla birlikte $N_R = 600$ ve DMF miktarı orta düzey olduğunda GLR-LRT ve GLR-Wald'ın yetersiz, GLK ve Lordif'in ise yüksek istatistiksel güç oranlarına sahip oldukları tespit edilmiştir. Son olarak $N_R = 600$ ve DMF miktarı yüksek olduğunda GLR-LRT ve GLR-Wald'ın kabul edilebilir, GLK ve Lordif'in ise yüksek düzey performans ortaya koydukları saptanmıştır. Ayrıca TBO(*ab*) DMF miktarı önemli olmaksızın $N_R = 1800$ iken dikkate alınan çoklu grup tekniklerine ilişkin istatistiksel güç oranlarının yükselmesi nedeniyle tekniklerin performansları arasındaki farkların azaldığı gözlemlenmiştir. TBO(*ab*) DMF türünde Lordif için hesaplanan istatistiksel güç oranlarının GLK'dan daha düşük olması, DMF'nin hem güçlük hem de ayırt edicilik parametresinden kaynaklanması durumunda saflaştırmanın Lordif'in performansına beklenen düzeyde katkı yapmadığını göstermiştir.

GLR-LRT, GLR-Wald ve Lordif tekniklerinin TB DMF türünde yüksek, TBO(a) ve TBO(ab) DMF türlerinde düşük istatistiksel güç oranlarına sahip olmaları, bu tekniklerin güçlük parametresinden kaynaklı DMF'li maddeleri daha kolay belirleyebildiğini ortaya koymuştur. GLK tekniğinin ise TB ve TBO(a) DMF türlerinde düşük, TBO(ab) DMF türünde yüksek istatistiksel güç oranlarına sahip olması, bu tekniğin performansının güçlük ile ayırt edicilik parametrelerinin etkileşiminden olumlu yönde etkilendiğini göstermiştir. TBO(a) ve TBO(ab) DMF türleri için DMF'li dört maddenin düşük DMF miktarı içerdiği koşullarda, Lordif'in .80'in oldukça altında istatistiksel güç oranlarına sahip olmasına rağmen tekniğin yalnızca TBO(a) DMF türünde diğer çoklu grup tekniklerinden daha yüksek performans göstermesi, saflaştırma işleminin bu DMF türünde Lordif'in performansına katkı sağladığına işaret etmiştir.

Tip III Hata Oranlarına İlişkin Sonuçlar

Testteki dört maddenin TB DMF'li olması durumunda; GLR-LRT ve GLR-Wald tekniklerine ilişkin performansların bütün koşullarda liberal kritere göre iyi kontrol edildiği, GLK ve Lordif teknikleri için hesaplanan Tip III hata oranlarının ise kabul edilebilir sınır değerden yüksek olduğu tespit edilmiştir. TB DMF miktarı ile referans grubun örneklem büyüklüğü arttığında GLK için hesaplanan Tip III hata oranlarının yükseldiği görülmüne rağmen belirtilen tekniğin performansının grup sayısından etkilenmediği gözlemlenmiştir. GLR-LRT, GLR-Wald ve Lordif tekniklerine ilişkin performansların ise TB DMF miktarı, referans grubun örneklem büyüklüğü ve grup sayısına göre farklılaşmadığı görülmüştür. TB DMF türünde Lordif için hesaplanan Tip III hata oranlarının dikkate alınan faktörlerden etkilenmemesi nedeniyle saflaştırma işleminin Lordif'in performansına olumlu yönde katkı sağladığı belirlenmiştir.

Testteki dört maddenin TBO(a) DMF'li olması durumunda; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının liberal kritere göre kabul edilebilir olmadığı görülmüştür. Öte yandan referans grubun örneklem büyüklüğü ile TBO(a) DMF miktarı arttıkça çoklu grup teknikleri için elde edilen değerlerin yükseldiği belirlenmesine rağmen, bu tekniklere ilişkin performansların grup sayısından etkilenmediği gözlemlenmiştir. Ayrıca araştırmada GLK için hesaplanan Tip III hata oranlarının GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerden daha düşük olduğu belirlenmiştir. TBO(a) DMF türünde Lordif'in GLR-LRT, GLR-Wald ve GMH

teknikleri ile benzer performans ortaya koyması, bu DMF türünde saflaştırma işleminin Lordif'in performansında iyileşmeye yol açmadığını göstermiştir.

Testteki dört maddenin TBO(ab) DMF'li olması durumunda ise GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri için hesaplanan Tip III hata oranlarının liberal kritere göre kabul edilebilir sınırın oldukça üzerinde olduğu belirlenmiştir. Aynı zamanda TBO(ab) DMF türü için DMF miktarının düşük olduğu durumda referans grubun örneklem büyüklüğü arttıkça dikkate alınan çoklu grup teknikleri için Tip III hata oranlarının yükseldiği görülmeye rağmen elde edilen değerlerin grup sayısından etkilenmediği tespit edilmiştir. DMF miktarı orta veya yüksek düzey iken referans grubun örneklem büyüklüğü önemli olmaksızın çoklu grup tekniklerine ilişkin Tip III hata oranlarının yükselmesi sonucunda dikkate alınan tekniklerin performansları arasındaki farkların azaldığı gözlemlenmiştir. Bu DMF türündeki bütün koşullarda Lordif için hesaplanan Tip III hata oranlarının ya diğer tekniklere benzer ya da onlardan yüksek olması, saflaştırma işleminin Lordif'in performansına olumlu yönde katkı sağlamadığını göstermiştir.

GLK tekniğinin TBO(a) DMF türünde düşük, TB ve TBO(ab) DMF türlerinde yüksek Tip III hata oranlarına sahip olması, bu tekniğin performansının güçlü parametresinden olumsuz etkilendiğini göstermiştir. GLR-LRT, GLR-Wald ve Lordif tekniklerinin ise TB DMF türünde düşük, TBO(a) ve TBO(ab) DMF türlerinde yüksek Tip III hata oranları elde edilmiş ve bu tekniklerin performanslarının ayırt edicilik parametresinden olumsuz yönde etkilendiği belirlenmiştir.

Çoklu Grup Tekniklerinin Performanslarına İlişkin Genel Sonuçlar

Yapılan bu simülasyon çalışmasında da çoklu grup tekniklerinin Tip I hata, istatistiksel güç ve Tip III hata oranları birlikte değerlendirilerek bu teknikler arasından performansı en yüksek olan teknik veya teknikler tespit edilmiştir. Guilera ve diğerleri (2013) tarafından yapılan araştırmada farklı koşulların dikkate alındığı simülasyon çalışmalarından elde edilen sonuçların bütünleştirilerek ilgilenilen tekniklerin performansları hakkında genel sonuçlara ulaşılmasının kolay olmadığı belirtildiğinden bu araştırma kapsamında da ulaşılan sonuçlar TB, TBO(a) ve TBO(ab) olmak üzere üç farklı DMF türü için özetlenmiştir.

TB DMF içeren testlerde, Lordif'in gerçekte DMF'li olmayan maddeleri yanlışlıkla DMF'li olarak işaretleme olasılığının düşük, TB DMF miktarı önemli

olmaksızın gerçekte DMF'li olan maddeleri doğru belirleme olasılığının yüksek ve DMF'nin türünü yanlış tespit etme olasılığının düşük olduğu gözlemlenmiştir. GLK'nın ise gerçekte DMF'li olmayan maddeleri yanlışlıkla DMF'li olarak işaretleme olasılığının yüksek, TB DMF miktarı düşük iken gerçekte DMF'li olan maddeleri doğru belirleme olasılığının düşük ve DMF'nin türünü yanlış tespit etme olasılığının yüksek olduğu görülmüştür. Son olarak GLR-LRT, GLR-Wald, GMH tekniklerinin gerçekte DMF'li olmayan maddeleri yanlışlıkla DMF'li olarak işaretleme olasılığının yüksek, TB DMF miktarı düşük iken gerçekte DMF'li olan maddeleri doğru belirleme olasılığının düşük ve DMF'nin türünü yanlış tespit etme olasılığının düşük olduğu belirlenmiştir. Böylece testteki dört madde TB DMF içerdiğinde dikkate alınan teknikler arasında istatistiksel gücü daha yüksek, Tip I ve Tip III hatası ise daha düşük olması nedeniyle diğer tekniklerden daha iyi performans gösteren tekniğin Lordif olduğu sonucuna ulaşılmıştır.

TBO(a) DMF içeren testlerde, Lordif'in gerçekte DMF'li olmayan maddeleri yanlışlıkla DMF'li olarak işaretleme olasılığının düşük, TBO(a) DMF miktarı düşük iken gerçekte DMF'li olan maddeleri doğru belirleme olasılığının düşük ve DMF'nin türünü yanlış tespit etme olasılığının yüksek olduğu gözlemlenmiştir. GLK'nın ise gerçekte DMF'li olmayan maddeleri yanlışlıkla DMF'li olarak işaretleme olasılığının düşük, TBO(a) DMF miktarı düşük iken gerçekte DMF'li olan maddeleri doğru belirleme olasılığının düşük ve DMF'nin türünü yanlış tespit etme olasılığının diğer tekniklerden düşük olduğu görülmüştür. Son olarak GLR-LRT ve GLR-Wald tekniklerinin gerçekte DMF'li olmayan maddeleri yanlışlıkla DMF'li olarak işaretleme olasılığının düşük, TBO(a) DMF miktarı düşük iken gerçekte DMF'li olan maddeleri doğru belirleme olasılığının düşük ve DMF'nin türünü yanlış tespit etme olasılığının yüksek olduğu belirlenmiştir. Böylece testteki dört madde TBO(a) DMF içerdiğinde dikkate alınan teknikler arasında istatistiksel gücü daha yüksek, Tip I ve Tip III hatası ise daha düşük olması nedeniyle diğer tekniklerden daha iyi performans gösteren tekniğin GLK olduğu sonucuna ulaşılmıştır.

TBO(ab) DMF içeren testlerde, Lordif'in gerçekte DMF'li olmayan maddeleri yanlışlıkla DMF'li olarak işaretleme olasılığının yüksek, TBO(ab) DMF miktarı düşük iken gerçekte DMF'li olan maddeleri doğru belirleme olasılığının düşük ve DMF'nin türünü yanlış tespit etme olasılığının yüksek olduğu gözlemlenmiştir. GLK'nın ise gerçekte DMF'li olmayan maddeleri yanlışlıkla DMF'li olarak işaretleme olasılığının yüksek, TBO(ab) DMF miktarı önemli olmaksızın gerçekte DMF'li olan maddeleri doğru belirleme olasılığının yüksek ve DMF'nin türünü yanlış tespit etme olasılığının yüksek

olduğu tespit edilmiştir. Son olarak GLR-LRT ve GLR-Wald tekniklerinin gerçekte DMF'li olmayan maddeleri yanlışlıkla DMF'li olarak işaretleme olasılığının düşük, TBO(ab) DMF miktarı önemli olmaksızın gerçekte DMF'li olan maddeleri doğru belirleme olasılığının düşük ve DMF'nin türünü yanlış tespit etme olasılığının yüksek olduğu belirlenmiştir. Böylece TBO(ab) DMF türü için Tip I hata, istatistiksel güç ve Tip III hata oranları birlikte incelendiğinde, herhangi bir tekniğin diğerlerine kıyasla ön plana çıkacak performans sergilemediği görülmüştür. Bu bağlamda yapılan simülasyon sonuçlarına dayalı olarak DMF belirleme tekniklerinin farklı DMF türlerinden etkilendiği anlaşıldığından ölçme uygulamalarında birden fazla çoklu grup tekniğinin birlikte kullanılmasının son derece önemli olduğu sonucuna ulaşılmıştır.

Öneriler

Bu bölümde GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin performanslarını incelemek amacıyla gerçekleştirilen simülasyon sonucunda ulaşılan bulgular çerçevesinde hem DMF ile ilgilenen araştırmacılara hem de test uygulayıcılarına yönelik önerilere yer verilmiştir.

Araştırmacılara Öneriler

- Araştırmada çoklu grupların farklı koşullar altındaki performansları Tip I hata, istatistiksel güç ve Tip III hata açısından incelenmiştir. Bu çalışmada dikkate alınan faktörlerin düzeyleri farklılaştırılarak; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin performansları grup sayısı (3, 6, 9), referans grubun örneklem büyüklüğü (300, 600, 1200), odak grupların örneklem büyüklüklerinin referans gruba oranı (1:1/2, 1:1/3 ve 1:1/4), DMF miktarı [$a(0.0, 0.3, 0.6)$, $b(0.0, 0.3, 0.6)$] ve DMF türü [DMF Yok, TB, TBO(a), TBO(ab)] koşullarında incelenebilir.
- Bu araştırmada grup sayısının üç veya dört olduğu durumda çoklu grup tekniklerinin performansları incelenmiş ve bazı tekniklerin performanslarının grup sayısına göre bir miktar farklılaştığı belirlenmesine rağmen bu farklılığın manidar olmadığı görülmüştür. İlerleyen araştırmalarda 5, 10, 15 veya 20 grup dikkate alınarak çoklu grupların performansları karşılaştırılabilir.

- Bu arařtırmada odak gruplardaki en kk rneklem bklnn 300 olduėu durumun dikkate alınması nedeniyle GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinin 100 ve 200 gibi daha kk rneklemelerden oluřan veri setlerinde kullanılmasının uygun olup olmadıėının incelenmesi nemli olabilir.
- Arařtırma kapsamında gerekleřtirilen simlasyonun tamamlanma sresinin uzamasından dolayı deėiřimlenen faktrler oklu grup teknikleri, grup sayısı, referans grubun rneklem bkl, odak grupların rneklem bkllerinin referans gruba oranı, DMF miktarı ve DMF tr ile sınırlı tutulmuřtur. İleride testteki DMF’li madde yzdesi, test uzunluėu, veri retmede kullanılan model, yetenek ve madde parametrelerinin daėılımı gibi faktrlerden biri veya birkaçı daha deėiřimlenerek, oklu grup tekniklerinin performanslarını belirlemek amacıyla yeni simlasyon alıřmaları yapılabilir.
- Yapılan bu arařtırmada incelenenlere benzer veya farklı kořullar dikkate alınarak GLK, GLR-LRT, GLR-Wald, GMH ve Lordif teknikleri ile LGGW, MG-NCDIF, LRL, Rasch Aėacı, BGGW tekniklerinin performansları karřılařtırılabilir.
- Bu simlasyon alıřmasında Finch’in (2016) arařtırmasında kullandıėı gerek test uygulamasından elde edilen madde parametreleri dikkate alınarak drt maddenin madde gk ile madde ayırt edicilik parametreleri gruplar arasında farklılařtırılmıřtır. Bu alanda alıřma yapan arařtırmacılar yapacakları alıřmalarda referans grubun madde parametrelerinin oklu grup tekniklerinin performansları zerindeki etkisini inceleyebilir.
- Yapılan bu alıřmada referans grubun odak gruplardan avantajlı olduėu kabul edilmiřtir. Bařka arařtırmalarda DMF’nin yn tersine evrilerek referans grubun bazı odak gruplardan dezavantajlı olduėu daha karmařık DMF belirleme alıřmaları yapılabilir.
- Bu arařtırmada tek boyutlu 2PLM’ye dayalı veri retilerek gk parametresinin yanı sıra ayırt edicilik parametresinin de farklılařması saėlanmasına raėmen oklu grup tekniklerine iliřkin performansların 3PLM’ye dayalı olarak retilen verilerde farklılařabileceėi belirtilmiřtir. Bu nedenle ileriki arařtırmalarda 3PLM’ye dayalı veri setleri retilerek řans parametresinin de dikkate alındıėı arařtırmalar planlanabilir.
- Psikolojik zelliklerin genellikle ok boyutlu olduėu dřnldėnde, GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerine iliřkin performansların

gerçek veya simüle edilmiş çok boyutlu veri setleri kullanılarak karşılaştırılmasının da alanyazına katkı sağlayabileceği söylenebilir.

- Yapılan bu simülasyon çalışmasının bulguları, gerçek test uygulamalarından elde edilen veri setleri kullanılarak yapılacak araştırmalar ile desteklenebilir.
- Bu çalışmada Lordif tekniği için saflaştırma işlemi otomatik olarak gerçekleştirilmiştir. Diğer teknikler için de farklı saflaştırma yaklaşımları kullanılarak çoklu grup tekniklerin performansları saflaştırmanın yapıldığı koşullarda da karşılaştırılabilir.
- Bu araştırmada ikili puanlanan maddeler için geliştirilen çoklu grup tekniklerinin performansları araştırılmıştır. İleriki araştırmalarda çoklu puanlanan maddeler için geliştirilen çoklu grup tekniklerinin performansları bir simülasyon çalışması kapsamında karşılaştırılabilir. Benzer bir çalışma geniş ölçekli uygulamalarda sıklıkla kullanılan karma testler için de yapılabilir.
- Bu simülasyon çalışmasında dikkate alınan her bir koşul için 100 tekrar yapılmıştır. DMF ile ilgilenen araştırmacıların yapacakları yeni çalışmalarda tekrar sayısını da değişimlemeyerek farklı tekrar sayılarını dikkate almaları önerilebilir.
- Bu çalışmada .05'lik bir p değeri kullanılmış ve araştırmanın sonuçlarını etkileyen başka bir değişkenin de nominal manidarlık düzeyi olduğu düşünülmüştür. Bu nedenle yapılacak araştırmalarda farklı bir alfa düzeyi dikkate alınarak çoklu grup tekniklerinin Tip I ve Tip III hata açısından performansları incelenebilir.
- İlerleyen araştırmalarda çoklu grup teknikleri için hem istatistiksel sonuçların hem de grafiksel gösterimlerin birleştirilerek tek bir paket veya kullanıcı dostu bir ara yüze sahip program ile elde edilmesini sağlayacak yazılımlar geliştirilebilir.

Uygulayıcılara Öneriler

- Gerçek veriler üzerinde yapılan analizlerde bir tekniğin çok sayıda maddeyi DMF'li olarak işaretlemesi, o tekniğin daha iyi olduğu anlamına gelmeyebilir. Benzer şekilde bir tekniğin az sayıda maddeyi DMF'li olarak belirlemesi de ilgili tekniğin yetersiz olduğu anlamına gelmemelidir. Bu nedenle tekniklerin nasıl performans göstereceğini bilmeyen uygulayıcıların daha doğru sonuçlara

ulařabilmeleri iin simlasyon alıřmalarının sonularını incelemeleri gerekmektedir. Bylece ikiden fazla grubun dikkate alındığı gerek test uygulamalarındaki DMF analizleri srecinde; GLK, GLR-LRT, GLR-Wald, GMH ve Lordif tekniklerinden biri veya birkaını kullanmak isteyen arařtırmacıların, veri setlerine en uygun tekniğı belirleyebilmeleri iin bu simlasyon alıřmasının bulgularından yararlanmaları onerilebilir.

- lek geliřtirme ve uyarlama alıřmalarında psikometristler ve test geliřtiricilerin, testte DMF’li madde olup olmadığını analizlerin ncesinde bilmeleri mmkn değıldir. Bu nedenle oklu grupların dikkate alındığı gerek test uygulamalarındaki testlerde DMF’li madde olabileceğı dřnlerek, ek bir saflařtırma iřleminin gerekleřtirilmesine gerek olmayan Lordif tekniğinin tercih edilmesi onerilebilir.
- Arařtırma kapsamında testteki maddelerin TB DMF iermesi durumunda Lordif’in, TBO(*a*) DMF iermesi durumunda ise GLK’nın yksek performans gstermesi nedeniyle bu simlasyon alıřmasının kořullarını yansıtan gerek uygulamalardaki analiz srecinde bu iki tekniğın kullanılması onerilebilir.
- Gerek uygulamalarda testte DMF’li madde olup olmadığı ve DMF’li madde var ise DMF’nin tr ve miktarı analizlerin ncesinde tahmin edilememesine rağmen rneklem byklğ hakkında bilgi sahibi olunması nedeniyle, zellikle kk rneklemelerin dikkate alındığı gerek test uygulamalarında arařtırmacıların veri setindeki rneklem byklğne uygun oklu grup tekniklerini kullanmaları onerilebilir.
- lme uygulamalarında DMF gsteren maddeleri belirlemek hikyenin sonu değıl, aksine bařlangıcıdır. Nitekim aynı kořullarda hem istatistiksel gc hem de Tip III hatası yksek olan oklu grup teknikleri tarafından gerekte DMF’li maddelerin DMF trnn doğru tespit edilememesi nedeniyle, DMF’nin hangi parametreden kaynaklandığının belirlenebilmesi iin arařtırmacıların MKE’leri de izdirmeleri tavsiye edilebilir. İstatistiksel tekniklerin sonularına ek olarak grafiksel gsterimlerden de yararlanılmasının daha doğru bir yaklařım olduğı onerilebilir.

KAYNAKLAR

- Acar, T. (2008). *Maddenin Farklı Fonksiyonlaşmasını Belirlemede Kullanılan Genelleştirilmiş Aşamalı Doğrusal Modelleme, Lojistik Regresyon ve Olabilirlik Oranı Tekniklerinin Karşılaştırılması* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara, Türkiye.
- Acar, T. (2011). Maddenin Farklı Fonksiyonlaşmasında Örneklem Büyüklüğü: Genelleştirilmiş Aşamalı Doğrusal Modelleme Uygulaması. *Kuram ve Uygulamada Eğitim Bilimleri*, 11(1), 279-288.
- Acar, T. ve Kelecioğlu, H. (2010). Maddenin Farklı Fonksiyonlaşmasını Belirleme Tekniklerinin Karşılaştırılması: GADM, LR ve MTK-OO. *Kuram ve Uygulamada Eğitim Bilimleri*, 10(2), 623-649.
- Ackerman, T. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29(1), 67–91. Retrieved from <http://www.jstor.org/stable/1434777>
- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*, 22(3), 37-51. doi: 10.1111/j.1745-3992.2003.tb00136.x
- Agresti, A. (2007). *Introduction to categorical data analysis* (2nd ed.). New York: Wiley.
- Ahmadi, A., & Bazvand, A. D. (2016). Gender differential item functioning on a national field-specific test: The case of PhD entrance exam of TEFL in Iran. *Iranian Journal of Language Teaching Research* 4(1), 63-82. doi: 10.30466/IJLTR.2016.20378
- Alfons, A., Templ, M., & Filzmoser, P. (2010). An object-oriented framework for statistical simulation: The r package simFrame. *Journal of Statistical Software*, 37(3), 1-35. doi: 10.18637/jss.v037.i03
- Allahyari, E. (2020). Differential item functioning analysis of the bar-on Emotional Quotient Inventory (EQ-I) across gender and age groups: An iterative hybrid ordinal logistic regression. *Psychologia*, 14(2), 95-105 doi: 10.21500/19002386.4894
- American Educational Research Association, American Psychological Association., & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Angoff, W. H., & Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement*, 10(2), 95-106. doi: 10.1111/j.1745-3984.1973.tb00787.x

- 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(4), 277-300. doi: 10.1111/j.1745-3984.1999.tb00558.x
- Arim, R. G., & Ercikan, K. (2014). Comparability between the American and Turkish versions of the TIMSS mathematics test results. *Education and Science*, 39(172), 33-48.
- Atar, B. (2007). *Differential item functioning analyses for mixed response data using IRT likelihood-ratio test, logistic regression, and GLLAMM procedures* (Unpublished doctoral dissertation). Florida State University, Florida.
- Bakan Kalaycıoğlu, D. ve Kelecioğlu, H. (2011). Öğrenci Seçme Sınavı'nın Madde Yanlılığı Açısından İncelenmesi. *Eğitim ve Bilim*, 36(161), 3-12.
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37, 379–384. doi: 10.3758/BF03192707
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *Fitting linear mixed-effects models using lme4*. *Journal of Statistical Software*, 67(1), 1-48. doi: 10.18637/jss.v067.i01
- Battauz, M. (2019). On wald tests for differential item functioning detection. *Statistical Methods & Applications*, 28(1), 103–118. doi:/10.1007/s10260-018-00442-w.
- Bechger, T. M., & Maris, G. A. (2015) Statistical test for differential item pair functioning. *Psychometrika*, 80(2), 317-340. doi: 10.1007/s11336-014-9408-y.
- Berends, P., & Romme, G. (1999). Simulation as a research tool in management studies. *European Management Journal*, 17(6), 576-583. doi: 10.1016/S0263-2373(99)00048-1
- Bergan, J. R. (2010). *Assessing the relative fit of alternative item response theory models to the data*. Arizona: Assessment Technology Inc.
- Bilican Demir, S. (2014). *Çok Kategorili Puanlanan Maddelerde Madde İşlev Farklılığının Mantel Test ve Olabilirlik Oran Testi ile Karşılaştırılması* (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara, Türkiye.
- Bilir, M. K. (2009). *Mixture item response theory-mimic model: Simultaneous estimation of differential item functioning for manifest groups and latent classes* (Unpublised doctoral dissertation). Florida State University, Florida.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-479). Reading, MA: Addison-Wesley.
- Bishara, A. J., & Hittner, J. B. (2012) Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods*, 17(3), 399–417. doi: 10.1037/a0028087

- Blanca, M. J., Alarcon, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, 29(4), 552-557. doi: 10.7334/psicothema2016.383
- Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, 39(4), 331- 348. doi: 10.1111/j.1745-3984.2002.tb01146.x
- Bowe, A. (2017). The cultural fairness of the 12-Item general health questionnaire among diverse adolescents. *Psychological Assessment*, 29(1), 87-97. doi: 10.1037/pas0000323
- Bradley, J. V. (1978). Robustness? *The British Journal of Mathematical and Statistical Psychology*, 31(2), 144–152. doi: 10.1111/j.2044-8317.1978.tb00581.x
- Brown, J. D. (2008). Effect size and eta squared. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 12(2), 38-43. Retrieved from https://hosted.jalt.org/test/bro_28.htm
- Bulut, O., & Sünbül, Ö. (2017). Monte carlo simulation studies in item response theory with the R programming language. *Journal of Measurement and Evaluation in Education and Psychology*, 8(3), 266-287. doi: 10.21031/epod.305821
- Camilli, G. (1992). A conceptual analysis of differential item functioning in terms of a multidimensional item response model. *Applied Psychological Measurement*, 16(2), 129-147. doi: 10.1177/014662169201600203
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221-256). Westport: American Council on Education & Praeger Publishers.
- Camilli, G. (2013). Ongoing issues in test fairness. *Educational Research and Evaluation*, 19(2), 104–120. doi: 10.1080/13803611.2013.767602
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. London: Sage Publications.
- Candell, G. L., & Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement*, 12(3), 253-260. doi: 10.1177/014662168801200304
- Cao, M., Tay, L., & Liu, Y. (2017). A monte carlo study of an iterative wald test procedure for DIF analysis. *Educational and Psychological Measurement*, 77(1), 104–118. doi: 10.1177/0013164416637104.
- Carroll, H. C. (2015). *An examination of the improved wald test for differential item functioning detection with multiple groups* (Unpublished doctoral dissertation). University of Kansas.
- Carter, N. T., & Zickar, M. J. (2011). The influence of dimensionality on parameter estimation accuracy in the generalized graded unfolding model. *Educational and Psychological Measurement*, 71(5), 765-788. doi: 10.1177/0013164410387594

- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29. doi: 10.18637/jss.v048.i06
- 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. doi: 10.1111/j.1745-3984.1996.tb00496.x
- Cheema, J. R. (2019). Cross-country gender DIF in PISA science literacy items. *European Journal of Developmental Psychology*, 16(2), 152–166. doi: 10.1080/17405629.2017.1358607
- Cheng, Y., Chen, P., Qian, J., & Chang, H. H. (2013). Equated pooled booklet method in DIF testing. *Applied Psychological Measurement*, 37, 276–288. doi: 10.1177/0146621612471889
- Cho, S. J., Suh, Y., & Lee, W. Y. (2016). An NCME instructional module on latent DIF analysis using mixture item response models. *Educational Measurement: Issues and Practice*, 35(1), 48–61. doi: 10.1111/emip.12093
- Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and monte carlo simulations. *Journal of Statistical Software*, 39(8), 1–30. doi: 10.18637/jss.v039.i08
- Choi, S. W., Gibbons, L. E., & Crane, P. K. (2016). *Lordif: Logistic ordinal regression differential item functioning using IRT* [Computer software]. Retrieved from <https://CRAN.R-project.org/package=lordif>.
- Choi, Y. J., & Asilkalkan, A. (2019) R packages for item response theory analysis: descriptions and features. *Measurement: Interdisciplinary Research and Perspectives*, 17(3), 168–175, doi: 10.1080/15366367.2019.1586404.
- Chun, S. (2014). *Using MIMIC methods to detect and identify sources of DIF among multiple groups* (Unpublished master's thesis). University of South Florida.
- Chun, S., Stark, S., Kim, E. S., & Chernyshenko, O. S. (2016). MIMIC methods for detecting DIF among multiple groups: Exploring a new sequential-free baseline procedure. *Applied Psychological Measurement*, 40(7), 486–499. doi: 10.1177/0146621616659738
- Chung, C. A. (2004). *Simulation modeling handbook: A practical approach*. Boca Raton: CRC Press.
- Clark, P. C. (2010). *An examination of Type I errors and power for two differential item functioning indices* (Unpublished master's thesis). Wright State University.
- Clark, P. C., & LaHuis, D. M. (2012). An examination of power and type I errors for two differential item functioning indices using the graded response model. *Organizational Research Methods*, 15(2), 229–246. doi: 10.1177/1094428111403815

- 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. doi: 10.1111/j.1745-3992.1998.tb00619.x
- Clauser, B. E., Mazor, K. M., & 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. doi: 10.1207/s15324818ame0604_2
- Clauser, B. E., Mazor, K. M., & Hambleton, R. K. (1994). The effects of score group width on the Mantel-Haenszel procedure. *Journal of Educational Measurement*, 31(1), 67-78. doi: 10.1111/j.1745-3984.1994.tb00435.x
- Cochran, W. G. (1954). Some methods for strengthening the common χ^2 test. *Biometrics*, 10(4), 417-451. doi: 10.2307/3001616
- Cohen, A. S., & Kim, S. H. (1992). *Detecting calculator effects on a standardized mathematics test*. Madison: University of Wisconsin-Madison, Center for Placement Testing.
- Cohen, A. S., & Kim, S. H. (1993). A comparison of Lord's χ^2 and Raju's area measures in detection of. *Applied Psychological Measurement*, 17(1), 39-52. doi: 10.1177/014662169301700109
- Cohen, A. S., Kim, S. H., & Wollack, J. A. (1996). An investigation of the likelihood ratio test for detection of differential item functioning. *Applied Psychological Measurement*, 20(1), 15–26. doi: 10.1177/014662169602000102
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, L., Manion, L., & Morrison, A. K (2018). *Research methods in education* (Eight Ed.). New York, NY: Routledge.
- Cook, K. F., Bamer, A. M., Amtmann, D., Molton, I. R., & Jensen, M. P. (2012). Six patient-reported outcome measurement information system short form measures have negligible age-or diagnosis-related differential item functioning in individuals with disabilities. *Archives of Physical Medicine and Rehabilitation*, 93(7), 1289-1291. doi: 10.1016/j.apmr.2011.11.022
- Crane, P. K., Belle, G., & Larson, E. B. (2004). Test bias in a cognitive test: Differential item functioning in the CASI. *Statistics in Medicine*, 23(2), 241-256. doi: 10.1002/sim.1713
- Crane, P. K., Gibbons, L. E., Jolley, L., & van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIF detect and difwithpar. *Medic Care*, 44, 115-123. doi: 10.1097/01.mlr.0000245183.28384.ed
- Crane, P. K., Hart, D. L., Gibbons, L. E., & Cook, K. F. (2006). A 37-item shoulder functional status item pool had negligible differential item functioning. *Journal of Clinical Epidemiology*, 59(5), 478-484. doi: 10.1016/j.jclinepi.2005.10.007

- Crins, M. H. P., van der Wees, P. J., Klausch, T., van Dulmen, S. A., Roorda, L. D., & Terwee, C. B. (2018). Psychometric properties of the PROMIS Physical Function item bank in patients receiving physical therapy. *PLoS ONE* 13(2), 1-14. doi: 10.1371/journal.pone.0192187.
- Croom, S., & Svetina, M. (2021). Psychometric properties of the psychopathic personality inventory: Application to high-functioning business population. *Current Psychology*, doi: 10.1007/s12144-021-01413-3
- Cuevas, M., & Cervantes, V. H. (2012). Differential item functioning detection with logistic regression. *Mathematics and Social Science*, 3, 45-59. doi: 10.4000/msh.12274
- Çelik, M. ve Özer Özkan, Y. (2020). PISA 2015 Matematik Alt Testinin Cinsiyet ve Bölgelere Göre Değişen Madde Fonksiyonunun İncelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 11(3), 286-301. doi: 10.21031/epod.715020.
- Çepni, Z. (2011). *Değişen Madde Fonksiyonlarının Sıbttest, Mantel Haenszel, Lojistik Regresyon ve Madde Tepki Kuramı Yöntemleriyle İncelenmesi* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara, Türkiye.
- Çıkrıkçı Demirtaşlı, N. ve Ulutaş, S. (2015). A study on detecting differential item functioning of PISA 2006 science literacy items in Turkish and American samples. *Eurasian Journal of Educational Research*, 58, 41-60. doi: 10.14689/ejer.2015.58.3
- Davidson, M. J., Wortzman, B., Ko, A. J., & Li, M. (2021, March). *Investigating item bias in a CS1 exam with differential item functioning*. In Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE'21), 13–20, Virtual Event, USA.
- Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2007). Developing theory through simulation methods. *Academy of Management Review*, 32(2), 480-499. doi: 10.5465/amr.2007.24351453
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. NY: The Guildford Press.
- De Ayala, R. J., Kim, S. H., Stapleton, L. M., & Dayton, C. M. (2002). Differential item functioning: A mixture distribution conceptualization. *International Journal of Testing*, 2, 243–276. doi: 10.1080/15305058.2002.9669495
- De Gruijter, D. N. M., & van der Kamp, L. J. T. (2008). *Statistical test theory for the behavioral sciences*. Boca Raton: Taylor & Francis Group.
- De Micheaux, P. L., & Tran, V. A. (2016). Power: A reproducible research tool to ease monte carlo power simulation studies for goodness-of-fit tests in R. *Journal of Statistical Software*, 69(3), 1-44. doi: 10.18637/jss.v069.i03

- Dell-Ross, T. L. (2021). *Investigating the performance of (multiple-factor) multiple-group methods for the detection of differential item functioning* (Unpublished doctoral dissertation). Georgia State University, Atlanta.
- DeMars, C. E. (2003). Sample size and recovery of nominal response model item parameters. *Applied Psychological Measurement*, 27(4), 275-288. doi: 10.1177/0146621603027004003
- DeMars, C. E. (2010). *Item response theory*. New York: Oxford University Press, Inc.
- DeMars, C. E., & Lau, A. (2011). Differential item functioning detection with latent classes: How accurately can we detect who is responding differentially? *Educational and Psychological Measurement*, 71(4), 597-616. doi: 10.1177/0013164411404221
- Demir, E. (2018). As a potential source of error, measuring the tendency of university students to copy the answers: A scale development study. *Eurasian Journal of Educational Research*, 75, 37-58. doi: 10.14689/ejer.2018.75.3
- Demir, S. ve Köse, İ. A. (2014). Mantel-Haenszel, SIBTEST ve Lojistik Regresyon Yöntemleri ile Değişen Madde Fonksiyonunu Analizi. *Journal of Human Sciences*, 11(1), 700-714. doi: 10.14687/ijhs.v11i1.2798
- Descôteaux, J. (2007). Statistical power: An historical introduction. *Tutorials in Quantitative Methods for Psychology*, 3(2), 28-34. doi: 10.20982/tqmp.03.2.p028
- Desjardins, C. D., & Bulut, O. (2018). *Handbook of educational measurement and psychometrics using R*. Boca Raton: Taylor & Francis Group.
- Dodeen, H. (2004). Stability of differential item functioning over a single population in survey data. *Journal of Experimental Education*, 72(3), 181-193. doi: 10.3200/JEXE.72.3.181-193
- Dodeen, H., & Johanson, G. A. (2003). An analysis of sex-related differential item functioning in attitude assessment. *Assessment & Evaluation in Higher Education*, 28(2), 129-134. doi: 10.1080/02602930301667
- Doğan, N. ve Öğretmen, T. (2008). Değişen Madde Fonksiyonunu Belirlemede Mantel-Haenszel, Ki-Kare ve Lojistik Regresyon Tekniklerinin Karşılaştırılması. *Eğitim ve Bilim*, 33(148), 100-112.
- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. P. W. Holland, and H. Wainer (Eds.). *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Erlbaum.
- Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23(4), 355-368. doi: 10.1111/j.1745-3984.1986.tb00255.x

- Drabinová, A., & Martinková, P. (2017). Detection of differential item functioning with non-linear regression: A non-IRT approach accounting for guessing. *Journal of Educational Measurement*, 54(4), 498-517. doi: 10.1111/jedm.12158
- Ellis, B. B., & Kimmel, H. (1992). Identification of unique cultural response patterns by means of item response theory. *Journal of Applied Psychology*, 77(2), 177-184. doi: 10.1037/0021-9010.77.2.177
- Ellis, B. B., & Mead, A. D. (2000). Assessment of the measurement equivalence of a Spanish translation of the 16PF questionnaire. *Educational and Psychological Measurement*, 60(5), 787-807. doi: 10.1177/00131640021970781
- Ellis, B. B., & Raju, N. S. (2003). *Test and item bias: What they are, what they aren't, and how to detect them*. ERIC Document ED 480042. Rockville, MD: Educational Resources Information Center.
- Ellis, S. M., & Steyn, H. S. (2003). Practical significance (effect sizes) versus or in combination with statistical significance (p-values). *Management Dynamics*, 12(4), 51-53.
- Elosua, P., & Wells, C. S. (2013). Detecting DIF in polytomous items using MACS, IRT and ordinal logistic regression. *Psicológica*, 34(2), 327-342.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Ercikan, K., & Koh, K. (2005). Examining the construct comparability of the English and French versions of TIMSS. *International Journal of Testing*, 5(1), 23-35. doi: 10.1207/s15327574ijt0501_3
- Fallon, L. M., Cathcart, S. C., & Johnson, A. H. (2021). Assessing differential item functioning in a teacher self-assessment of cultural responsiveness. *Journal of Psychoeducational Assessment*, 39(7), 816-831. doi: 10.1177/07342829211026464
- Fan, X. (2001). Statistical significance and effect size in education research: two sides of a coin. *The Journal of Educational Research*, 94(5), 275-282. doi: 10.1080/00220670109598763
- Farrington, A. L., & Lonigan, C. J. (2015). Examining the measurement precision and invariance of the revised get ready to read! *Journal of Learning Disabilities*, 48(3), 227-238. doi: 10.1177/0022219413495568
- Fay, K., & Boyd, M. J. (2010). Eta-squared. In N. J. Salkind (Eds.), *Encyclopedia of Research Design* (pp. 423-425). Thousand Oaks: SAGE Publications, Inc.
- Feinberg, R. A., & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice*, 35(2), 36-49. doi:10.1111/emip.12111.

- Feuerherd, M., Knuth, D., Muehlan, H., & Schmidt, S. (2014). Differential item functioning (DIF) analyses of the impact of event scale-revised (IES-R): Results from a large european study on people with disaster experiences. *Traumatology*, 20(4), 313-320. doi: 10.1037/h0099858
- Fidalgo, A. M. (2011). A new approach for differential item functioning detection using Mantel-Haenszel methods. The GMGDIF program. *The Spanish Journal of Psychology*, 14(2), 1018-1022. doi: 10.5209/rev_sjop.2011.v14.n2.47
- Fidalgo, A. M., & Madeira, J. M. (2008). Generalized Mantel-Haenszel methods for differential item functioning detection. *Educational and Psychological Measurement*, 68(6), 940-958. doi: 10.1177/0013164408315265
- Fidalgo, A. M., & Scalón, J. D. (2010). Using generalized Mantel-Haenszel statistics to assess DIF among multiple groups. *Journal of Psychoeducational Assessment*, 28(1), 60-69. doi: 10.1177/0734282909337302
- Fidalgo, A. M., Ferreres, D., & Muniz, J. (2004a). Liberal and conservative differential item functioning detection using Mantel-Haenszel and SIBTEST: Implications for Type I and Type II error rates. *Journal of Experimental Education*, 73(1), 23-39. doi: 10.3200/JEXE.71.1.23-40
- Fidalgo, A. M., Ferreres, D., & Muniz, J. (2004b). Utility of the Mantel-Haenszel procedure for detecting differential item functioning in small samples. *Educational and Psychological Measurement*, 64(6), 925-936. doi: 10.1177/0013164404267288
- Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE.
- Fikis, D. R. J., & Oshima, T. C. (2017). Effect of purification procedures on DIF analysis in IRTPRO. *Educational and Psychological Measurement*, 77(3), 415-428. doi:10.1177/0013164416645844.
- Finch, W. H., Immekus, J. C., & French, B. F. (2016). *Applied psychometrics using SPSS and AMOS*. Charlotte, NC: Information Age Publishing Inc.
- 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. doi: 10.1177/0146621605275728
- Finch, W. H. (2012). The MIMIC model as a tool for differential bundle functioning detection. *Applied Psychological Measurement*, 36(1), 40-59. doi: 10.1177/0146621611432863
- Finch, W. H. (2016). Detection of item functioning for more than two groups: A monte carlo comparison of methods. *Applied Measurement in Education*, 29(1), 30-45. doi: 10.1080/08957347.2015.1102916
- Finch, W. H., & French, B. F. (2007). Detection of crossing differential item functioning: A comparison of four methods. *Educational and Psychological Measurement*, 67(4), 565-582. doi: 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(5), 742–759. doi: 10.1177/0013164407313370
- Finch, W. H., & French, B. F. (2015). Modeling of nonrecursive structural equation models with categorical indicators. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(3), 416–428. doi: 10.1080/10705511.2014.937380.
- Finch, W. H., & French, B. F. (2019). *Educational and psychological measurement*. New York: Taylor & Francis.
- Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology*, 10, 1–18. doi: 10.3389/fpsyg.2019.01507
- Fleishman, J., Spector, W., & Altman, B. (2002). Impact of differential item functioning on age and gender differences in functional disability. *The Journal of Gerontology Social Sciences*, 57(5), 275–284. doi: 10.1093/geronb/57.5.S275
- Flowers, C. P., Oshima, T. C., & Raju, N. S. (1999). A description and demonstration of the Polytomous-DFIT framework. *Applied Psychological Measurement*, 23(4), 309–326. doi: 10.1177/01466219922031437
- Frederickx, S., Tuerlinckx, F., De Boeck, P., & Magis, D. (2010). RIM: A random item mixture model to detect differential item functioning. *Journal of Educational Measurement*, 47(4), 432–457. doi: 10.1111/j.1745-3984.2010.00122.x
- French, A. W., & Miller, T. R. (1996). Logistic regression and its use in detecting differential item functioning in polytomous items. *Journal of Educational Measurement*, 33(3), 315–332. doi: 10.1111/j.1745-3984.1996.tb00495.x
- 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. doi: 10.1177/0013164406294781
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2–18. doi: 10.1037/a0024338
- Gamerman, D., Gonçalves, F. B., & Soares, T. M. (2018). Differential item functioning. W. J. van der Linden (Eds). *Handbook of item response theory* (vol.3, pp. 67-84). Boca Raton, FL: CRC Press.
- Gao, X. (2019). *A comparison of six DIF detection methods* (Unpublished master's thesis). University of Connecticut Graduate School.
- Gelin, M. N., & Zumbo, B. D. (2003). Differential item functioning results may change depending on how an item is scored: An illustration with the center for epidemiologic studies depression scale. *Educational and Psychological Measurement*, 63(1), 65–74. doi: 10.1177/0013164402239317

- Gelin, M. N. (2001). *Type I error rates of the DIF MIMIC approach using Joreskog's covariance matrix with ML and WLS estimation* (Unpublished doctoral dissertation). The University of British Columbia.
- George, D., & Mallery, P. (2020). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge: Taylor & Francis Group.
- Geramipour, M., & Shahmirzadi, N. (2019). A gender-related differential item functioning study of an English test. *The Journal of Asia TEFL*, 16(2), 674-682. doi: 10.18823/asiatefl.2019.16.2.15.674
- Gierl, M. J. (2005). Using dimensionality based DIF analysis to identify and interpret constructs that elicit group differences. *Educational Measurement Issues and Practice*, 24(1), 3-13. doi: 10.1111/j.1745-3992.2005.00002.x
- Gierl, M. J., & Bolt, D. M. (2001). Illustrating the use of nonparametric regression to assess differential item and bundle functioning between multiple groups. *International Journal of Testing*, 1, 249-270. doi: 10.1080/15305058.2001.9669474
- Gierl, M. J., Bisanz, J., Bisanz, G. L., Boughton, K. A., & Khaliq, S. N. (2001). Illustrating the utility of differential bundle functioning analysis to identify and interpret group differences on achievement tests. *Educational Measurement: Issues and Practices*, 20(2), 26-36. doi: 10.1111/j.1745-3992.2001.tb00060.x
- Gierl, M., Gotzmann, A., & Boughton, K. A. (2004). Performance of SIBTEST when the percentage of DIF items is large. *Applied Measurement in Education*, 17(3), 241-264. doi: 10.1207/s15324818ame1703_2
- Gilbert, N. (1999). Simulation: A new way of doing social science. *American Behavioral Scientist*, 42(10), 1485-1487. doi:10.1177/0002764299042010002.
- Gierl, M. J. & Khaliq, S. N. (2001). Identifying sources of differential item and bundle functioning on translated achievement tests: A confirmatory analysis. *Journal of Educational Measurement*, 38(2), 164-187. doi: 10.1111/j.1745-3984.2001.tb01121.x
- Glas, C., & Jehangir, K. (2013). Modeling Country-specific differential item functioning. In L. Rutkowski, M. Von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 97-116). London: Chapman Hall/CRC Press.
- Glass, G. V, Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumption underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, 42(3), 237-288. doi: 10.3102/00346543042003237
- González-Romá, V., Hernández, A., & Gómez-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. doi: 10.1207/s15327906mbr4101_3

- Goodman, J., Wilse, J., Allen, N., & Klaric, J. (2011). Identification of differential item functioning in assessment booklet designs with structurally missing data. *Educational and Psychological Measurement*, 71(1), 80–94. doi: 10.1177/0013164410387341
- Gök, B., Atalay Kabasakal, K. ve Kelecioğlu, H. (2014). PISA 2009 Öğrenci Anketi Tutum Maddelerinin Kültüre Göre Değişen Madde Fonksiyonu Açısından İncelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(1), 72-87. doi: 10.21031/epod.64124
- Gök, B., Kelecioğlu, H. ve Doğan, N. (2010). Değişen Madde Fonksiyonunu Belirlemede Mantel–Haenszel ve Lojistik Regresyon Tekniklerinin Karşılaştırılması. *Eğitim ve Bilim*, 35(156), 3-16.
- Gravetter, F. J., & Wallnau, L. B. (2012). *Statistics for the behavioral sciences* (9th ed.), Belmont, CA: Wadsworth/Cengage Learning.
- Gray, C. D., & Kinnear, P. R. (2012). *IBM SPSS statistics 19 made simple*. New York: Psychology Press, Taylor & Francis Group.
- Gregory, J. J., Werth, P. M., Reilly, C. A., & Jevsevar, D. S. (2021). Cross-specialty PROMIS-global health differential item functioning. *Quality of Life Research*, 30(8), 2339-2348. doi: 10.1007/s11136-021-02812-6
- Gu, F. (2016). A review of Monte Carlo simulation and resampling methods for social science. *Journal of Educational and Behavioral Statistics*, 41(4), 448-450. doi: 10.3102/1076998615625963
- Guilera, G., Gómez-Benito, J., Hidalgo, M. D., & Sánchez-Meca, J. (2013). Type I error and statistical power of the Mantel-Haenszel procedure for detecting DIF: A meta-analysis. *Psychological Methods*, 18(4), 553-571. doi: 10.1037/a0034306
- Güler, N., & Penfield, R. D. (2009). A comparison of logistic regression and contingency table methods for simultaneous detection of uniform and nonuniform DIF. *Journal of Educational Measurement*, 46(3), 314-329. doi: 10.1111/j.1745-3984.2009.00083.x
- Gómez-Benito, J., Sireci, S., Padilla, J., Hidalgo, M. D., & Benítez, I. (2018). Differential item functioning: Beyond validity evidence based on internal structure. *Psicothema*, 30(1), 104-109. doi: 10.7334/psicothema2017.183
- Hahn, E. A., Kallen, M. A., Jacobs, E. A., Ganschow, P. S., Garcia, S. F., & Burns, J. L. (2014). English-Spanish equivalence of the health literacy assessment using talking touchscreen technology (Health LiTT). *Journal of Health Communication*, 19(2), 285-301. doi: 10.1080/10810730.2014.936567
- Hair J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (eighth edition). United Kingdom: Cengage Learning, EMEA.
- Hallgren, K. A. (2013). Conducting simulation studies in the R programming environment. *Tutorials in Quantitative Methods for Psychology*, 9(2), 43–60. doi: 10.20982/tqmp.09.2.p043

- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage Publication.
- Hambleton, R. K. (2006). Good practices for identifying differential item functioning. *Medical Care*, 44(11), 182-188. doi: 10.1097/01.mlr.0000245443.86671.c4
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and application*. Boston: Kluwer-Nijhoff Publishing.
- Hanson, B. A. (1998). Uniform DIF and DIF defined by differences in item response functions. *Journal of Educational and Behavioral Statistics*, 23(3), 244-253. doi: 10.3102/10769986023003244
- Happach, R. M., & Tilebein, M. (2015). Simulation as research method: Modeling social interactions in management science. In C. Misselhorn (Eds.), *Collective agency and cooperation in natural and artificial systems* (pp. 239-259). New York: Springer
- Harwell, M. (2018). A strategy for using bias and RMSE as outcomes in monte carlo studies in statistics. *Journal of Modern Applied Statistical Methods*, 17(2), 29-38. doi: 10.22237/jmasm/1551907966
- Harwell, M. R., Kohli, N., & Peralta, Y. (2017). Experimental design and data analysis in computer simulation studies in the behavioral sciences. *Journal of Modern Applied Statistical Methods*, 16(2), 3-28. doi: 10.22237/jmasm/1509494520
- Harwell, M. R., Kohli, N., & Peralta-Torres, Y. (2018). A survey of reporting practices of computer simulation studies in statistical research. *The American Statistician*, 72(4), 321-327. doi: 10.1080/00031305.2017.1342692
- Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing monte carlo results in methodological research: The one-and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics*, 17, 315-339. doi:10.2307/1165127.
- Harwell, M. R., Stone, C. A., Hsu, T., & Kirisci, L. (1996). Monte carlo studies in item response theory. *Applied Psychological Measurement*, 20(2), 101-125. doi: 10.1177/014662169602000201
- Hidalgo, M. D., & Lopez-Pina, J. A. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement*, 64(6), 903-915. doi: 10.1177/0013164403261769
- Hidalgo, M. D., Gómez-Benito, J., & Zumbo, B. D. (2014). Binary logistic regression analysis for detecting differential item functioning: Effectiveness of R^2 and delta log odds ratio effect size measures. *Educational and Psychological Measurement*, 74(6), 927-949. doi: 10.1177/0013164414523618
- Hladká, A., & Martinková, P. (2020). difNLR: Generalized logistic regression models for DIF and DDF detection. *The R Journal*, 12(1), 300-323. doi: 10.32614/RJ-2020-014

- Hofert, M., & Mächler, M. (2016). Parallel and other simulations in R made easy: An end-to-end study. *Journal of Statistical Software*, 69(4), 1-44. doi:10.18637/jss.v069.i04.
- Holland, P. W., & Thayer, D. T. (1988). Differential item functioning and the Mantel-Haenszel procedure. H. Wainer, and H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum.
- Hong, I., Yoo, E., Kazley, A. S., Lee, D., Li C., & Reistetter, T. A. (2018). Development and validation of the activities of daily living short form for community-dwelling korean stroke survivors. *Evaluation & The Health Professions*, 41(1), 44-66. doi: 10.1177/0163278717695863
- Hope, D., Adamson, K., McManus, I. C., Chris, L., & Elder, A. (2018). Using differential item functioning to evaluate potential bias in a high stakes postgraduate knowledge based assessment. *BMC Medical Education*, 18, 1-7. doi:10.1186/s12909-018-1143-0
- Hopkins, B. (1973). Educational research and Type III errors. *The Journal of Experimental Education*, 41(4), 31-32. doi: 10.1080/00220973.1973.11011423
- Howell, D. C. (2014). *Fundamental statistics for the behavioral sciences*. Belmont, CA: Cengage Learning.
- Huang, X., Wilson, M., & Wang, L. (2016). Exploring plausible causes of differential item functioning in the PISA science assessment: Language, curriculum or culture. *Educational Psychology*, 36(2), 378–390. doi: 10.1080/01443410.2014.946890.
- Huggins, A. C. (2014). The effect of differential item functioning in anchor items on population invariance of equating. *Educational and Psychological Measurement*, 74(4), 627-658. doi: 10.1177/0013164413506222
- Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: A monte carlo study. *Applied Psychological Measurement*, 6, 249–260. doi: 10.1177/014662168200600301.
- Iwata, N., Turner, R. J., & Lloyd, D. A. (2002). Race/ethnicity and depressive symptoms in community-dwelling young adults: A differential item functioning analysis. *Psychiatry Research*, 110(3), 281–289. doi: 10.1016/s0165-1781(02)00102-6
- Jafari, P., Allahyari, E., Salarzadeh, M., & Bagheri, Z. (2015). Item-level informant discrepancies across obese–overweight children and their parents on the PedsQL™ 4.0 instrument: An iterative hybrid ordinal logistic regression. *Quality of Life Research*, 25(1), 25-33. doi: 10.1007/s11136-015-1046-z
- Jiang, H., & Stout, W. (1998). Improved Type I error control and reduced estimation bias for DIF detection using SIBTEST. *Journal of Educational and Behavioral Statistics*, 23(4), 291-322. doi: 10.2307/1165279

- Jin, K., & Chen, H. (2020). MIMIC approach to assessing differential item functioning with control of extreme response style. *Behavior Research Methods*, 52(1), 23-35. doi: 10.3758/s13428-019-01198-1
- Jin, K., Chen, H., & Wang, W. (2015). Assessing differential item functioning in multiple grouping variables with factorial logistic regression. In Millsap, Bolt, Ark & Wang, (Eds.), *Quantitative psychology research* (pp. 243-259). New York: Springer International Publishing.
- Jin, K., Chen, H., & Wang, W. (2018). Using odds ration to detect differential item functioning. *Applied Psychological Measurement*, 42(8), 613-629. doi: 10.1177/0146621618762738
- 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(4), 329-349. doi: 10.1207/S15324818AME1404_2
- Kamata, A., & Vaughn, B. K. (2004). An introduction to differential item functioning analysis. *Learning Disabilities: A Contemporary Journal*, 2(2), 49-69.
- Kan, A., Sünbül, Ö. ve Ömür, S. (2013). 6.-8. Sınıf Seviye Belirleme Sınavları Alt Testlerinin Çeşitli Yöntemlere Göre Değişen Madde Fonksiyonlarının İncelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 9(2), 207-222. doi: 10.17860/efd.55452
- Kanjee, A. (2007). Using logistic regression to detect bias when multiple groups are tested. *South African Journal of Psychology*, 37(1), 47-61. doi: 10.1177/008124630703700104
- Karakaya, İ. (2012). Seviye Belirleme Sınavındaki Fen ve Teknoloji ile Matematik Alt Testlerinin Madde Yanlılığı Açısından İncelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, 12(1), 215-229.
- Karami, H., & Nodoushan, M. A. S. (2011). Differential item functioning (DIF): Current problems and future directions. *International Journal of Language Studies (IJLS)*, 5(3), 133-142. Retrieved from <https://eric.ed.gov/?id=ED521872>
- Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson.
- Kim, J. (2010). *Controlling Type 1 error rate in evaluating differential item functioning for four DIF methods: Use of three procedures for adjustment of multiple item testing* (Unpublished doctoral dissertation). Georgia State University, Atlanta.
- Kim, J., & Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement*, 73(3), 458-470. doi: 10.1177/0013164412467033
- Kim, S. H., & Cohen, A. S. (1992). Effects of linking methods on detection DIF. *Journal of Educational Measurement*, 29(1), 51-66. doi: 10.1111/j.1745-3984.1992.tb00367.x

- Kim, S. H., & Cohen, A. S. (1995). A comparison of Lord's chi-square, Raju's area measures, and the likelihood ratio test on detection of differential item functioning. *Applied Psychological Measurement Education*, 8(4), 291-312. doi: 10.1207/s15324818ame0804_2
- Kim, S. H., Cohen, A. S., & Kim, H. A. (1994). An investigation of Lord's procedure for the detection of differential item functioning. *Applied Psychological Measurement*, 18(3), 217-228. doi: 10.1177/014662169401800303
- Kim, S. H., Cohen, A. S., & Park, T. H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement*, 32(3), 261-276. doi: 10.1111/j.1745-3984.1995.tb00466.x
- Kim, S. H., Cohen, A. S., Alagoz, C., & Kim, S. (2007). DIF detection and effect size measures for polytomously scored items. *Journal of Educational Measurement*, 44(2), 93-116. doi: 10.1111/j.1745-3984.2007.00029.x
- Kim, W. (2003). *Development of a differential item functioning (DIF) procedure using the hierarchical generalized linear model: A comparison study with logistic regression procedure* (Unpublished doctoral dissertation). The Pennsylvania State University, Pennsylvania.
- Kimball, A. W. (1957). Errors of the third kind in statistical consulting. *Journal of the American Statistical Association*, 52(278), 133-142. doi: 10.1080/01621459.1957.10501374
- Kirk, R. E. (2008). *Statistics an introduction* (Fifth edition). United States: Thomson Higher Education.
- Kleijnen, J. P. C. (1995). Theory and methodology verification and validation of simulation models. *European Journal of Operational Research* 82(1), 145-162. doi: 10.1016/0377-2217(94)00016-6
- Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, 53(6), 2576-2590. doi: 10.3758/s13428-021-01587-5.
- Kopf, J., Zeileis, A., & Strobl, C. (2015a). A framework for anchor methods and an iterative forward approach for DIF detection. *Applied Psychological Measurement*, 39(2), 83-103. doi: 10.1177/0146621614544195.
- Kopf, J., Zeileis, A., & Strobl, C. (2015b). Anchor selection strategies for DIF analysis: Review, assessment, and new approaches. *Educational and Psychological Measurement*, 75(1), 22-56. doi: 10.1177/0013164414529792.
- Korkmaz, M. (2005). *Madde Cevap Kuramına Dayalı Olarak Çok Kategorili Maddelerde Madde ve Test Yanlılığının (İşlevsel Farklılığın) İncelenmesi* (Yayımlanmamış Doktora Tezi). Ege Üniversitesi, İzmir, Türkiye.
- Kristjansson, E. A., Desrochers, A., & Zumbo, B. D. (2003). Translating and adapting measurement instruments for cross-cultural research: A guide for practitioners. *Canadian Journal of Nursing Research*, 35(2), 127-142.

- Kristjansson, E., Aylesworth, R., McDowell, I., & Zumbo, B. D. (2005). A comparison of four methods for detecting DIF in ordered response items. *Educational and Psychological Measurement*, 65(6), 935-953. doi: 10.1177/0013164405275668
- Kulick, E., & Hu, P. G. (1989). *Examining the relationship between differential item functioning and item difficulty* (College Board Report No. 89-5; ETS RR-89-18). New York, NY: College Entrance Examination Board.
- Kuzucu, Y. ve Sariot Ertürk, Ö. (2020). Psychometric properties of Turkish version of aggression questionnaire short form: Measurement invariance and differential item functioning across sex and age. *Journal of Measurement and Evaluation in Education and Psychology* 11(3), 243-265. doi: 10.21031/epod.683176
- Lai, J. S., Teresi, J., & Gershon, R. (2005). Procedures for the analysis of differential item functioning (DIF) for small sample sizes. *Evaluation and The Health Professions*, 28(3), 283-294. doi: 10.1177/0163278705278276
- Lakens., D. (2013) Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology* 4, 1–12. doi: 10.3389/fpsyg.2013.00863
- Lan, M. C. (2014). *Exploring gender differential item functioning (DIF) in eight grade mathematics items for the United States and Taiwan* (Unpublished doctoral dissertation). University of Washington.
- Landis, J. R., Heyman, E. R., & Koch, G. G. (1978). Average partial association in three-way contingency tables: A review and discussion of alternative tests. *International Statistical Review*, 46(3), 237-254. doi :10.2307/1402373
- Langer, M. (2008). *A reexamination of Lord's wald test for differential item functioning using item response theory and modern error estimation* (Unpublised doctoral dissertation). University of North Carolina, Chapel Hill.
- Law, A. M. (2003). *How to conduct a successful simulation study*. Proceedings of the 2003 Winter Simulation Conference, New Orleans, LA. U.S.A.
- Le, L. (2009). Investigating gender differential item functioning across countries and test languages for PISA science items. *International Journal of Testing*, 9(2), 122–133. doi: 10.1080/15305050902880769
- Lee, H., & Geisinger, K. F. (2016). The matching criterion purification for differential item functioning analyses in a large-scale assessment. *Educational and Psychological Measurement*, 76(1), 141-163. doi: 10.1177/0013164415585166
- Lee, S., Bulut, O., & Suh, Y. (2016). Multidimensional extension of multiple indicators multiple causes models to detect DIF. *Educational and Psychological Measurement*, 77(4), 545-569. doi: 10.1177/0013164416651116.
- Lee, Y., Cohen, A., & Toro, M. (2009). Examining Type I error and power for detection of differential item and testlet functioning. *Asia Pacific Education Review*, 10(3), 365-375. doi: 10.1007/s12564-009-9039-7

- Lei, P. W., & Li, H. (2013). Small-sample DIF estimation using SIBTEST, Cochran's Z, and log-linear smoothing. *Applied Psychological Measurement*, 37(5), 397-416. doi: 10.1177/0146621613478150
- Leventhal, L. (1999). Updating the debate on one-versus two-tailed test with the directional two-tailed test. *Psychological Reports*, 84(3), 707-718. doi: 10.2466/pr0.1999.84.3.707
- Leventhal, L., & Huynh, C. L. (1996). Directional decisions for two-tailed tests: Power, error rates, and sample size. *Psychological Methods*, 1(3), 278-292. doi: 10.1037/1082-989X.1.3.278
- León, J. A., Martínez-Huertas, J. A., Olmos, R., Moreno, J. D., & Escudero, I. (2019). Metacomprehension skills depend on the type of text: An analysis from Differential Item Functioning. *Psicothema*, 31(1), 66-72. doi: 10.7334/psicothema2018.163
- Li, H., & Stout, W. A. (1996). A new procedure for detection of crossing DIF. *Psychometrika*, 61(4), 647-677. doi: 10.1007/BF02294041
- 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. doi: 10.1177/0013164411432333
- Li, Y., Cohen, A. S., & Ibarra, R. A. (2004). Characteristics of mathematics items associated with gender DIF. *International Journal of Testing*, 4(2), 115-136. doi: 10.1207/s15327574ijt0402_2
- Li, Y., She, M., Tu, D., & Cai, Y. (2021). Computerized adaptive testing for schizotypal personality disorder: Detecting individuals at risk. *Frontiers in Psychology*, 11. doi: 10.3389/fpsyg.2020.574760
- Li, Z., & Zumbo, B. D. (2009). Impact of differential item functioning on subsequent statistical conclusions based on observed test score data. *Psicológica*, 30(2), 343-370.
- Linn, R. L. (1993). The use of differential item functioning statistics: A discussion of current practice and future implications. P. W. Holland and H. Wainer (Eds.). *Differential item functioning* (pp. 349-364). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Liu, H., Zhang, Y., & Luo, F. (2015). Mediation analysis for ordinal outcome variables. In Millsap, Bolt, Ark & Wang, (Eds.), *Quantitative psychology research* (pp. 429-450). New York: Springer International Publishing.
- Liu, J., & Dorans, J. N. (2013). Assessing a critical aspect of construct continuity when test specifications change or test forms deviate from specifications. *Educational Measurement: Issues and Practice*, 32(1), 15-22. doi: 10.1111/emip.12001

- Lopez Rivas, G. E. (2012). *Detection and classification of DIF types using parametric and nonparametric methods: A comparison of the IRT-likelihood ratio test, crossing-SIBTEST, and logistic regression procedures* (Unpublished doctoral dissertation). University of South Florida, Florida.
- Lopez Rivas, G. E., Stark, S., & Chernyshenko, O. S. (2009). The effects of referent item parameters on differential item functioning detection using the free baseline likelihood ratio test. *Applied Psychological Measurement*, 33(4), 251-265. doi: 10.1177/0146621608321760
- 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 and Zeitlinger.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Luo, Y., & Liang, X. (2019). Simultaneously modeling differential testlet functioning and differential item functioning: Addressing variance heterogeneity with a multigroup one-parameter testlet model. *Measurement: Interdisciplinary Research and Perspectives*, 17(2), 93-105. doi: 10.1080/15366367.2018.1533783.
- Lyons-Thomas, J., Sandilands, D. D., & Ercikan, K. (2014). Gender differential item functioning in mathematics in four international jurisdictions. *Education and Science*, 39(172), 20-32.
- MacDonald, P. (1999). Power, Type I, and Type III Error rates of parametric and non-parametric statistical tests. *The Journal of Experimental Education*, 67(4), 367-379. doi: 10.1080/00220979909598489
- Magis, D., & De Boeck, P. (2011). Identification of differential item functioning in multiple-group settings: A multivariate outlier detection approach. *Multivariate Behavioral Research*, 46, 733-755. doi:10.1080/00273171.2011.606757.
- Magis, D., & Facon, B. (2013). Item purification does not always improve DIF detection: A counterexample with Angoff's delta plot. *Journal of Educational and Psychological Measurement*, 73(2), 293-311. doi: 10.1177/0013164412451903
- Magis, D., Béland, S., & Raiche, G. (2016). *diffR: Collection of methods to detect dichotomous differential item functioning (DIF)*. R package version 4.7.
- Magis, D., Béland, S., Tuerlinckx, F., & De 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. doi: 10.3758/BRM.42.3.847
- Magis, D., Raïche, G., Béland, S., & Gérard, P. (2011). A generalized logistic regression procedure to detect differential item functioning among multiple groups. *International Journal of Testing*, 11(4), 365-386. doi: 10.1080/15305058.2011.602810

- Magis, D., Tuerlinckx, F., & De Boeck, P. (2015). Detection of differential item functioning using the lasso approach. *Educational and Behavioral Statistics*, 40(2), 111–135. doi: 10.3102/1076998614559747
- Mair, P. (2018). *Modern psychometrics with R*. Cambridge: Springer.
- Maller, S. J. (2001). Differential item functioning in the WISC-III: Item parameters for boys and girls in the national standardization sample. *Educational and Psychological Measurement*, 61(5), 793-817. doi: 10.1177/00131640121971527
- Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel–Haenszel procedure. *Journal of the American Statistical Association*, 58(303), 690–700. doi: 10.1080/01621459.1963.10500879
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4), 719-748. doi: 10.1093/jnci/22.4.719
- Mapuranga, R., Dorans, N. J., & Middleton, K. (2008). *A review of recent developments in differential item functioning* (Research Report No. ETS RR-08-43). Princeton, NJ: ETS.
- Martinková, P., Drabinová, A., Liaw, Y., Sanders, E. A., McFarland, J. L., & Price, R. M. (2017). Checking equity: Why differential item functioning analysis Should be a routine part of developing conceptual assessments. *CBE—Life Sciences Education*, 16(2), 1-13. doi:10.1187/cbe.16-10-0307.
- Masyn, K. E. (2017). Measurement invariance and differential item functioning in latent class analysis with stepwise multiple indicator multiple cause modeling. *Structural Equation Modeling*, 24(2), 180–197. doi: 10.1080/10705511.2016.1254049.
- Mazor, K. M., Clauser, R. E., & Hambleton, R. K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel procedure. *Educational and Psychological Measurement*, 54(2), 284–291. doi: 10.1177/0013164494054002003
- Mcnamara, T., & Roever, C. (2006). Psychometric approaches to fairness: Bias and DIF. *Language Learning*, 56(S2), 81-128. doi: 10.1111/j.1467-9922.2006.00381.x
- Meade, A. W., & Wright, N. A. (2012). Solving the measurement invariance anchor item problem in item response theory. *Journal of Applied Psychology*, 97(5), 1016-1031. doi: 10.1037/a0027934
- Meade, A. W., Lautenschlager, G. J., & Johnson, E. C. (2007). A monte carlo examination of the sensitivity of the differential functioning of items and tests framework for tests of measurement invariance with Likert data. *Applied Psychological Measurement*, 31(5), 430-455. doi: 10.1177/0146621606297316

- Meij, A. M. M., Kelderman, H., & van der Flier, H. (2010). Improvement in detection of differential item functioning using a mixture item response theory model. *Multivariate Behavioral Research*, 45(6), 975–999. doi: 10.1080/00273171.2010.533047
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics*, 7(2), 105–118. doi: 10.3102/10769986007002105
- Mendes-Barnett, S., & Ercikan, K. (2006). Examining sources of gender DIF in mathematics assessments using a confirmatory multidimensional model approach. *Applied Measurement in Education*, 19(4), 289–304. doi: 10.1207/s15324818ame1904_4
- Miller, T. R., & Spray, J. A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement*, 30(2), 107–122. doi: 10.1111/j.1745-3984.1993.tb01069.x
- Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, 17(4), 297–334. doi: 10.1177/014662169301700401
- Monahan, P. O., & Ankenmann, R. D. (2005). Effect of unequal variances in proficiency distributions on Type-I error of the Mantel-Haenszel chi-square test for differential item functioning. *Journal of Educational Measurement*, 42(2), 101–131. doi: 10.1111/j.1745-3984.2005.00006
- Mooney, C. Z. (1997). *Monte carlo simulation*. Thousand Oaks, CA: Sage.
- Morales, L. S, Reise, S. P., & Hays, R. D. (2000). Evaluating the equivalence of health care ratings by Whites and Hispanics. *Medical Care*, 38(5), 517–527. doi: 10.1097/00005650-200005000-00008
- Morales-Ortiz, M., & Peña-Cardenas, F. (2019). DIF in the Spanish version of the verbal selective reminding test using samples from Hispanics in the United States, Mexicans, and Spaniards. *Frontiers in Psychology*, 10, 1–8. doi: 10.3389/fpsyg.2019.02687.
- Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Tutorial in Biostatistics*, 38(11), 2074–2102. doi: 10.1002/sim.8086
- Mosteller, F. (1948). A k-sample slippage test for an extreme population. *The Annals of Mathematical Statistics*, 19, 58–65. Retrieved from https://link.springer.com/chapter/10.1007/978-0-387-44956-2_5
- Mucherah, W., Finch, W. H., & Keaitse, S. (2012). Differential bundle functioning analysis of the self-description questionnaire self-concept scale for Kenyan female and male students using the MIMIC model. *International Journal of Testing*, 12(1), 78–99. doi: 10.1080/15305058.2011.620724

- Mundform, D. J., Schaffer, J., Kim, M. J., Shaw, D., Thongteeraparp, A., Preecha, C., & Supawan, P. (2011). Number of replications required in monte carlo simulation studies: A synthesis of four studies. *Journal of Modern Applied Statistical Methods*, 10(1), 19-28. doi: 10.22237/jmasm/1304222580
- Muñiz, J., Hambleton, R. K., & Xing, D. (2001). Small sample studies to detect flaws in item translations. *International Journal of Testing*, 1(2), 115-135. doi: 10.1207/S15327574IJT0102_2
- Muraki, E. (1992). A generalized partial credit model: Applications of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159-176. doi: 10.1177/014662169201600206
- Muraki, E. (1999). Stepwise analysis of differential item functioning based on multiple-group partial credit model. *Journal of Educational Measurement*, 36(3), 217-232. doi: 10.1111/j.1745-3984.1999.tb00555.x
- Narayanan, P., & Swaminathan, H. (1996). Identification of items that nonuniform DIF. *Applied Psychological Measurement*, 20(3), 257-274. doi: 10.1177/014662169602000306
- Noortgate, W. V., & De Boeck, P. (2005). Assessing and explaining differential item functioning using logistic mixed models. *Journal of Educational and Behavioral Statistics*, 30(4), 443-464. doi: 10.3102/10769986030004443
- Norouzian, R., & Plonsky, L. (2018). Eta- and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. *Second Language Research*, 34(2), 257-271. doi: 10.1177/0267658316684904
- Okada, K. (2013). Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika*, 40(2), 129-147. doi: 10.2333/bhmk.40.129
- Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, 25(3), 241-286. doi: 10.1006/ceps.2000.1040
- Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8(4), 434-447. doi: 10.1037/1082-989X.8.4.434
- Oliveri, M. E., Ercikan, K., & Zumbo, B. (2013). Analysis of sources of latent class differential item functioning in international assessments. *International Journal of Testing*, 13(3), 272-293. doi: 10.1080/15305058.2012.738266
- Oliveri, M. E., Olson, B., Ercikan, K., & Zumbo, B. (2012). Methodologies for investigating item- and test-level construct comparability in international large-scale assessments. *International Journal of Testing*, 12(3), 203-223. doi: 10.1080/15305058.2011.617475

- Ong, M. L. (2012). *A comparison of differential item functioning detection for dichotomously scored items by using 2.1 IRTPRO, BILOG-MG 3 and IRTLRF V.2* (Unpublished master's thesis). The University of Georgia, Athens.
- Ong, Y. M., Williams, J. S., & Lamprianou, I. (2011). Exploration of the validity of gender differences in mathematics assessment using differential bundle functioning. *International Journal of Testing*, 11(3), 271–293. doi: 10.1080/15305058.2011.555574
- Opariuc-Dan, C., Grigoraş, M., Butucescu, A., Sirbu, A., & Dumbrava, A. C., (2017). Assessment of differential item functioning in personnel selection. *Psihologia Resurselor Umane*, 15(2), 138-148. doi: 10.24837/pru.2017.2.480
- Oshima, T. C., Wright, K., & White, N. (2015). Multiple-group noncompensatory differential item functioning in Raju's differential functioning of items and tests. *International Journal of Testing*, 15(3), 254-273. doi: 10.1080/15305058.2015.1009980
- Osterlind, J. S. (1983). *Test item bias*. London: Sage Publications.
- Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Panter, A. T., & Reeve, B. B. (2002). Assessing tobacco beliefs among youth using item response theory models. *Drug and Alcohol Dependence*, 68(1), 21-39. doi: 10.1016/s0376-8716(02)00213-2
- Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte carlo experiments: Design and implementation. *Structural Equation Modeling*, 8(2), 287–312. doi: 10.1207/S15328007SEM0802_7R
- Paz, S. H., Spritzer, K. L., Reise, S. P., & Hays, R. D. (2017). Differential item functioning of the patient-reported outcomes information system (PROMIS) pain interference item bank by language (Spanish versus English). *Qual Life Res*, 26(6), 1451-1462. doi: 10.1007/s11136-017-1499-3.
- Penfield, R. D. (2001). Assessing differential item functioning across multiple groups: A comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education*, 14(3), 235-259. doi: 10.1207/S15324818AME1403_3
- Penfield, R. D. (2003). Application of the Breslow-Day test of trend in odds ratio heterogeneity to the detection of nonuniform DIF. *Alberta Journal of Educational Research*, 49(3), 231-243. Retrieved from <https://psycnet.apa.org/record/2003-10912-002>
- Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. C. R. Rao, and S. Sinharay (Eds.). *Handbook of statistics* (Vol. 26, pp. 125-167). Amsterdam, Netherlands: Elsevier.

- Penfield, R. D., & Lam, T. C. M. (2000). Assessing differential item functioning in performance assessment: review and recommendations. *Educational Measurement: Issues and Practice*, 19(3), 5–15. doi: 10.1111/j.1745-3992.2000.tb00033.x
- Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor anova designs. *Educational and Psychological Measurement*, 64(6), 916–924. doi: 10.1177/0013164404264848
- Potenza, M. T., & Dorans, N. J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. *Applied Psychological Measurement*, 19(1), 23–37. doi: 10.1177/014662169501900104
- Price, E. A. (2014). *Item discrimination, model-data fit, and Type I error rates in DIF detection using Lord's χ^2 , the likelihood ratio test, and the Mantel-Haenszel procedure* (Unpublished doctoral dissertation). Ohio University.
- Puth, M. T., Neuhauser, M., & Ruxton, G. D. (2014). Effective use of Pearson's product-moment correlation coefficient. *Animal Behaviour*, 93, 183–189. doi: 10.1016/j.anbehav.2014.05.003
- R Development Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria. URL <http://www.R-project.org/>.
- Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, 53(4), 495–502. doi: 10.1007/BF02294403
- Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, 14(2), 197–207. doi: 10.1177/014662169001400208
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A Comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87(3), 517–529. doi: 10.1037/0021-9010.87.3.517
- Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). An IRT-based internal measure of test bias with application of differential item functioning. *Applied Psychological Measurement*, 19(4), 353–368. doi: 10.1177/014662169501900405
- Rasch, D. (2012). Hypothesis testing and the error of the third kind. *Psychological Test and Assessment Modeling*, 54(1), 90–99. Retrieved from <https://psycnet.apa.org/record/2012-13143-006>
- Revicki, D. A., Chen, W. H., & Tucker, C. (2015). Developing item banks for patient-reported health outcomes. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp.334–363). Routledge:Taylor & Francis Group.

- Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6(2), 135–147. doi: 10.1016/j.edurev.2010.12.001
- Rimehaug, S. A., Kaat, A. J., Nordvik, J. E., Klokkeud, M., & Robinson, H. S. (2022). Psychometric properties of the PROMIS-57 questionnaire, Norwegian version. *Qual Life Res*, 31(1), 269–280. doi: 10.1007/s11136-021-02906-1.
- Rizopoulos, D. (2018). *ltm: Latent trait models for item response theory analyses*. R package version 4.7.
- Robey, R. R., & Barcikowski, R. S. (1992). Type I error and the number of iterations in Monte Carlo studies of robustness. *British Journal of Mathematical and Statistical Psychology*, 45(2), 283–288. doi: 10.1111/j.2044-8317.1992.tb00993.x
- Rockoff, D. (2018). *A randomization test for the detection of differential item functioning* (Unpublished doctoral dissertation). The University of Arizona, Arizona.
- Rodebaugh, T. L., Woods, C. M., Heimberg, R. G., Liebowitz, M. R., & Schneier, F. R. (2006). The factor structure and screening utility of the social interaction anxiety scale. *Psychological Assessment*, 18(2), 231–237. doi: 10.1037/1040-3590.18.2.231
- Rogers, H. J., & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*, 17(2), 105–116. doi: 10.1177/014662169301700201
- Rollins III, J. D. (2018). *A comparison of observed score approaches to detecting differential item functioning among multiple groups* (Unpublished doctoral dissertation). The University of North Carolina at Greensboro, Greensboro.
- Roussos, L. A., & Stout, W. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel Type I error performance. *Journal of Educational Measurement*, 33(2), 215–230. doi: 10.1111/j.1745-3984.1996.tb00490.x
- Roussos, L. A., Schnipke, D. L., & Pashley, P. J. (1999). A generalized formula for the Mantel-Haenszel differential item functioning parameter. *Journal of Educational and Behavioral Statistics*, 24(3), 293–322. doi: 10.3102/10769986024003293
- Rubinstein, R. Y., & Kroese, D. P. (2017). *Simulation and the monte carlo method*. New Jersey: John Wiley & Sons.
- Rusch, T., Mair, P., & Hatzinger, R. (2013). *Psychometrics with R: A review of CRAN packages for item response theory*. Center for Empirical Research Methods: Discussion Paper Series

- Sachse, K. A., Roppelt, A., & Haag, N. (2016). A comparison of linking methods for estimating national trends in international comparative large-scale assessments in the presence of cross-national DIF. *Journal of Educational Measurement*, 53(2), 152–171. doi:10.1111/jedm.12106
- Salehi, M., & Tayebi, A. (2012). Differential item functioning: Implications for test validation. *Journal of Language Teaching and Research*, 3(1), 84-92. doi: 10.4304/jltr.3.1.84-92
- Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society.
- Sandilands, D., Oliveri, M. E., Zumbo, B. D., & Ercikan, K. (2013). Investigating sources of differential item functioning in international large-scale assessments using a confirmatory approach. *International Journal of Testing*, 13(2), 152–174. doi: 10.1080/15305058.2012.690140.
- Sarı, H. İ. (2013). *DIF detection across two methods of defining group comparisons: Pairwise and composite group comparisons* (Unpublished master's thesis). University of Florida, Florida.
- Sarı, H. İ., & Huggins, A. C. (2014). Differential item functioning detection across two methods of defining group comparisons: Pairwise and composite group comparisons. *Educational and Psychological Measurement*, 75(4), 648-676. doi: 10.1177/0013164414549764
- Schauberg, G., & Mair, P. (2020). A regularization approach for the detection of differential item functioning in generalized partial credit models. *Behavior Research Methods*, 52, 279–294. doi: 10.3758/s13428-019-01224-2
- Schauberg, G., & Tutz, G. (2016). Detection of differential item functioning in Rasch models by boosting techniques. *British Journal of Mathematical and Statistical Psychology*, 69(1), 80–103. doi: 10.1111/bmsp.12060
- Schmitt, A. P., & Dorans, N. J. (1990). Differential item functioning for minority examinees on the SAT. *Journal of Educational Measurement*, 27(1), 67–81. doi: 10.1111/j.1745-3984.1990.tb00735.x
- Scott, L. (2014). *Controlling Analytic selection of a valid subtest for DIF analysis when DIF has multiple potential causes among multiple groups* (Unpublished doctoral dissertation). Arizona State University, Arizona.
- Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., DeGraeff, A., Groenvold, M., Gundy, C., Koller, M., Petersen, M. A., & Sprangers, M. A. G. (2009). A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *Journal of Clinical Epidemiology*, 62(3), 288–295. doi: 10.1016/j.jclinepi.2008.06.003
- Seybert, J., & Stark, S. (2012). Iterative linking with the differential functioning of items and tests (DFIT) method: Comparison of testwide and item parameter replication (IPR) critical values. *Applied Psychological Measurement*, 36(6), 494-515. doi: 10.1177/0146621612445182

- Seybert, J., Stark, S. & Chernyshenko, O. S. (2014). Detecting DIF with ideal point models: A comparison of area and parameter difference methods. *Applied Psychological Measurement*, 38(2), 151-165. doi: 10.1177/0146621613508306
- Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, 46(5), 561-584. doi: 10.1146/annurev.ps.46.020195.003021
- Shealy, R., & Stout, W. F. (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. doi: 10.1007/BF02294572
- Shepard, L., Camilli, G., & Williams, D. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*, 22(2), 77–105. doi:10.1111/j.1745-3984.1985.tb01050.x
- Sigal, M. J., & Chalmers, R. P. (2016). Play it again: Teaching statistics with monte carlo simulation. *Journal of Statistics Education*, 24(3), 136-156. doi: 10.1080/10691898.2016.1246953
- Sinharay, S., Dorans, N. J., Grant, M. C., & Blew, E. O. (2009). Using past data to enhance small sample DIF estimation: bayesian approach. *Journal of Educational and Behavioral Statistics*, 34(1). 74-96. doi: 10.3102/1076998607309021
- Sireci, S. G., & Ríos, J. A. (2013). Decisions that make a difference in detecting differential item functioning. *Educational Research and Evaluation*, 19(2), 170-187. doi: 10.1080/13803611.2013.767621
- Sireci, S. G., Patsula, L., & Hambleton, R. K. (2017). Test Uyarlama Sürecinde Ortaya Çıkan Hataların Belirlenmesinde Kullanılan İstatistiksel Yöntemler (N. Koç ve A. Yıldırım, Çev.). *Eğitimde ve Psikolojide Kullanılan Testlerin Kültürlerarası Değerlendirme Amacıyla Uyarlanması* (ss. 95–120). Ankara: Pegem Akademi (Orijinal eserin basım yılı 2005).
- Skidmore, S. T., & Thompson, B. (2013). Bias and precision of some classical ANOVA effect sizes when assumptions are violated. *Behavior Research Methods*, 45(2), 536–546. doi: 10.3758/s13428-012-0257-2
- Skrondal, A. (2000). Design and analysis of monte carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, 35(2), 137–167. doi: 10.1207/S15327906MBR3502_1
- Smith, L. L., & Reise, S. P. (1998). Gender differences on negative affectivity: An IRT study of differential item functioning on the multidimensional personality questionnaire stress reaction scale. *Journal of Personality and Social Psychology*, 75(5), 1350-1362. doi: 10.1037//0022-3514.75.5.1350
- Smith, L. L. (2002). On the usefulness of item bias analysis to personality psychology. *Personality and Social Psychology Bulletin*, 28(6), 754-763. doi: 10.1177/0146167202289005

- Soares, T. M., Gonçalves, F. B., & Gamerman, D. (2009). An integrated bayesian model for DIF analysis. *Journal of Educational and Behavioral Statistics*, 34(3), 348-377. doi: 10.3102/1076998609332752
- Socha, A., DeMars, C. E., Zilberberg, A., & Phan, H. (2015). Differential item functioning detection with the Mantel-Haenszel procedure: The effects of matching types and other factors. *International Journal of Testing*, 15(3), 193-215. doi: 10.1080/15305058.2014.984066.
- Somes, G. W. (1986). The generalized Mantel–Haenszel statistic. *The American Statistician*, 40, 106–108. doi: 10.1080/00031305.1986.10475369
- Spence, I. (1983). Monte carlo simulation studies. *Applied Psychological Measurement*, 7(4), 405-425. doi: 10.1177/014662168300700403
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the effects of differential item (functioning and differential) test functioning on selection decisions: When are statistically significant effects practically important? *Journal of Applied Psychology*, 89(3), 497-508. doi: 10.1037/0021-9010.89.3.497
- Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91(1), 25-39. doi: 10.1037/0021-9010.91.1.25
- Steinberg, L. (2001). The consequences of pairing questions: Context effects in personality measurement. *Journal of Personality and Social Psychology*, 81(2), 332–342. doi: 10.1037//0022-3514.81.2.332
- Steinberg, L., & Thissen, D. (2006). Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods*, 11(4), 402–415. doi: 10.1037/1082-989X.11.4.402.
- Stevens, P. J. (2009). *Applied multivariate statistics for the social sciences* (5th ed.) New York, NY: Routledge Taylor and Francis Group.
- Strobl, C., Kopf, J., & Zeileis, A. (2015). Rasch Trees: A new method for detecting differential item functioning in the Rasch model. *Psychometrika*, 80(2), 289-316. doi: 10.1007/s11336-013-9388-3
- Stubbe, T. C. (2011). How do different versions of a test instrument function in a single language? A DIF analysis of the PIRLS 2006 German assessments. *Educational Research and Evaluation*, 17(6), 465–481. doi:10.1080/ 13803611.2011.630560
- Sullivan, G. M., & Feinn, R. (2012). Using effect size – or why the p value is not enough. *Journal of Graduate Medical Education*, 4(3), 279-282. doi: 10.4300/JGME-D-12-00156.1
- Suna, H. E. (2012). *TIMSS 2007 Fen Bilimleri Testindeki Maddelerin Dil ve Cinsiyet Yanlılığı Açısından İncelenmesi* (Yayımlanmamış Yüksek Lisans Tezi). Ankara Üniversitesi, Ankara, Türkiye.

- Sünbül, Ö. ve Ömür Sünbül, S. (2016). Değişen Madde Fonksiyonunun Belirlenmesinde Kullanılan Yöntemlerde I. Tip Hata ve Güç Çalışması. *İlköğretim Online*, 15(3), 882-897. doi: 10.17051/io.2016.10640
- Svetina, D., & Rutkowski, L. (2014). Detecting differential item functioning using generalized logistic regression in the context of large-scale assessments. *Large Scale Assessments in Education*, 2(4), 1-17. doi: 10.1186/s40536-014-0004-5
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370. doi: 10.1111/j.1745-3984.1990.tb00754.x
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (Sixth edition). United States: Pearson Education.
- Tan, Q., Cai, Y., Li, Q., Zhang, Y., & Tu, D. (2018). Development and validation of an item bank for depression screening in the chinese population using computer adaptive testing: A simulation study. *Frontiers in Psychology*, 9, 1-13. doi: 10.3389/fpsyg.2018.01225
- Tate, C. (2018). Type III error. In Bruce B. Frey (Eds.). *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 1744-1746). Thousand Oaks: SAGE Publications, Inc.
- Tay, L., Huang, Q., & Vermunt, J. K. (2016). Item response theory with covariates (IRT-C): Assessing item recovery and differential item functioning for the three-parameter logistic model. *Educational and Psychological Measurement*, 76(1), 22-42. doi: 10.1177/0013164415579488
- Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, 18(1), 3-46. doi:10.1177/1094428114553062.
- Tay, L., Vermunt, J. K., & Wang, C. (2013). Assessing the item response theory with covariate (IRT-C) procedure for ascertaining DIF. *International Journal of Testing*, 13, 201-222. doi:10.1080/15305058.2012.692415.
- Taylor, C. S., & Lee, Y. (2012). Gender DIF in reading and mathematics tests with mixed item formats. *Applied Measurement in Education*, 25(3), 246-280. doi: 10.1080/08957347.2012.687650
- Teymoori, A., Real, R., Gorbunova, A., Haghish, E. F., Andelic, N., Wilson, L., Asendorf, T. Menon, D., & von Steinbüchel, N. (2020). Measurement invariance of assessments of depression (PHQ-9) and anxiety (GAD-7) across sex, strata and linguistic backgrounds in a European-wide sample of patients after traumatic brain injury. *Journal of Affective Disorders*, 262, 278-285. doi: 10.1016/j.jad.2019.10.035
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. P. W. Holland, and H. Wainer (Eds.). *Differential item functioning*. Lawrence Erlbaum.

- Toprak, E. ve Yakar, L. (2017). SBS 2011 Alt Testindeki Maddelerin Değişen Madde Fonksiyonu Açısından Farklı Yöntemlerle İncelenmesi. *International Journal Of Eurasia Social Sciences*, 8(26), 220-231. Erişim Adresi: <https://atif.sobiad.com/index.jsp?modul=makaledetay&Alan=sosyal&Id=d9b8VHEBu-adCBSERt5X>
- Trzpiot, G. (2015). *Some remarks of Type III error for directional two-tailed test*. *Studia Ekonomiczne*.
- Tunç, E. B. (2016). *İki ve Çok Kategorili Puanlanan Maddelerde Değişen Madde Fonksiyonlarının Karşılaştırılması* (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara, Türkiye.
- Tureson, K., & Odland, A. (2018). Monte Carlo simulation studies. In Bruce B. Frey (Eds.). *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 1085-1089). Thousand Oaks: SAGE Publications, Inc.
- Tutz, G., & Schauberger, G. (2015). A penalty approach to differential item functioning in rasch models. *Psychometrika* 80(1), 21–43. doi: 10.1007/s11336-013-9377-6
- Tutz, G., & Berger, M. (2016). Item-focussed trees for the identification of items in differential item functioning. *Psychometrika* 81(3), 727–750. doi: 10.1007/s11336-015-9488-3
- Umesh, U. N., Peterson, R. A., McCann-Nelson, M., & Vaidyanathan, R. (1996). Type IV error in marketing research: The investigation of ANOVA interactions. *Journal of the Academy of Marketing Science*, 24(1), 17-26. doi: 10.1007/BF02893934
- Uyar, Ş. ve Kaya Uyanık, G. (2016). PISA 2012 Bilişsel Maddelerinin Kültüre Göre Değişen Madde Fonksiyonu Bakımından İncelenmesi. *Eğitim ve Öğretim Araştırmaları Dergisi*, 5(3), 230-240.
- Uzun, N. B. ve Gelbal, S. (2017). PISA Fen Başarı Testinin Madde Yanlılığının Kültür ve Dil Açısından İncelenmesi. *Kastamonu Eğitim Dergisi*, 25(6), 2427-2446. Erişim Adresi: <https://atif.sobiad.com/index.jsp?modul=makale-detay&Alan=sosyal&Id=C5SkKHMBu-adCBSEI-hR>
- Ünsal Özberk, E. B. ve Koç, N. (2017). WÇZÖ-IV Maddelerinin Cinsiyet ve Sosyo-Ekonomik Düzey Açısından İşlev Farklılığının Belirlenmesinde Kullanılan Yöntemlerin Karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(1), 112-127. doi: 10.21031/epod.287577
- Vaughn-Coaxum, R. A., Mair, P., & Weisz, J. R. (2016). Racial/ethnic differences in youth depression indicators. *Clinical Psychological Science*, 4(2), 239–253. doi: 10.1177/2167702615591768.
- Walker C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment*, 29(4), 364-376. doi:10.1177/0734282911406666

- Walker, C. M., & Beretvas, S. N. (2001). An empirical investigation demonstrating the multidimensional DIF paradigm: A cognitive explanation for DIF. *Journal of Educational Measurement*, 38(2), 147-163. doi: 10.1111/j.1745-3984.2001.tb01120.x
- Walker, C. M., & Beretvas, S. N. (2003). Comparing multidimensional and unidimensional proficiency classifications: Multidimensional IRT as a diagnostic aid. *Journal of Educational Measurement*, 40(3), 255-275. doi: 10.1111/j.1745-3984.2003.tb01107.x
- Walker, C. M., & Gocer Sahin, S. (2017). Using a multidimensional IRT framework to better understand differential item functioning (DIF): A tale of three DIF detection procedures. *Educational Psychological Measurement*, 77(6), 945–970. doi: 10.1177/0013164416657137
- Walker, C. M., Beretvas, S. N., & Ackerman, T. A. (2001). An examination of conditioning variables used in computer adaptive testing for DIF. *Applied Measurement in Education*, 14(1), 3-16. doi: 10.1207/S15324818AME1401_02
- Wang, M. (2016). *Longitudinal differential item functioning detection using bifactor models and the wald test* (Unpublished doctoral dissertation). University of Kansas.
- Wang, N., & Lane, S. (1996). Detection of gender-related differential item functioning in a mathematics performance assessment. *Applied Measurement in Education*, 9(2), 175-199. doi: 10.1207/s15324818ame0902_6
- Wang, W. C. (2000b). The simultaneous factorial analysis of differential item functioning. *Methods of Psychological Research Online*, 5(1), 51–76. Retrieved from https://www.researchgate.net/publication/286955980_The_simultaneous_factorial_analysis_of_differential_item_functioning
- Wang, W. C., & Chen, C. T. (2005). Item parameter recovery, standard error estimates, and fit statistics of the winsteps program for the family of Rasch models. *Educational and Psychological Measurement*, 65(3), 376-404. doi: 10.1177/0013164404268673
- Wang, W. C., & Shih, C. L. (2010). MIMIC methods for assessing differential item functioning in polytomous items. *Applied Psychological Measurement*, 34(3), 166-180. doi: 10.1177/0146621609355279
- Wang, W. C., & Su, Y. H. (2004). Effects of average signed area between two item characteristic curves and test purification procedures on the DIF detection via the Mantel–Haenszel method. *Applied Measurement in Education*, 17(2), 113–144. doi: 10.1207/s15324818ame1702_2
- Wang, W. C., & Yeh, Y. L. (2003). Effects of anchor item methods on differential item functioning detection with the likelihood ratio test. *Applied Psychological Measurement*, 27(6), 479-498. doi: 10.1177/0146621603259902
- Webster, N. S., & Bowe, A. G. (2021). Examining measurement equivalency within the Latin American module of the International Civic Competence Study 2009. *Journal of Community Psychol.* 2021, 49(7), 2972-2982. doi: 10.1002/jcop.22515.

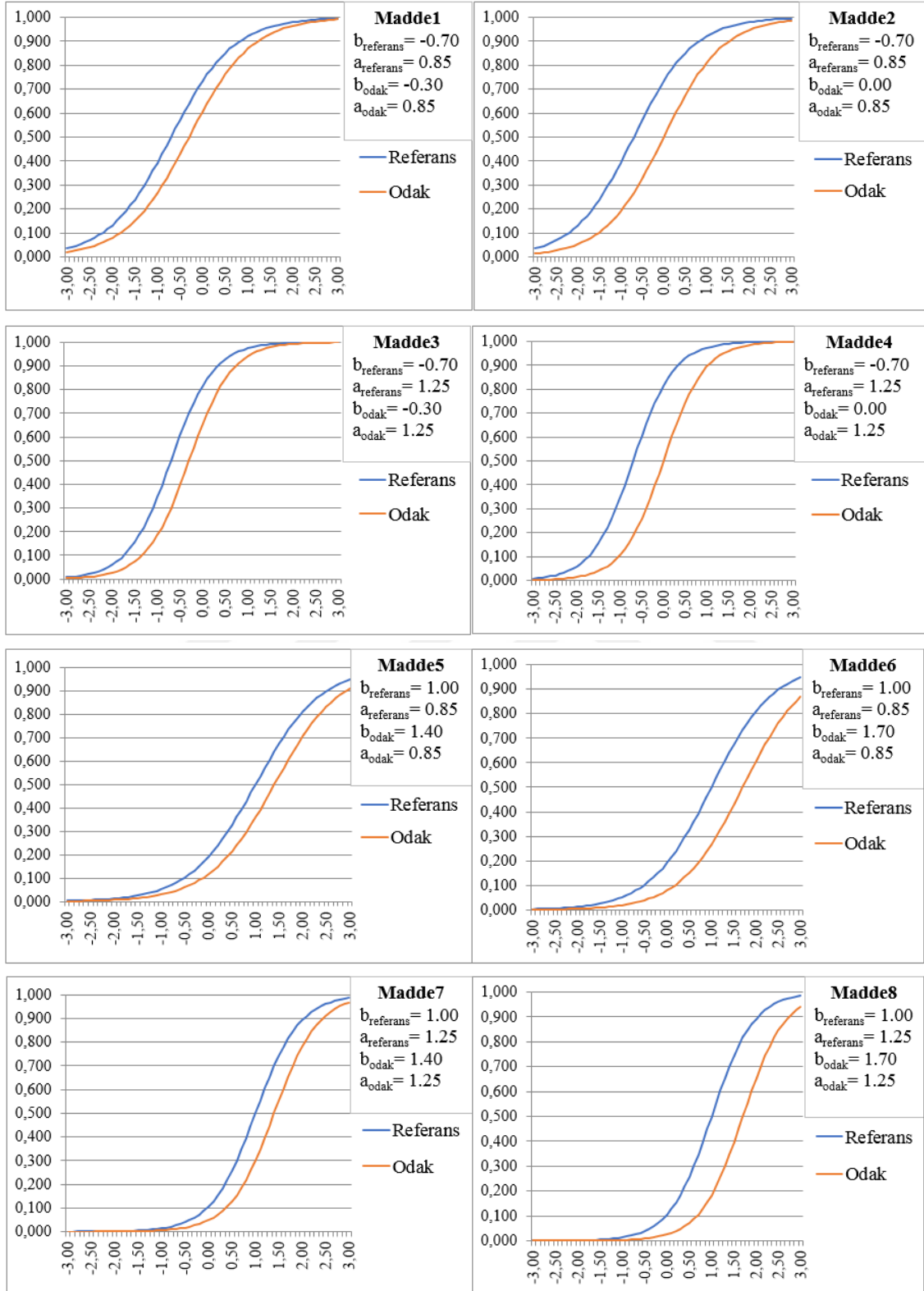
- Wen, Y. (2014). *DIF analyses in multilevel data: Identification and effects on ability estimates* (Unpublished doctoral dissertation). The University of Wisconsin, Milwaukee.
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis* (2nd ed.). New York, NY: Springer.
- Wood, W. S. (2011). *Differential item functioning procedures for polytomous items when examinee sample sizes are small* (Unpublished doctoral dissertation). Graduate College of The University of Iowa, Iowa.
- Woods, C. M. (2008). Likelihood-ratio DIF testing: Effects of nonnormality. *Applied Psychological Measurement*, 32(7), 511-526. doi: 10.1177/0146621607310402
- Woods, C. M. (2009a). Empirical selection of anchors for tests of differential item functioning. *Applied Psychological Measurement*, 33(1), 42-57.
- Woods, C. M. (2009b). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research*, 44(1), 1-27. doi: 10.1080/00273170802620121
- Woods, C. M., & Grimm, K. J. (2011). Testing of nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement*, 35(5), 339-361. doi: 10.1177/0146621611405984
- Woods, C. M., Cai, L., & Wang, M. (2013). The Langer-improved Wald test for DIF testing with multiple groups: Evaluation and comparison to two-group IRT. *Educational and Psychological Measurement*, 73(3), 532-547. doi:10.1177/0013164412464875.
- Wu, M. L., Tam, H. P., & Jen, T. (2016). *Educational measurement for applied researchers: Theory into practice*. Singapore: Springer.
- Wyse, A. E., & Mapuranga, R. (2009). Differential item functioning analysis using Rasch item information functions. *International Journal of Testing*, 9(4), 333- 357. doi: 10.1080/15305050903352040
- Yadav, S. B., & Korukonda, A. (1985). Management of Type III error in problem identification. *Interfaces*, 15(4), 55-61. doi: 10.1287/inte.15.4.55
- Yalçın, S. (2015). *TIMSS 2011 Fen Uygulamasında Cinsiyete Göre Farklılaşan Madde Fonksiyonunu, Madde, Öğrenci ve Okul Düzeyinde Açıklayan Değişkenler* (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara.
- Yao, L., & Li, F. (2015). A DIF detection procedure in multidimensional item response theory framework using MCMC technique. *Int. J. Quantitative Research in Education*. 2(3-4), 285–304. doi: 10.1504/IJQRE.2015.071734.
- Yu, L., Moses, T., Puhon, G., & Dorans, N. (2008). *DIF detection with small samples: applying smoothing techniques to frequency distributions in the Mantel-Haenszel procedure* (ETS research report, RR-08-44). Princeton, NJ.

- Yuan, K.-H., Tong, X., & Zhang, Z. (2015). Bias and efficiency for SEM with missing data and auxiliary variables: Two-stage robust method versus two-stage ML. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(2), 178-192. doi: 10.1080/10705511.2014.935750.
- Yurdugül, H., ve Aşkar, P. (2004). Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'nın Cinsiyete Göre Madde Yanlılığı Açısından İncelenmesi. *Eğitim Bilimleri ve Uygulama Dergisi*, 3(5), 3-20.
- Zampetakis, L. A., Bakatsaki, M., Litos, C., Kafetsios K. G., & Moustakis, V. (2017). Gender-based differential item functioning in the application of the theory of planned behavior for the study of entrepreneurial intentions. *Frontiers in Psychology*, 8, 1-8. doi: 10.3389/fpsyg.2017.00451
- Zenisky, A. L., Hambleton, R. K., & Robin, F. (2003). Detection of DIF in large-scale state assessments: A study evaluating a two-stage approach. *Educational and Psychological Measurement*, 63(1), 51-64. doi: 10.1177/0013164402239316
- Zhang, Y., Dorans, N. J., & Matthews-López, J. L. (2005). *Using DIF dissection method to assess effects of item deletion*. (College Board Research Report No. 2005-10). New York, NY: The College Board.
- Zhong, Q., Gelaye, B., Fann, J. R., Sanchez, S. E., & Williams, M. A. (2014). Cross-cultural validity of the Spanish version of PHQ-9 among pregnant Peruvian women: A Rasch item response theory analysis. *Journal of Affective Disorders*, 158, 148-153. doi: 10.1016/j.jad.2014.02.012.
- Zhu, X., & Aryadoust, V. (2020). An investigation of mother tongue differential item functioning in a high-stakes computerized academic reading test. *Computer Assisted Language Learning*, 33, 1-24. doi:10.1080/09588221.2019.17047884
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland, and H. Wainer (Eds.), *Differential item functioning* (pp. 337-347). Hillsdale, NJ: Lawrence Erlbaum.
- Zieky, M. (2006). Fairness reviews in assessment. In S. M. Downing, & T. M. Haladyna (Eds.), *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Zumbo, B. D. A. (1999). *Handbook on the theory and methods of differential item functioning: Logistic regression modeling as a unitary framework for binary and likert-type item scores*. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zumbo, B. D. A. (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. doi: 10.1080/15434300701375832

- Zumbo, B. D. A., & 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. Retrieved from <https://eric.ed.gov/?id=EJ846827>
- Zumbo, B. D. A., & Thomas, D. R. (1997). *A measure of effect size for a model-based approach for studying DIF* (Working paper of the Edgeworth Laboratory for Quantitative Behavioral Science). Canada: University of Northern British Columbia.
- Zumbo, B. D. A., Liu, Y., Wu, A. D., Shear, B. R., Olvera, O. L., & Tanvinder, K. A. (2015). A methodology for Zumbo's third generation DIF Analysis and the Ecology of Item Responding. *Language Assessment Quarterly*, 12(1), 136-151. doi: 10.1080/15434303.2014.972559.
- Zwick, R. (1990). When do item response function and Mantel–Haenszel definitions of differential item functioning coincide? *Journal of Educational Statistics*, 15(3), 185-198. doi: 10.3102/10769986015003185
- Zwick, R. (2012). *A review of ETS differential item functioning assessment procedures: flagging rules, minimum sample size requirements, and criterion refinement* (Research Report; ETS RR-89-18). New Jersey, NJ: Educational Testing Service.
- Zwick, R., & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement*, 26(1), 55-66. doi: 10.1111/j.1745-3984.1989.tb00318.x
- Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30(3), 233–251. doi: 10.1111/j.1745-3984.1993.tb00425.x

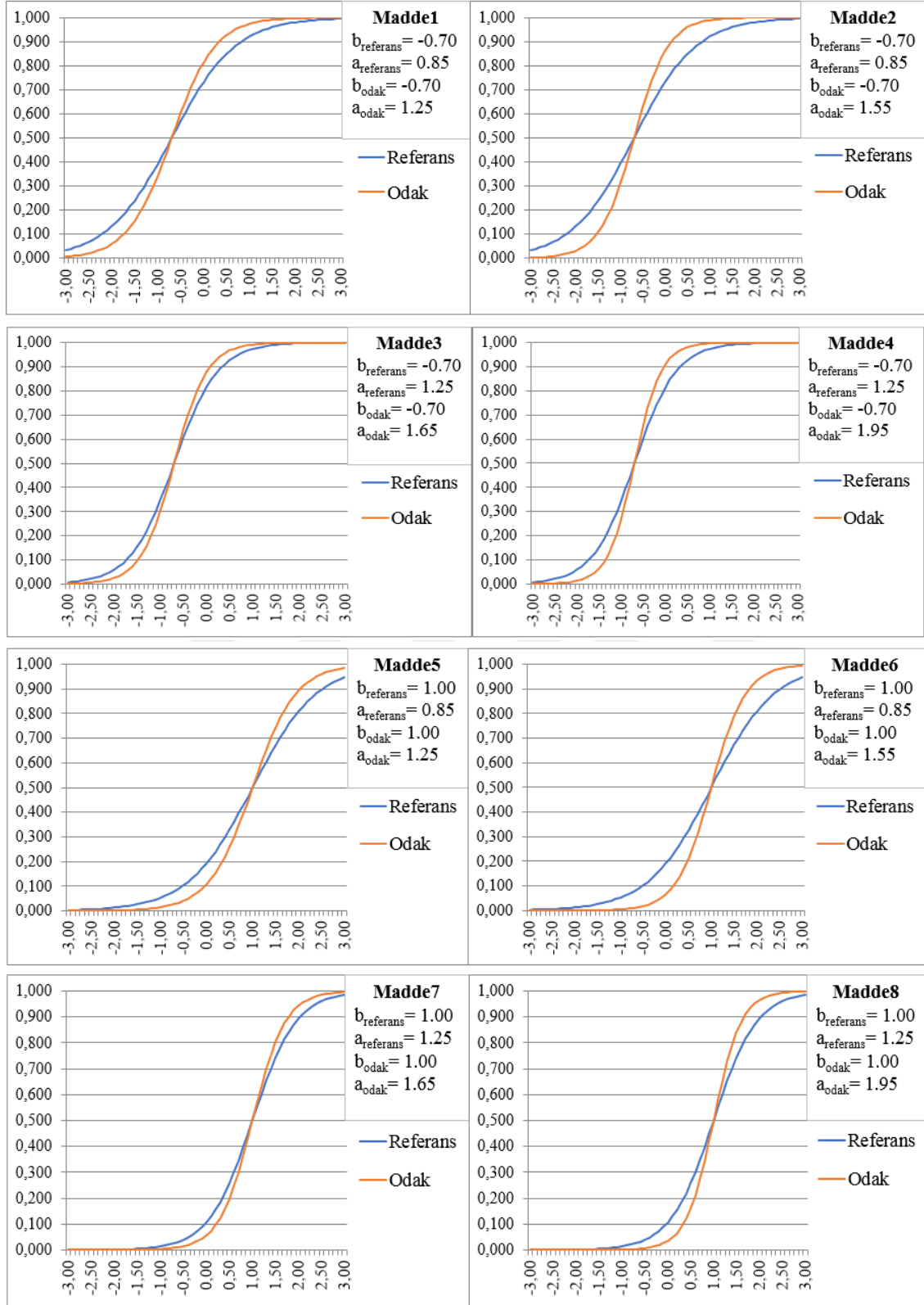


EK 1. TB DMF Türü İçin Örnek MKE'ler



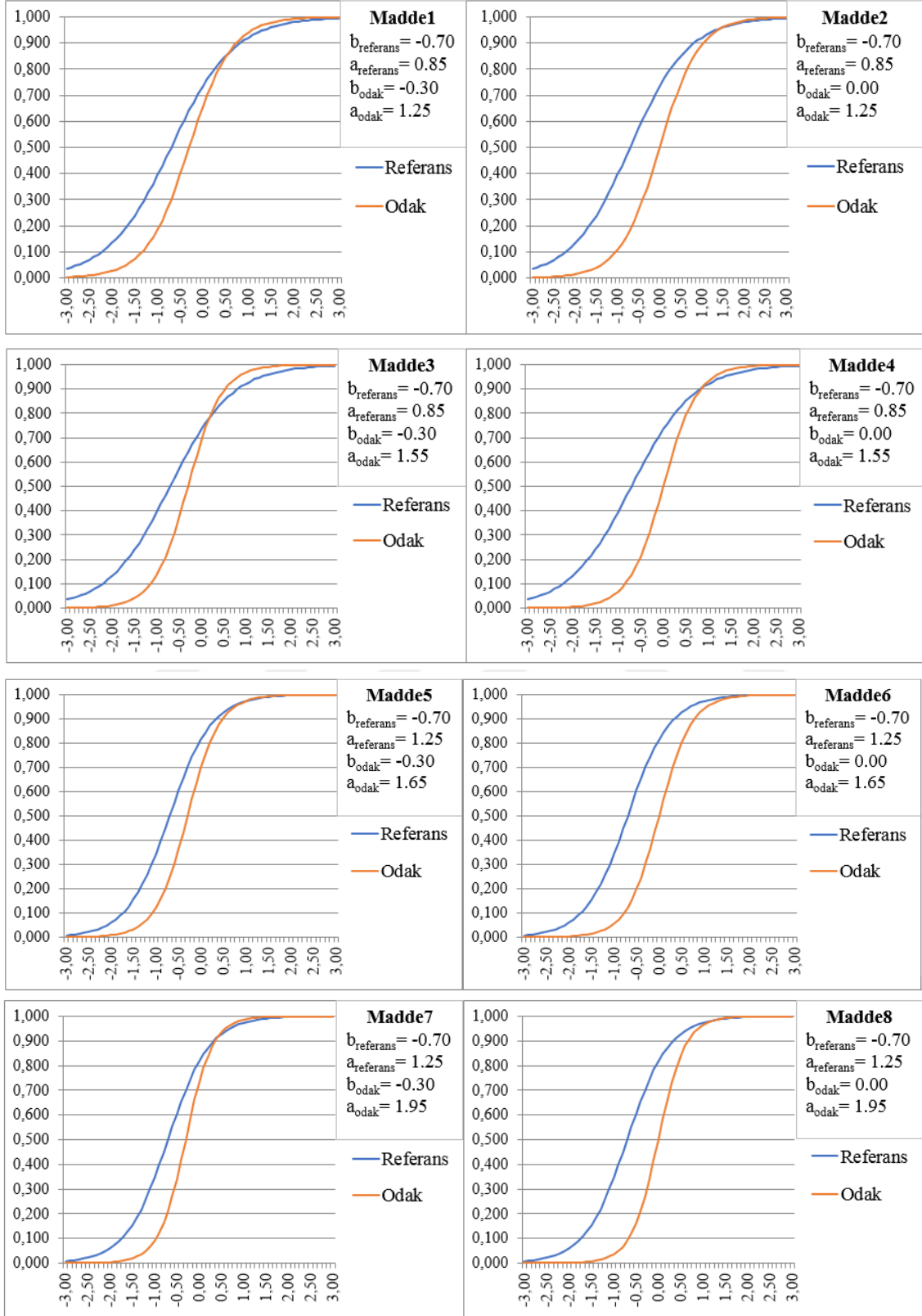
Şekil 1.1. TB DMF Türü İçin Madde Karakteristik Eğrileri

EK 2. TBO(a) DMF Türü İçin Örnek MKE'ler



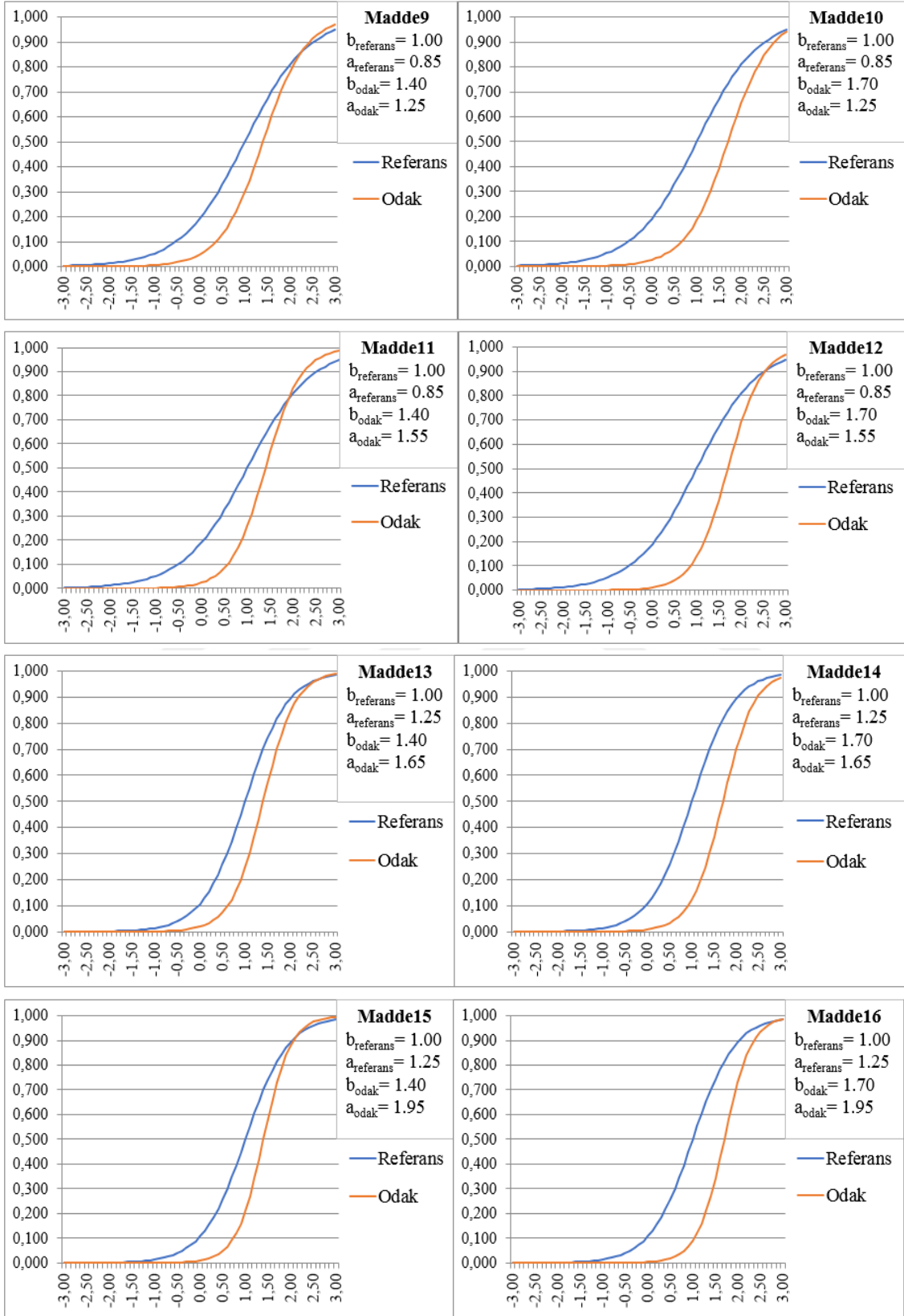
Şekil 2.1. TBO(a) DMF Türü İçin Madde Karakteristik Eğrileri

EK 3. TBO(ab) DMF Türü İçin Örnek MKE'ler



Şekil 3.1. TBO(ab) DMF Türü İçin Madde Karakteristik Eğrileri

EK 3. TBO(ab) DMF Türü İçin Örnek MKE'ler (Devam)



Şekil 3.1. TBO(ab) DMF Türü İçin Madde Karakteristik Eğrileri (Devam)

EK 4. Referans ve Odak Grupların Örneklem Büyüklükleri

Tablo 4.1

Üç Grup İçin Örneklem Büyüklüğü ve Örneklem Büyüklüğü Oranları

Koşul	Referans	Odak1	Odak2	Koşul	Referans	Odak1	Odak2
1	600	600	600	10	1800	1800	1800
2	600	600	400	11	1800	1800	1200
3	600	600	300	12	1800	1800	900
4	600	400	600	13	1800	1200	1800
5	600	400	400	14	1800	1200	1200
6	600	400	300	15	1800	1200	900
7	600	300	600	16	1800	900	1800
8	600	300	400	17	1800	900	1200
9	600	300	300	18	1800	900	900

Tablo 4.2

Dört Grup İçin Örneklem Büyüklüğü ve Örneklem Büyüklüğü Oranları

Koşul	Referans	Odak1	Odak2	Odak3	Koşul	Referans	Odak1	Odak2	Odak3
1	600	600	600	600	28	1800	1800	1800	1800
2	600	600	600	400	29	1800	1800	1800	1200
3	600	600	600	300	30	1800	1800	1800	900
4	600	600	400	600	31	1800	1800	1200	1800
5	600	600	400	400	32	1800	1800	1200	1200
6	600	600	400	300	33	1800	1800	1200	900
7	600	600	300	600	34	1800	1800	900	1800
8	600	600	300	400	35	1800	1800	900	1200
9	600	600	300	300	36	1800	1800	900	900
10	600	400	400	400	37	1800	1200	1200	1200
11	600	400	400	300	38	1800	1200	1200	900
12	600	400	300	300	39	1800	1200	900	900
13	600	400	600	600	40	1800	1200	1800	1800
14	600	400	400	600	41	1800	1200	1200	1800
15	600	400	600	400	42	1800	1200	1800	1200
16	600	400	600	300	43	1800	1200	1800	900
17	600	400	300	600	44	1800	1200	900	1800
18	600	400	300	400	45	1800	1200	900	1200
19	600	300	600	600	46	1800	900	1800	1800
20	600	300	600	400	47	1800	900	1800	1200
21	600	300	600	300	48	1800	900	1800	900
22	600	300	400	600	49	1800	900	1200	1800
23	600	300	400	400	50	1800	900	1200	1200
24	600	300	400	300	51	1800	900	1200	900
25	600	300	300	600	52	1800	900	900	1800
26	600	300	300	400	53	1800	900	900	1200
27	600	300	300	300	54	1800	900	900	900

EK 5. Odak Gruplara Eklenen DMF Miktarları

Tablo 5.1

Üç Grup İçin Odak Gruplara Eklenen DMF Miktarları

Koşul Sayısı	Odak1		Odak2	
	Güçlük Parametresi	Ayırt Edicilik Parametresi	Güçlük Parametresi	Ayırt Edicilik Parametresi
1	0.00	0.00	0.00	0.00
2	0.40	0.00	0.40	0.00
3	0.40	0.00	0.70	0.00
4	0.70	0.00	0.40	0.00
5	0.70	0.00	0.70	0.00
6	0.00	0.40	0.00	0.40
7	0.00	0.40	0.00	0.70
8	0.00	0.70	0.00	0.40
9	0.00	0.70	0.00	0.70
10	0.40	0.40	0.40	0.40
11	0.40	0.40	0.40	0.70
12	0.40	0.40	0.70	0.40
13	0.40	0.70	0.40	0.40
14	0.70	0.40	0.40	0.40
15	0.40	0.40	0.70	0.70
16	0.40	0.70	0.40	0.70
17	0.70	0.40	0.40	0.70
18	0.70	0.40	0.70	0.40
19	0.70	0.70	0.40	0.40
20	0.40	0.70	0.70	0.40
21	0.40	0.70	0.70	0.70
22	0.70	0.40	0.70	0.70
23	0.70	0.70	0.40	0.70
24	0.70	0.70	0.70	0.40
25	0.70	0.70	0.70	0.70

Tablo 5.2

Dört Grup İçin Odak Gruplara Eklenen DMF Miktarları

Koşul Sayısı	Odak1		Odak2		Odak3	
	Güçlük Parametresi	Ayırt Edicilik Parametresi	Güçlük Parametresi	Ayırt Edicilik Parametresi	Güçlük Parametresi	Ayırt Edicilik Parametresi
1	0.00	0.00	0.00	0.00	0.00	0.00
2	0.40	0.00	0.40	0.00	0.40	0.00
3	0.40	0.00	0.40	0.00	0.70	0.00
4	0.40	0.00	0.70	0.00	0.40	0.00
5	0.70	0.00	0.40	0.00	0.40	0.00
6	0.70	0.00	0.40	0.00	0.70	0.00
7	0.70	0.00	0.70	0.00	0.40	0.00
8	0.40	0.00	0.70	0.00	0.70	0.00
9	0.70	0.00	0.70	0.00	0.70	0.00
10	0.00	0.40	0.00	0.40	0.00	0.40
11	0.00	0.40	0.00	0.40	0.00	0.70
12	0.00	0.40	0.00	0.70	0.00	0.40
13	0.00	0.70	0.00	0.40	0.00	0.40
14	0.00	0.70	0.00	0.40	0.00	0.70
15	0.00	0.70	0.00	0.70	0.00	0.40
16	0.00	0.40	0.00	0.70	0.00	0.70
17	0.00	0.70	0.00	0.70	0.00	0.70
18	0.40	0.40	0.40	0.40	0.40	0.40
19	0.40	0.40	0.40	0.40	0.40	0.70

(Devam ediyor)

Tablo 5.2 (Devam)

Dört Grup İçin Odak Gruplara Eklenen DMF Miktarları

Koşul Sayısı	Odak1		Odak2		Odak3	
	Güçlük Parametresi	Ayırt Edicilik Parametresi	Güçlük Parametresi	Ayırt Edicilik Parametresi	Güçlük Parametresi	Ayırt Edicilik Parametresi
20	0.40	0.40	0.40	0.40	0.70	0.40
21	0.40	0.40	0.40	0.40	0.70	0.70
22	0.40	0.40	0.40	0.70	0.40	0.40
23	0.40	0.40	0.40	0.70	0.40	0.70
24	0.40	0.40	0.40	0.70	0.70	0.40
25	0.40	0.40	0.40	0.70	0.70	0.70
26	0.40	0.40	0.70	0.40	0.40	0.40
27	0.40	0.40	0.70	0.40	0.40	0.70
28	0.40	0.40	0.70	0.40	0.70	0.40
29	0.40	0.40	0.70	0.40	0.70	0.70
30	0.40	0.40	0.70	0.70	0.40	0.40
31	0.40	0.40	0.70	0.70	0.40	0.70
32	0.40	0.40	0.70	0.70	0.70	0.40
33	0.40	0.40	0.70	0.70	0.70	0.70
34	0.40	0.70	0.40	0.40	0.40	0.40
35	0.40	0.70	0.40	0.40	0.40	0.70
36	0.40	0.70	0.40	0.40	0.70	0.40
37	0.40	0.70	0.40	0.40	0.70	0.70
38	0.40	0.70	0.40	0.70	0.40	0.40
39	0.40	0.70	0.40	0.70	0.40	0.70
40	0.40	0.70	0.40	0.70	0.70	0.40
41	0.40	0.70	0.40	0.70	0.70	0.70
42	0.40	0.70	0.70	0.40	0.40	0.40
43	0.40	0.70	0.70	0.40	0.40	0.70
44	0.40	0.70	0.70	0.40	0.70	0.40
45	0.40	0.70	0.70	0.40	0.70	0.70
46	0.40	0.70	0.70	0.70	0.40	0.40
47	0.40	0.70	0.70	0.70	0.40	0.70
48	0.40	0.70	0.70	0.70	0.70	0.40
49	0.40	0.70	0.70	0.70	0.70	0.70
50	0.70	0.40	0.40	0.40	0.40	0.40
51	0.70	0.40	0.40	0.40	0.40	0.70
52	0.70	0.40	0.40	0.40	0.70	0.40
53	0.70	0.40	0.40	0.40	0.70	0.70
54	0.70	0.40	0.40	0.70	0.40	0.40
55	0.70	0.40	0.40	0.70	0.40	0.70
56	0.70	0.40	0.40	0.70	0.70	0.40
57	0.70	0.40	0.40	0.70	0.70	0.70
58	0.70	0.40	0.70	0.40	0.40	0.40
59	0.70	0.40	0.70	0.40	0.40	0.70
60	0.70	0.40	0.70	0.40	0.70	0.40
61	0.70	0.40	0.70	0.40	0.70	0.70
62	0.70	0.40	0.70	0.70	0.40	0.40
63	0.70	0.40	0.70	0.70	0.40	0.70
64	0.70	0.40	0.70	0.70	0.70	0.40
65	0.70	0.40	0.70	0.70	0.70	0.70
66	0.70	0.70	0.40	0.40	0.40	0.40
67	0.70	0.70	0.40	0.40	0.40	0.70
68	0.70	0.70	0.40	0.40	0.70	0.40
69	0.70	0.70	0.40	0.40	0.70	0.70
70	0.70	0.70	0.40	0.70	0.40	0.40
71	0.70	0.70	0.40	0.70	0.40	0.70
72	0.70	0.70	0.40	0.70	0.70	0.40
73	0.70	0.70	0.40	0.70	0.70	0.70
74	0.70	0.70	0.70	0.40	0.40	0.40
75	0.70	0.70	0.70	0.40	0.40	0.70
76	0.70	0.70	0.70	0.40	0.70	0.40
77	0.70	0.70	0.70	0.40	0.70	0.70
78	0.70	0.70	0.70	0.70	0.40	0.40
79	0.70	0.70	0.70	0.70	0.40	0.70
80	0.70	0.70	0.70	0.70	0.70	0.40
81	0.70	0.70	0.70	0.70	0.70	0.70

EK 6. Üç Grup İçin Oluşturulan Simülasyon Deseni

Tablo 6.1

Grup Sayısının Üç Olduğu Durum İçin Simülasyon Deseni

Örneklem Büyüklüğü		DMF		DMF		DMF Teknikleri	Tekrar Sayısı	Üretilen Veri Sayısı
Referans Grup	Odak Gruplar	DMF Türü	DMF Miktarı	DMF Miktarına Göre Koşul Sayısı ^a				
	Odak1: Odak2		Güçlük	Ayrıt Edicilik				
600	600:600	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100	500
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100	2000
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100	2000
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100	8000
600	600:400	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100	500
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100	2000
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100	2000
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100	8000
600	600:300	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100	500
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100	2000
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100	2000
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100	8000
600	400:600	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100	500
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100	2000
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100	2000
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100	8000
600	300:600	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100	500
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100	2000
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100	2000
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100	8000
600	400:400	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100	500
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100	2000
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100	2000
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100	8000
600	400:300	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100	500
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100	2000
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100	2000
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100	8000
600	300:400	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100	500
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100	2000
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100	2000
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100	8000
600	300:300	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100	500
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100	2000
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100	2000
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100	8000
1800	1800:1800	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100	500
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100	2000
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100	2000
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100	8000
1800	1800:1200	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100	500
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100	2000
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100	2000
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100	8000

(Devam ediyor)

Tablo 6.1 (Devam)

Grup Sayısının Üç Olduğu Durum İçin Simülasyon Deseni

Örneklem Büyüklüğü		DMF		DMF		Tekrar Sayısı	Üretilen Veri Sayısı
Referans Grup	Odak Gruplar	DMF Türü	DMF Miktarı	DMF Miktarına Göre Koşul Sayısı ^a	DMF Teknikleri		
	Odak1: Odak2		Güçlük	Ayrıt Edicilik			
1800	1800:900	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100
1800	1200:1800	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100
1800	900:1800	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100
1800	1200:1200	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100
1800	1200:900	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100
1800	900:1200	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100
1800	900:900	DMF Yok	b (0.00, 0.00)	a (0.00, 0.00)	1	5	100
		TB(b)	b (0.40, 0.70)	a (0.00, 0.00)	4	5	100
		TBO(a)	b (0.00, 0.00)	a (0.40, 0.70)	4	5	100
		TBO(ab)	b (0.40, 0.70)	a (0.40, 0.70)	16	5	100
Toplam							225,000

^a Odak 1 ve Odak 2 için DMF miktarlarının kombinasyonlarına göre oluşturulan koşul sayıları

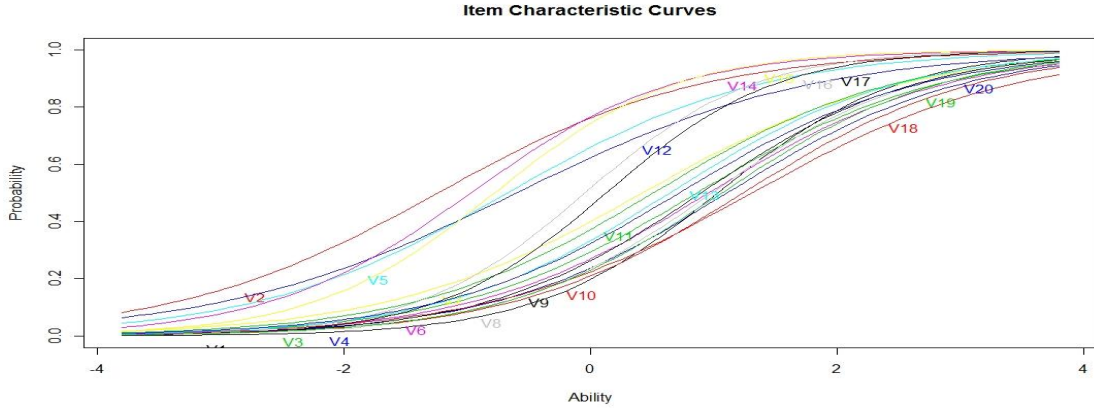
EK 7. Referans ve Odak Grupların Madde Parametreleri

Tablo 7.1

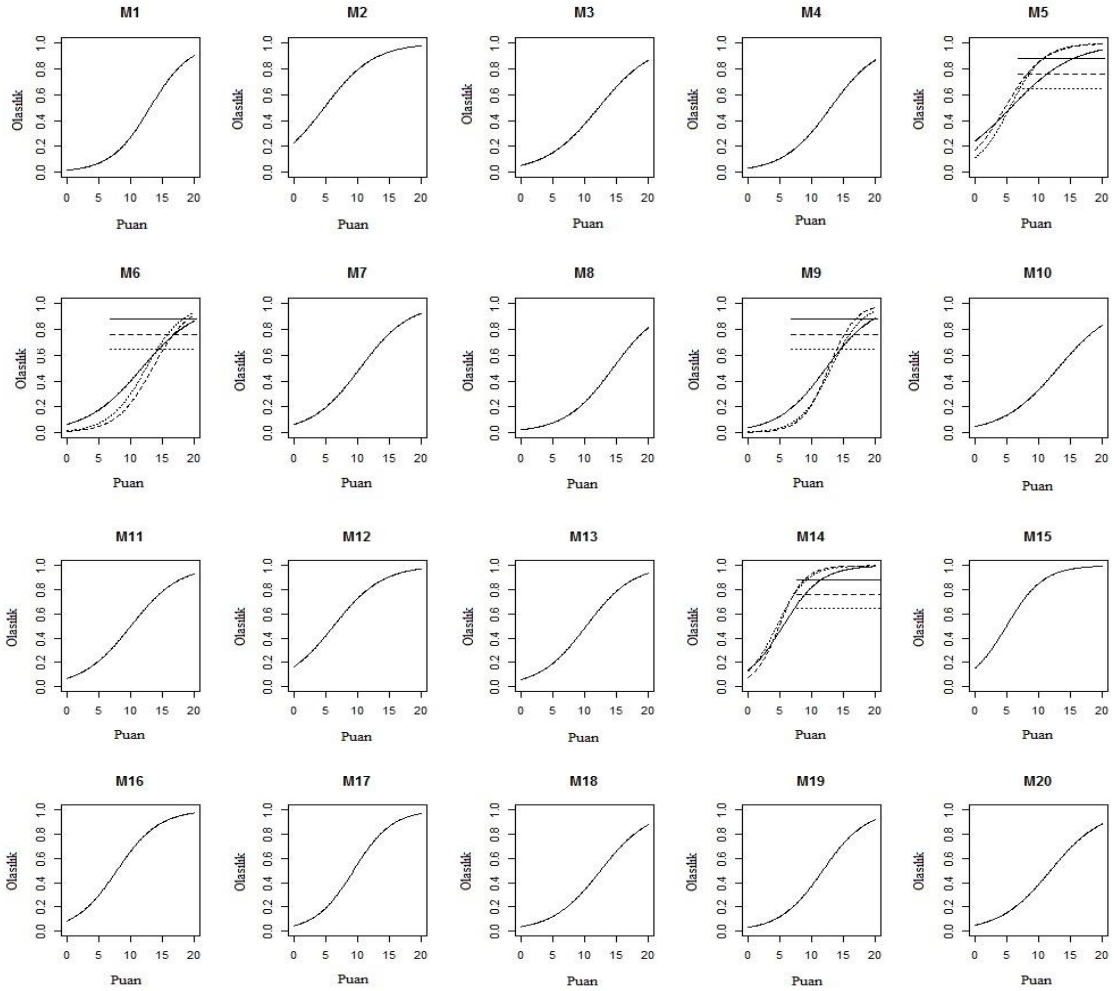
Referans Grubun Madde Parametreleri ile Odak Gruplara Eklenen DMF Miktarları

Madde Sırası	Referans Grubun Madde Parametreleri		Odak Grupların Madde Parametrelerine Eklenen DMF Miktarları		
	Güçlük	Ayırt edicilik	Güçlük	Ayırt edicilik	Güçlük ve Ayırt Edicilik
1	1.03	1.35			
2	-1.00	0.96			
3	1.09	0.92			
4	0.97	1.23			
5	-0.71	0.86	<i>b(+0.40;+0.7)</i>	<i>a(+0.40; +0.70)</i>	<i>b(+0.40; +0.70) a(+0.40; +0.70)</i>
6	1.00	0.88	<i>b(+0.40;+0.7)</i>	<i>a(+0.40; +0.70)</i>	<i>b(+0.40; +0.70) a(+0.40; +0.70)</i>
7	0.53	0.88			
8	1.15	1.04			
9	1.00	1.22	<i>b(+0.40;+0.7)</i>	<i>a(+0.40; +0.70)</i>	<i>b(+0.40; +0.70) a(+0.40; +0.70)</i>
10	1.29	0.95			
11	0.43	1.05			
12	-0.52	0.96			
13	0.51	1.14			
14	-0.79	1.31	<i>b(+0.40;+0.7)</i>	<i>a(+0.40; +0.70)</i>	<i>b(+0.40; +0.70) a(+0.40; +0.70)</i>
15	-0.81	1.32			
16	-0.07	1.47			
17	0.14	1.36			
18	1.07	1.07			
19	0.93	1.29			
20	0.77	1.05			

EK 8. Referans ve Odak Gruplar için Madde Karakteristik Eğrileri

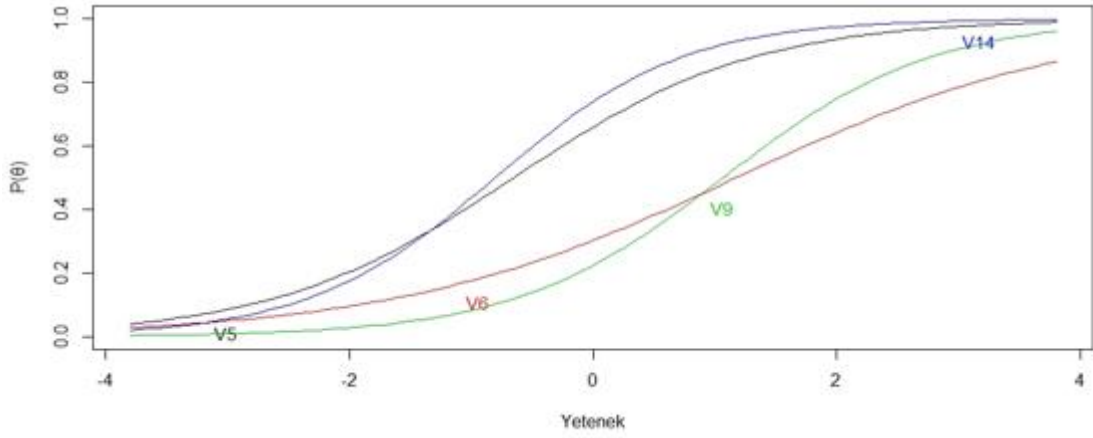


Şekil 8.1. Referans grup için Madde Karakteristik Eğrileri

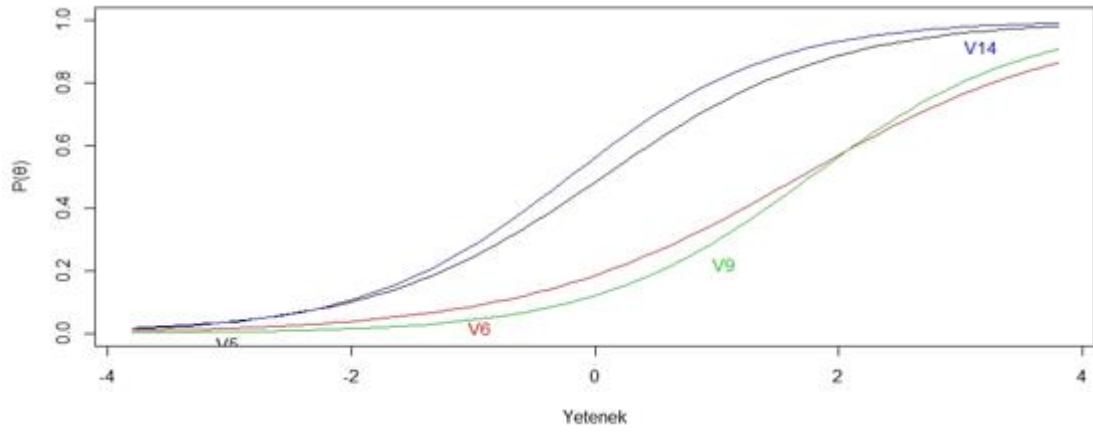


Şekil 8.2. Testteki 20 Maddenin Referans ve Odak Gruplar İçin Madde Karakteristik Eğrileri

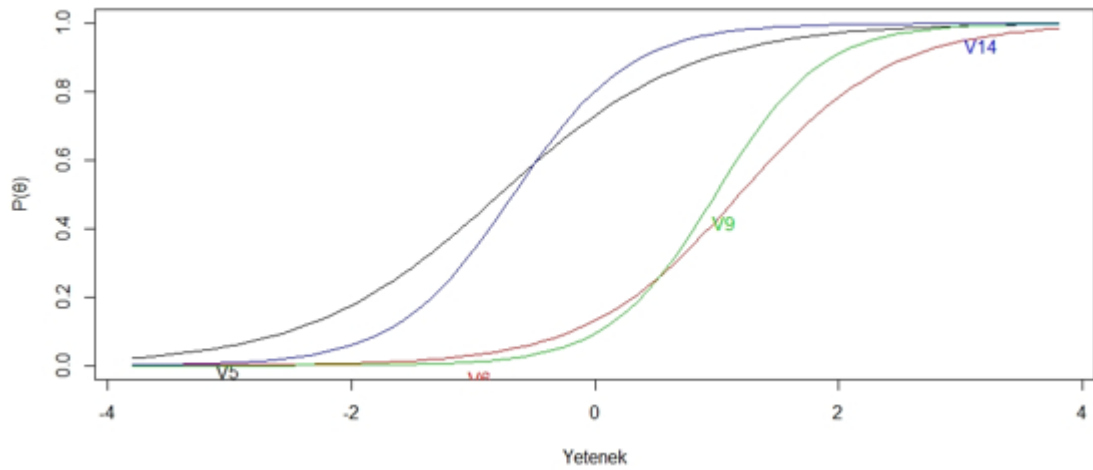
EK 9. DMF'li Dört Maddenin Referans ve Odak Gruplardaki MKE'leri



Şekil 9.1. DMF Olmayan Durum İçin MKE'ler (M5, M6, M9, M14)

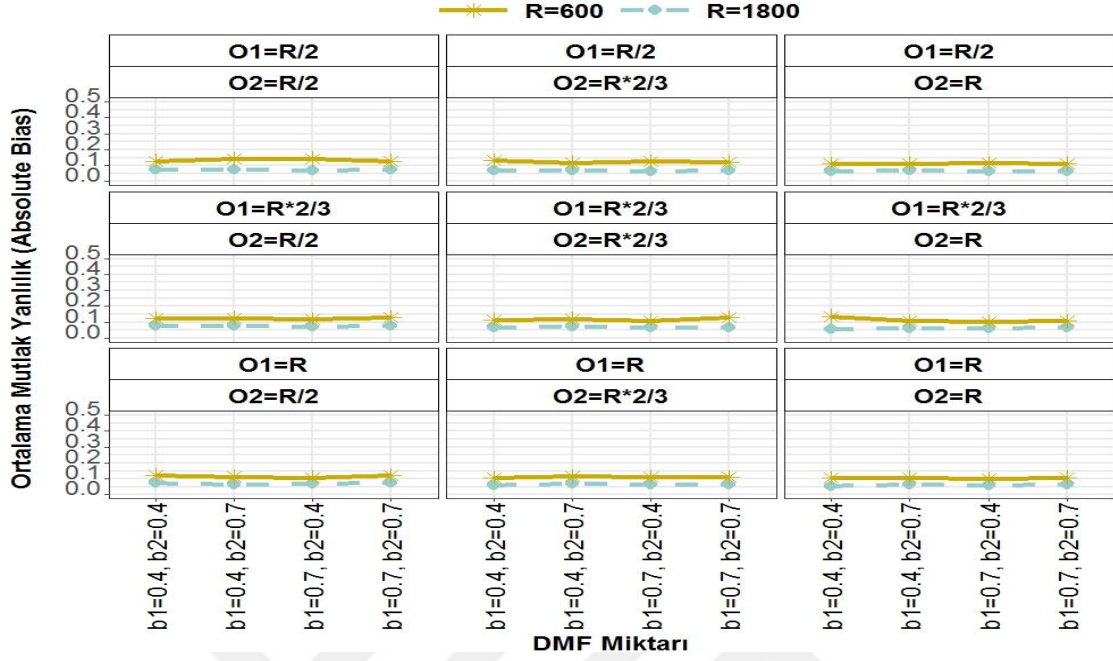


Şekil 9.2. Güçlük Parametresinden Kaynaklı DMF (0.70) İçin MKE'ler (M5, M6, M9, M14)

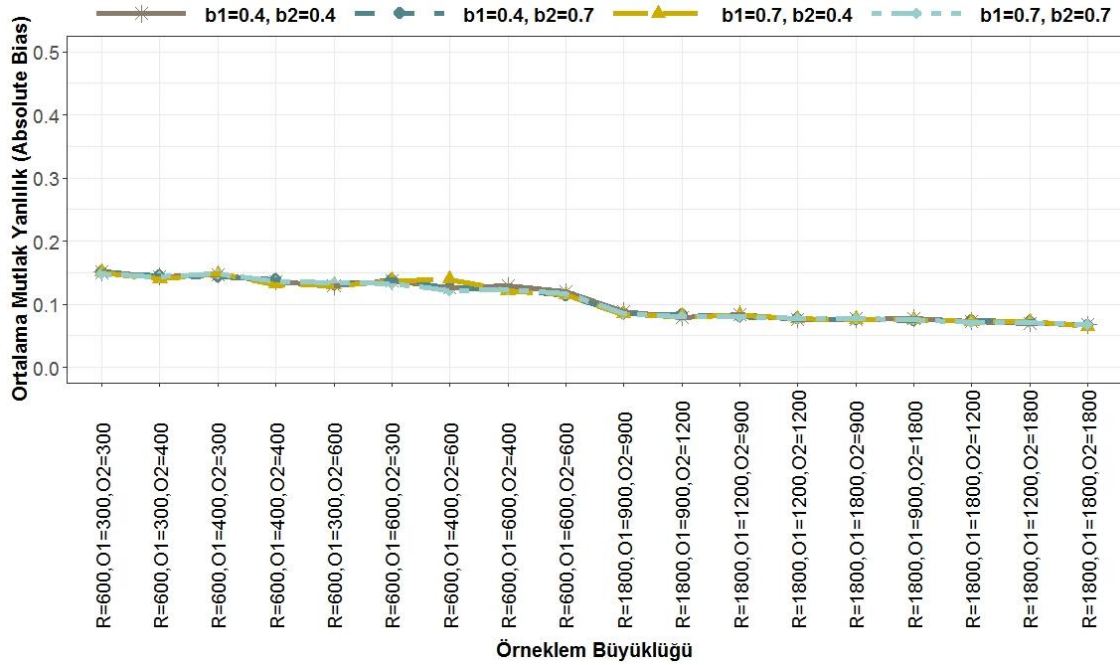


Şekil 9.3. Ayırt Edicilik Parametresinden Kaynaklı DMF (0.70) İçin MKE'ler (M5, M6, M9, M14)

EK 10. Madde Güçlük Parametresi İçin Ortalama Mutlak Yanlılık



Şekil 10.1. Odak Grupların Örneklem Büyüklüğü Oranlarına ile DMF Miktarlarına Göre Madde Güçlük Parametresi Etkileşimi İçin Elde Edilen Ortalama Mutlak Yanlılık Değerleri



Şekil 10.2. Grupların Örneklem Büyüklükleri ile DMF Miktarlarına Göre Madde Güçlük Parametresi Etkileşimi İçin Elde Edilen Ortalama Mutlak Yanlılık Değerleri

**EK 11. Madde Güçlük Parametresine İlişkin Ortalama Mutlak Yanlılık ve
Ortalama RMSE Sonuçları**

Tablo 11.1

Madde Güçlük Parametresi İçin Mutlak Yanlılık ve RMSE Sonuçları

Referans Örnekleme Büyükliği	Odak1 Örnekleme Büyükliği	Odak2 Örnekleme Büyükliği	Odak1 DMF Miktarı	Odak2 DMF Miktarı	Ortalama Mutlak yanlılık	Ortalama RMSE
600	300	300	0.40	0.40	0.1284	0.1708
600	300	300	0.70	0.40	0.1422	0.1905
600	300	300	0.40	0.70	0.1398	0.1787
600	300	300	0.70	0.70	0.1288	0.1725
600	300	400	0.40	0.40	0.1316	0.1738
600	300	400	0.70	0.40	0.1255	0.1671
600	300	400	0.40	0.70	0.1172	0.1540
600	300	400	0.70	0.70	0.1195	0.1544
600	300	600	0.40	0.40	0.1100	0.1446
600	300	600	0.70	0.40	0.1151	0.1475
600	300	600	0.40	0.70	0.1135	0.1489
600	300	600	0.70	0.70	0.1121	0.1441
600	400	300	0.40	0.40	0.1227	0.1565
600	400	300	0.70	0.40	0.1188	0.1504
600	400	300	0.40	0.70	0.1251	0.1657
600	400	300	0.70	0.70	0.1295	0.1777
600	400	400	0.40	0.40	0.1156	0.1479
600	400	400	0.70	0.40	0.1088	0.1372
600	400	400	0.40	0.70	0.1191	0.1525
600	400	400	0.70	0.70	0.1274	0.1671
600	400	600	0.40	0.40	0.1370	0.1748
600	400	600	0.70	0.40	0.1031	0.1294
600	400	600	0.40	0.70	0.1077	0.1351
600	400	600	0.70	0.70	0.1065	0.1389
600	600	300	0.40	0.40	0.1218	0.1533
600	600	300	0.70	0.40	0.1062	0.1374
600	600	300	0.40	0.70	0.1139	0.1487
600	600	300	0.70	0.70	0.1213	0.1596
600	600	400	0.40	0.40	0.1051	0.1359
600	600	400	0.70	0.40	0.1102	0.1427
600	600	400	0.40	0.70	0.1160	0.1466
600	600	400	0.70	0.70	0.1102	0.1406
600	600	600	0.40	0.40	0.1054	0.1341
600	600	600	0.70	0.40	0.1003	0.1294
600	600	600	0.40	0.70	0.1063	0.1363
600	600	600	0.70	0.70	0.1064	0.1367
1800	900	900	0.40	0.40	0.0747	0.0952
1800	900	900	0.70	0.40	0.0699	0.0905
1800	900	900	0.40	0.70	0.0736	0.0935
1800	900	900	0.70	0.70	0.0762	0.0965
1800	900	1200	0.40	0.40	0.0683	0.0866
1800	900	1200	0.70	0.40	0.0669	0.0850
1800	900	1200	0.40	0.70	0.0701	0.0910
1800	900	1200	0.70	0.70	0.0703	0.0908
1800	900	1800	0.40	0.40	0.0665	0.0859
1800	900	1800	0.70	0.40	0.0666	0.0867
1800	900	1800	0.40	0.70	0.0700	0.0883

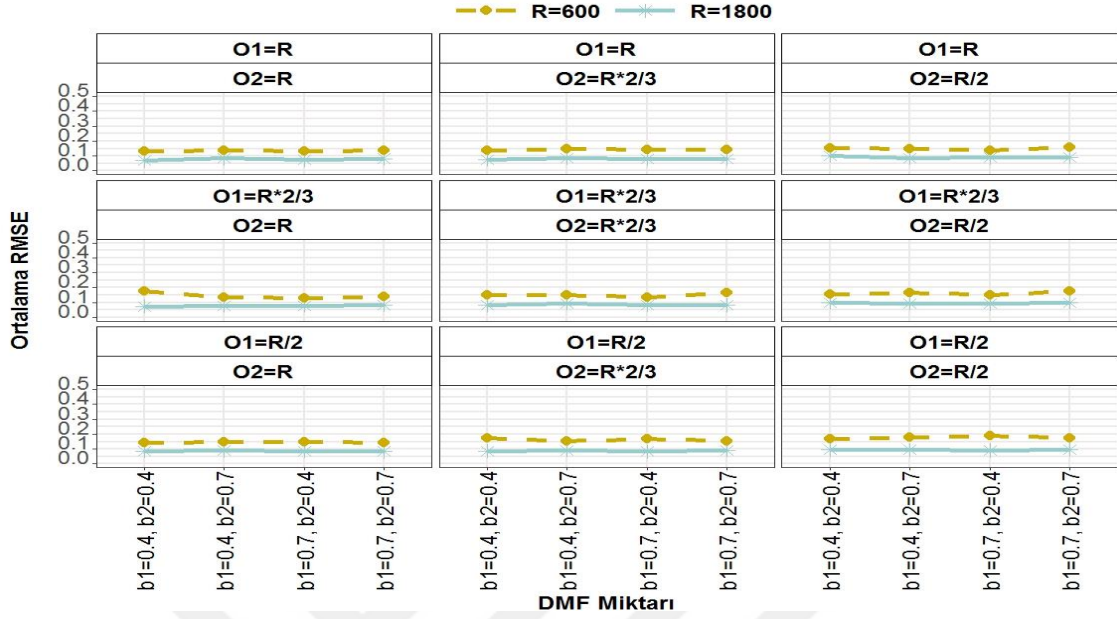
(Devam ediyor)

Tablo 11.1 (Devam)

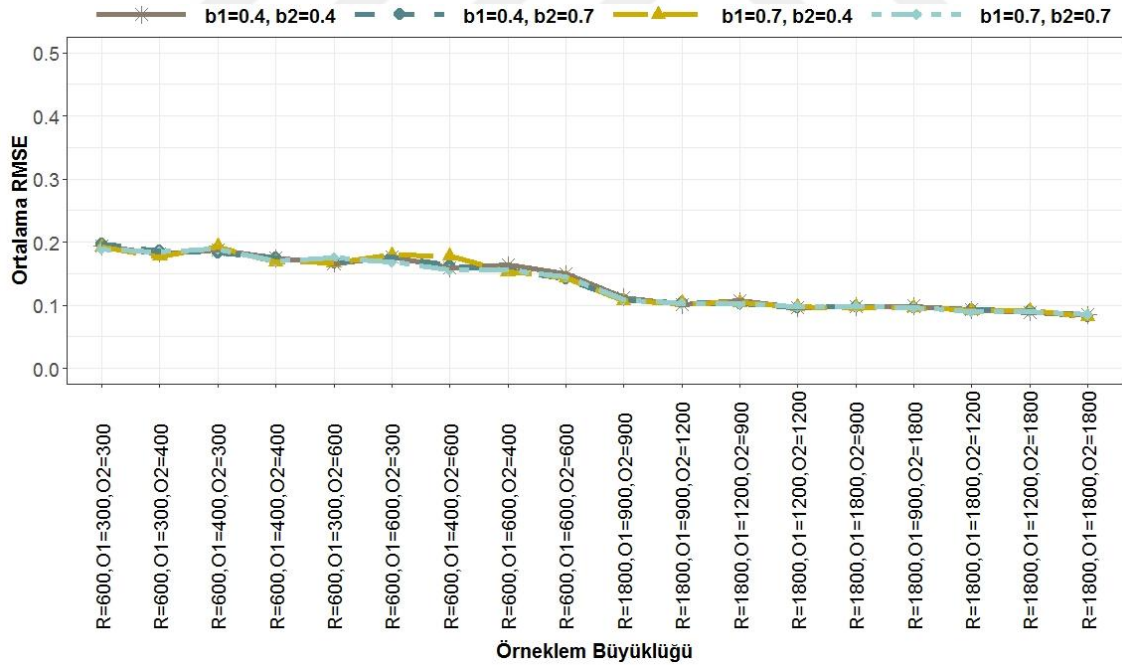
Madde Güçlük Parametresi İçin Mutlak Yanlılık ve RMSE Sonuçları

Referans Örneklem Büyükliği	Odak1 Örneklem Büyükliği	Odak2 Örneklem Büyükliği	Odak1 DMF Miktarı	Odak2 DMF Miktarı	Ortalama Mutlak yanlılık	Ortalama RMSE
1800	900	1800	0.70	0.70	0.0655	0.0835
1800	1200	900	0.40	0.40	0.0793	0.1000
1800	1200	900	0.70	0.40	0.0706	0.0926
1800	1200	900	0.40	0.70	0.0757	0.0950
1800	1200	900	0.70	0.70	0.0800	0.0991
1800	1200	1200	0.40	0.40	0.0651	0.0823
1800	1200	1200	0.70	0.40	0.0658	0.0848
1800	1200	1200	0.40	0.70	0.0746	0.0954
1800	1200	1200	0.70	0.70	0.0659	0.0827
1800	1200	1800	0.40	0.40	0.0589	0.0746
1800	1200	1800	0.70	0.40	0.0626	0.0793
1800	1200	1800	0.40	0.70	0.0611	0.0787
1800	1200	1800	0.70	0.70	0.0648	0.0829
1800	1800	900	0.40	0.40	0.0766	0.1028
1800	1800	900	0.70	0.40	0.0717	0.0896
1800	1800	900	0.40	0.70	0.0671	0.0868
1800	1800	900	0.70	0.70	0.0729	0.0922
1800	1800	1200	0.40	0.40	0.0604	0.0762
1800	1800	1200	0.70	0.40	0.0639	0.0809
1800	1800	1200	0.40	0.70	0.0674	0.0856
1800	1800	1200	0.70	0.70	0.0631	0.0806
1800	1800	1800	0.40	0.40	0.0545	0.0686
1800	1800	1800	0.70	0.40	0.0586	0.0735
1800	1800	1800	0.40	0.70	0.0670	0.0831
1800	1800	1800	0.70	0.70	0.0658	0.0825

EK 12. Madde Güçlük Parametresi İçin Ortalama RMSE

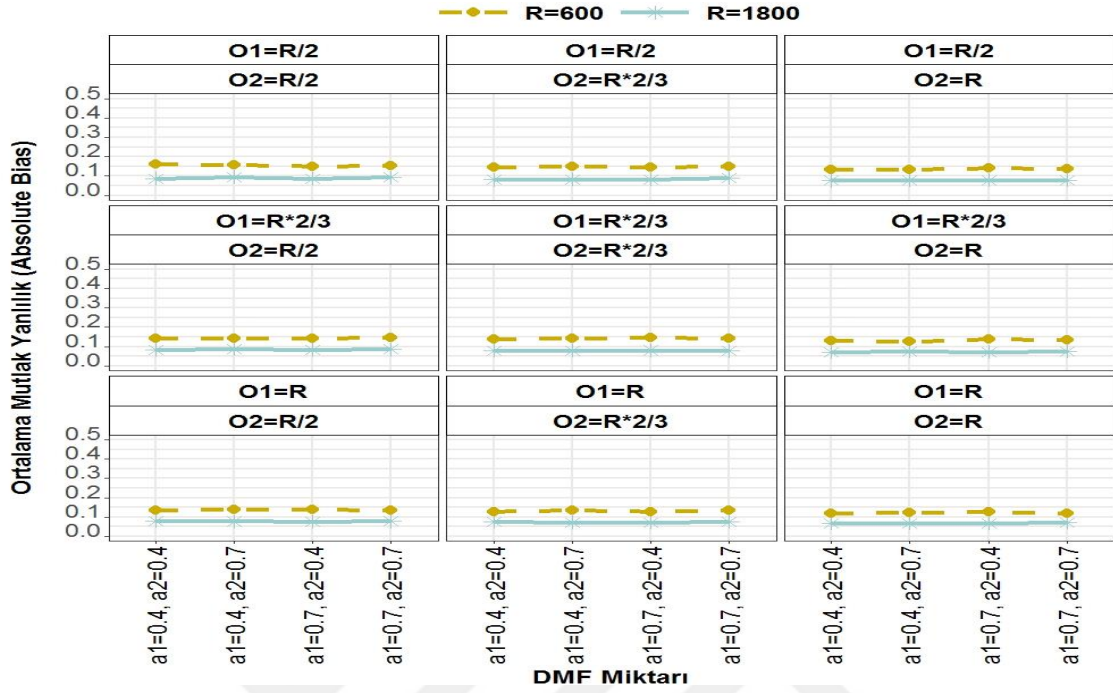


Şekil 12.1. Odak Grupların Örneklem Büyüklüğü Oranlarına ile DMF Miktarlarına Göre Madde Güçlük Parametresi Etkileşimi İçin Elde Edilen Ortalama RMSE Değerleri

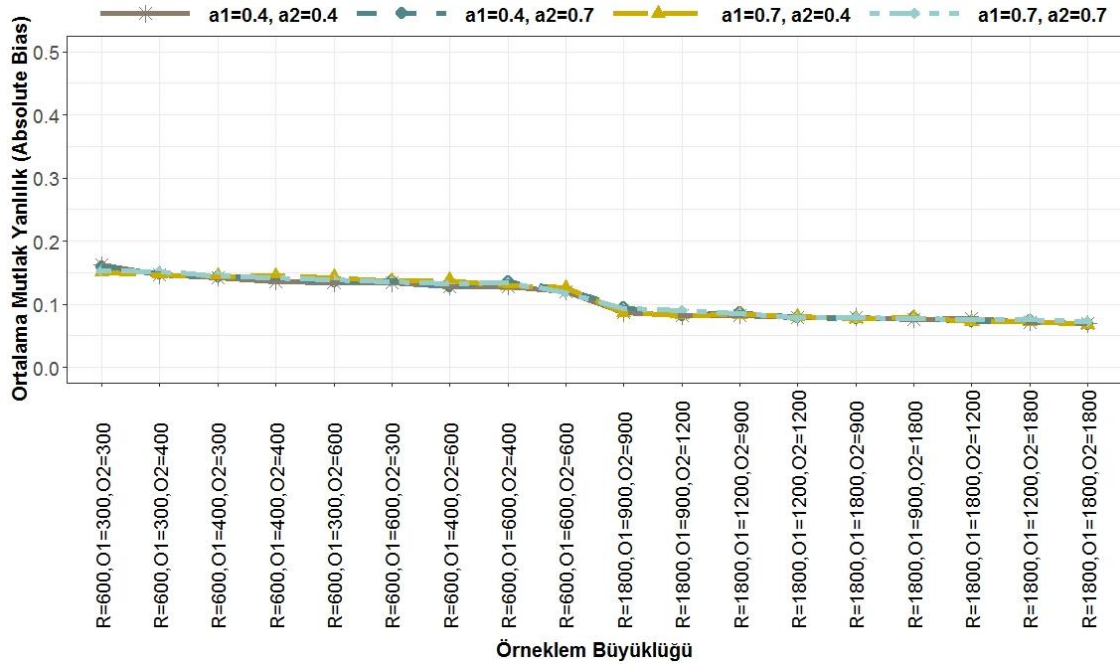


Şekil 12.2. Grupların Örneklem Büyüklükleri ile DMF Miktarlarına Göre Madde Güçlük Parametresi Etkileşimi İçin Elde Edilen Ortalama RMSE Değerleri

EK 13. Madde Ayırt Edicilik Parametresi İçin Ortalama Mutlak Yanlılık



Şekil 13.1. Odak Grupların Örneklem Büyüklüğü Oranlarına ile DMF Miktarlarına Göre Madde Ayırt Edicilik Parametresi Etkileşimi İçin Elde Edilen Ortalama Mutlak Yanlılık Değerleri



Şekil 13.2. Grupların Örneklem Büyüklükleri ile DMF Miktarlarına Göre Madde Ayırt Edicilik Parametresi Etkileşimi İçin Elde Edilen Ortalama Mutlak Yanlılık Değerleri

EK 14. Madde Ayırt Edicilik Parametresine İlişkin Ortalama Mutlak Yanlılık ve Ortalama RMSE Sonuçları

Tablo 14.1

Madde Ayırt Edicilik Parametresi İçin Mutlak Yanlılık ve RMSE Sonuçları

Referans Örnekleme Büyükliği	Odak1 Örnekleme Büyükliği	Odak2 Örnekleme Büyükliği	Odak1 DMF Miktarı	Odak2 DMF Miktarı	Ortalama Mutlak yanlılık	Ortalama RMSE
600	300	300	0.40	0.40	0.1618	0.2100
600	300	300	0.70	0.40	0.1502	0.1922
600	300	300	0.40	0.70	0.1584	0.2037
600	300	300	0.70	0.70	0.1530	0.1984
600	300	400	0.40	0.40	0.1461	0.1899
600	300	400	0.70	0.40	0.1470	0.1868
600	300	400	0.40	0.70	0.1486	0.1885
600	300	400	0.70	0.70	0.1508	0.1961
600	300	600	0.40	0.40	0.1334	0.1734
600	300	600	0.70	0.40	0.1425	0.1822
600	300	600	0.40	0.70	0.1350	0.1721
600	300	600	0.70	0.70	0.1379	0.1779
600	400	300	0.40	0.40	0.1423	0.1828
600	400	300	0.70	0.40	0.1439	0.1841
600	400	300	0.40	0.70	0.1427	0.1840
600	400	300	0.70	0.70	0.1455	0.1893
600	400	400	0.40	0.40	0.1363	0.1724
600	400	400	0.70	0.40	0.1446	0.1826
600	400	400	0.40	0.70	0.1416	0.1793
600	400	400	0.70	0.70	0.1411	0.1821
600	400	600	0.40	0.40	0.1282	0.1636
600	400	600	0.70	0.40	0.1379	0.1754
600	400	600	0.40	0.70	0.1280	0.1629
600	400	600	0.70	0.70	0.1324	0.1724
600	600	300	0.40	0.40	0.1334	0.1711
600	600	300	0.70	0.40	0.1368	0.1764
600	600	300	0.40	0.70	0.1378	0.1778
600	600	300	0.70	0.70	0.1354	0.1749
600	600	400	0.40	0.40	0.1279	0.1644
600	600	400	0.70	0.40	0.1287	0.1652
600	600	400	0.40	0.70	0.1357	0.1741
600	600	400	0.70	0.70	0.1342	0.1691
600	600	600	0.40	0.40	0.1203	0.1528
600	600	600	0.70	0.40	0.1255	0.1590
600	600	600	0.40	0.70	0.1235	0.1552
600	600	600	0.70	0.70	0.1181	0.1501
1800	900	900	0.40	0.40	0.0862	0.1091
1800	900	900	0.70	0.40	0.0864	0.1111
1800	900	900	0.40	0.70	0.0947	0.1230
1800	900	900	0.70	0.70	0.0927	0.1182
1800	900	1200	0.40	0.40	0.0814	0.1038
1800	900	1200	0.70	0.40	0.0826	0.1056
1800	900	1200	0.40	0.70	0.0819	0.1042
1800	900	1200	0.70	0.70	0.0903	0.1150
1800	900	1800	0.40	0.40	0.0760	0.0984
1800	900	1800	0.70	0.40	0.0782	0.1007
1800	900	1800	0.40	0.70	0.0784	0.1013

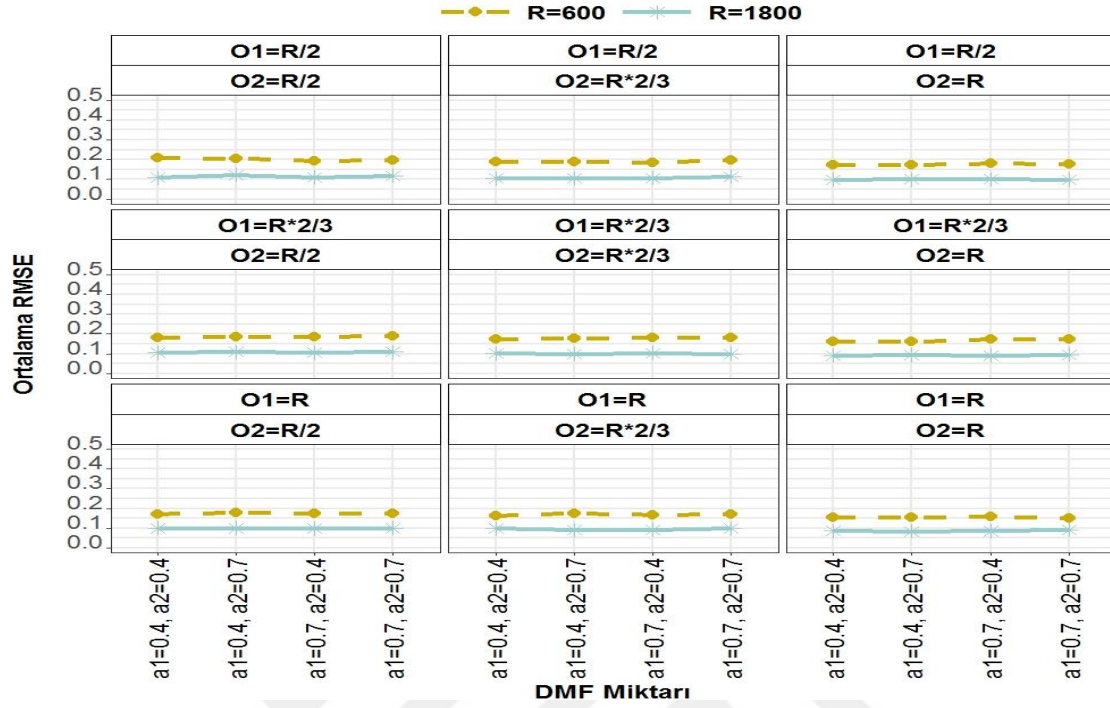
(Devam ediyor)

Tablo 14.1 (Devam)

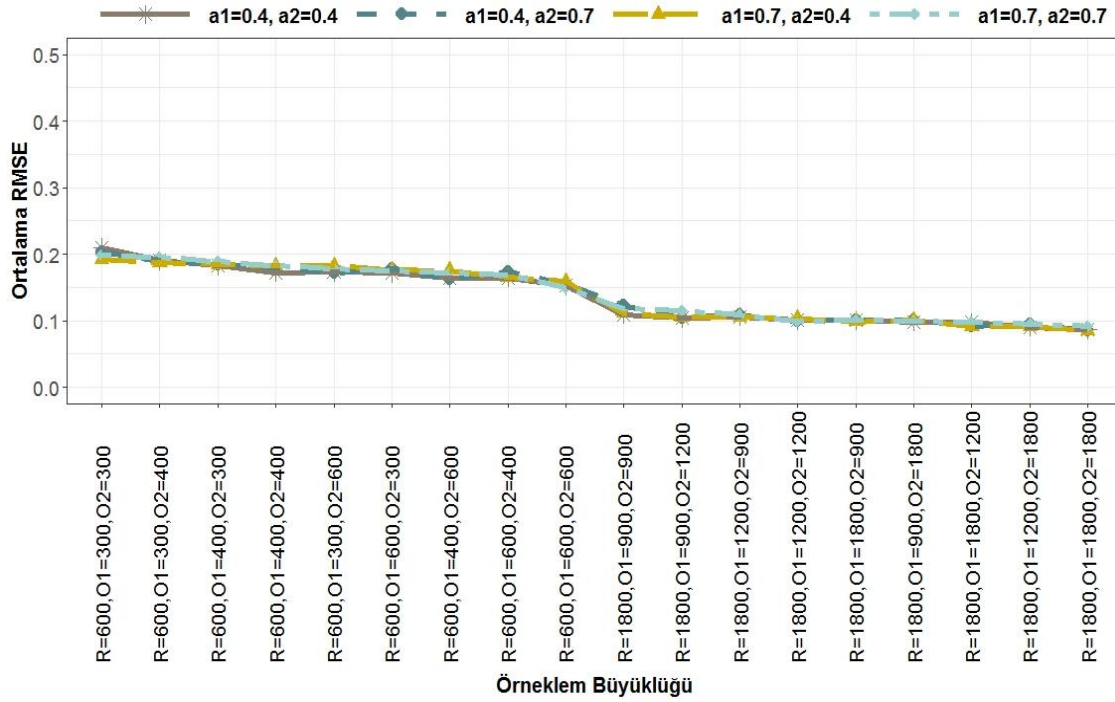
Madde Ayırt Edicilik Parametresi İçin Mutlak Yanlılık ve RMSE Sonuçları

Referans Örneklem Büyükliği	Odak1 Örneklem Büyükliği	Odak2 Örneklem Büyükliği	Odak1 DMF Miktarı	Odak2 DMF Miktarı	Ortalama Mutlak yanlılık	Ortalama RMSE
1800	900	1800	0.70	0.70	0.0779	0.0993
1800	1200	900	0.40	0.40	0.0828	0.1059
1800	1200	900	0.70	0.40	0.0828	0.1050
1800	1200	900	0.40	0.70	0.0866	0.1091
1800	1200	900	0.70	0.70	0.0853	0.1091
1800	1200	1200	0.40	0.40	0.0789	0.1008
1800	1200	1200	0.70	0.40	0.0804	0.1029
1800	1200	1200	0.40	0.70	0.0791	0.0994
1800	1200	1200	0.70	0.70	0.0784	0.0998
1800	1200	1800	0.40	0.40	0.0713	0.0905
1800	1200	1800	0.70	0.40	0.0719	0.0908
1800	1200	1800	0.40	0.70	0.0746	0.0944
1800	1200	1800	0.70	0.70	0.0756	0.0954
1800	1800	900	0.40	0.40	0.0793	0.1008
1800	1800	900	0.70	0.40	0.0770	0.0981
1800	1800	900	0.40	0.70	0.0773	0.1003
1800	1800	900	0.70	0.70	0.0786	0.1005
1800	1800	1200	0.40	0.40	0.0765	0.0976
1800	1800	1200	0.70	0.40	0.0721	0.0914
1800	1800	1200	0.40	0.70	0.0720	0.0910
1800	1800	1200	0.70	0.70	0.0764	0.0973
1800	1800	1800	0.40	0.40	0.0692	0.0865
1800	1800	1800	0.70	0.40	0.0680	0.0852
1800	1800	1800	0.40	0.70	0.0678	0.0850
1800	1800	1800	0.70	0.70	0.0721	0.0913

EK 15. Madde Ayırt Edicilik Parametresi İçin Ortalama RMSE



Şekil 15.1. Odak Grupların Örneklem Büyüklüğü Oranları ile DMF Miktarlarına Göre Madde Ayırt Edicilik Parametresi Etkileşimi İçin Elde Edilen Ortalama RMSE Değerleri



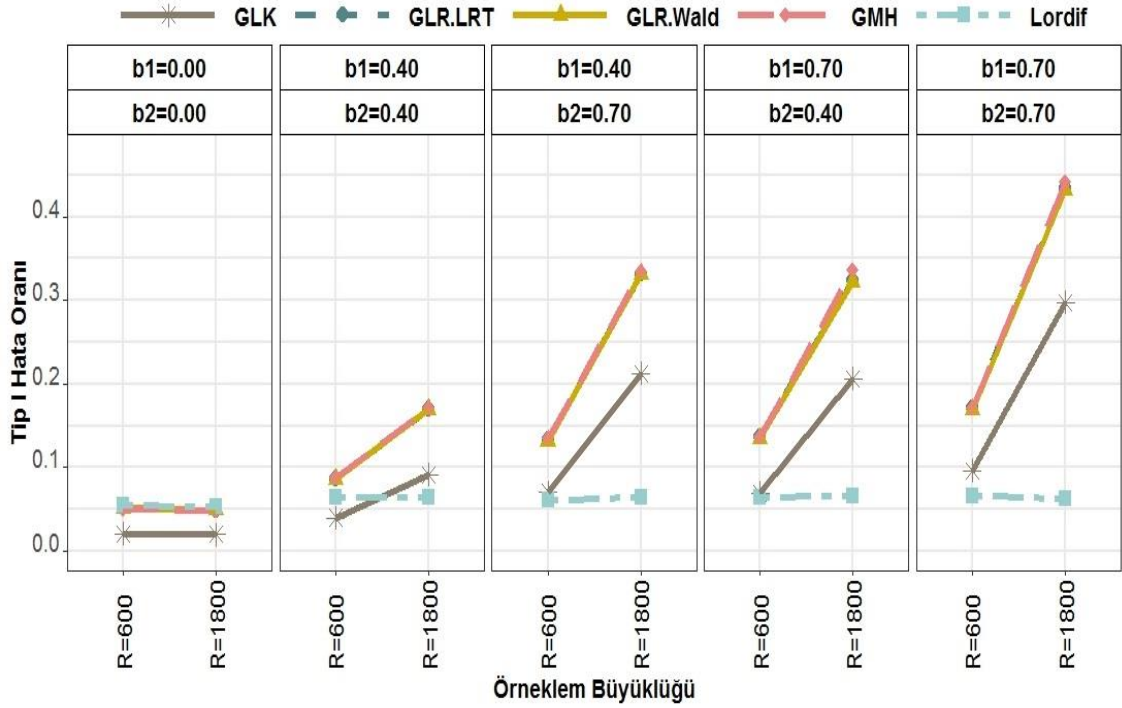
Şekil 15.2. Grupların Örneklem Büyüklükleri ile DMF Miktarlarına Göre Madde Ayırt Edicilik Parametresi Etkileşimi İçin Elde Edilen Ortalama RMSE Değerleri

EK 16. Üç ve Dört Grupta TB DMF Türü İçin Tip I Hata Oranları

Tablo 16.1

Üç Grupta TB DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip I Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı	Odak2 DMF Miktarı	GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
			Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata
$N_R=600$	0.00	0.00	.020	.002	.051	.003	.050	.002	.049	.003	.055	.003
	0.40	0.40	.040	.002	.086	.003	.086	.003	.088	.002	.064	.001
	0.40	0.70	.070	.003	.133	.005	.132	.005	.135	.004	.061	.002
	0.70	0.40	.069	.005	.137	.007	.135	.007	.137	.006	.064	.003
	0.70	0.70	.096	.006	.171	.004	.169	.004	.172	.004	.066	.002
$N_R=1800$	0.00	0.00	.020	.003	.050	.002	.049	.002	.048	.002	.054	.002
	0.40	0.40	.091	.004	.170	.006	.169	.007	.172	.005	.064	.002
	0.40	0.70	.212	.011	.333	.014	.332	.014	.335	.016	.065	.002
	0.70	0.40	.206	.011	.324	.014	.323	.014	.337	.013	.066	.001
	0.70	0.70	.298	.000	.434	.009	.433	.009	.441	.010	.063	.002

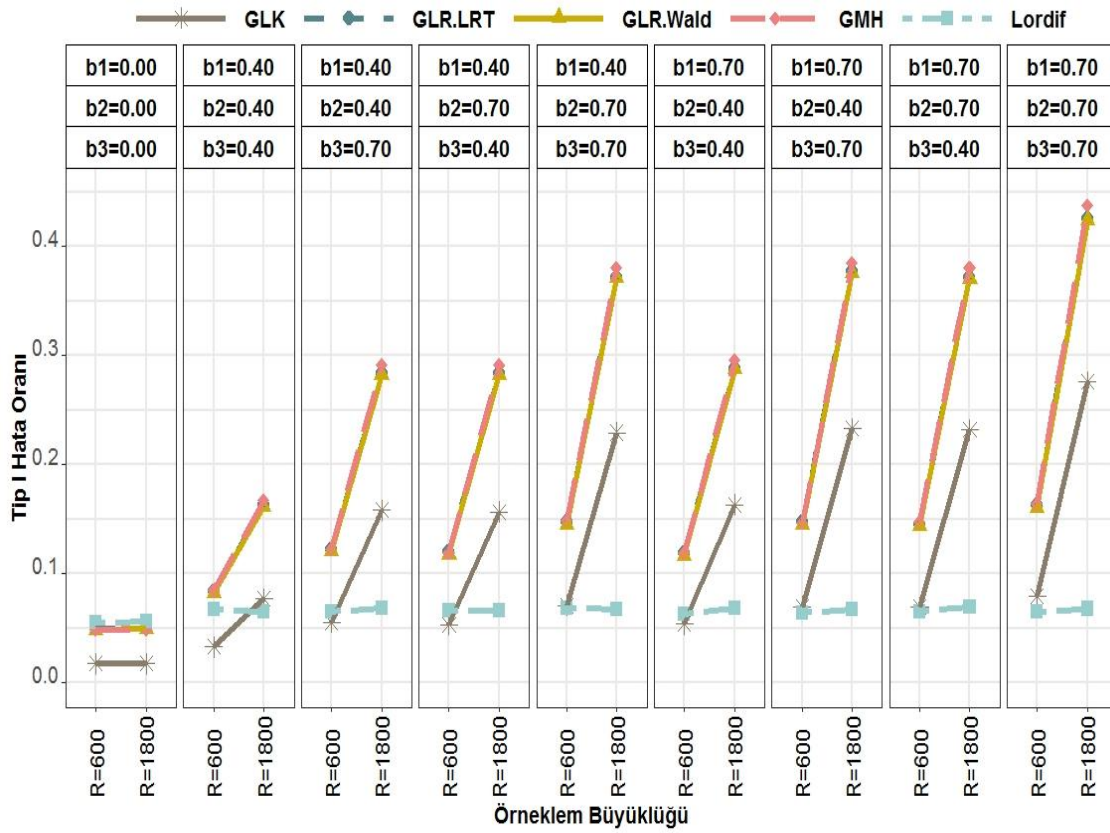


Şekil 16.1. Üç Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TB DMF Miktarı Etkileşimi İçin Tip I Hata Oranları

Tablo 16.2

Dört Grupta TB DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip I Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı	Odak2 DMF Miktarı	Odak3 DMF Miktarı	GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
				Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata
N _R =600	0.00	0.00	0.00	.018	.001	.050	.001	.049	.001	.049	.001	.056	.001
	0.40	0.40	0.40	.033	.001	.085	.001	.083	.001	.085	.001	.068	.001
	0.40	0.40	0.70	.056	.002	.123	.002	.121	.002	.122	.002	.065	.001
	0.40	0.70	0.40	.054	.001	.120	.002	.118	.002	.119	.002	.066	.001
	0.40	0.70	0.70	.071	.002	.148	.002	.146	.002	.148	.002	.068	.001
	0.70	0.40	0.40	.054	.002	.119	.002	.117	.002	.120	.002	.064	.001
	0.70	0.40	0.70	.070	.001	.147	.002	.145	.002	.147	.002	.065	.001
	0.70	0.70	0.40	.069	.002	.146	.003	.144	.003	.148	.003	.065	.001
	0.70	0.70	0.70	.080	.002	.163	.002	.160	.002	.164	.002	.066	.001
N _R =1800	0.00	0.00	0.00	.018	.001	.050	.001	.049	.001	.049	.001	.056	.001
	0.40	0.40	0.40	.077	.001	.163	.002	.162	.002	.168	.002	.065	.001
	0.40	0.40	0.70	.158	.005	.284	.007	.283	.007	.292	.007	.068	.001
	0.40	0.70	0.40	.157	.005	.284	.007	.283	.007	.291	.007	.066	.002
	0.40	0.70	0.70	.230	.004	.372	.006	.371	.006	.381	.006	.068	.001
	0.70	0.40	0.40	.163	.004	.289	.006	.288	.006	.297	.006	.069	.001
	0.70	0.40	0.70	.233	.006	.378	.006	.376	.006	.385	.006	.068	.001
	0.70	0.70	0.40	.232	.005	.372	.006	.371	.006	.380	.006	.070	.001
	0.70	0.70	0.70	.276	.003	.426	.004	.424	.004	.438	.004	.068	.001



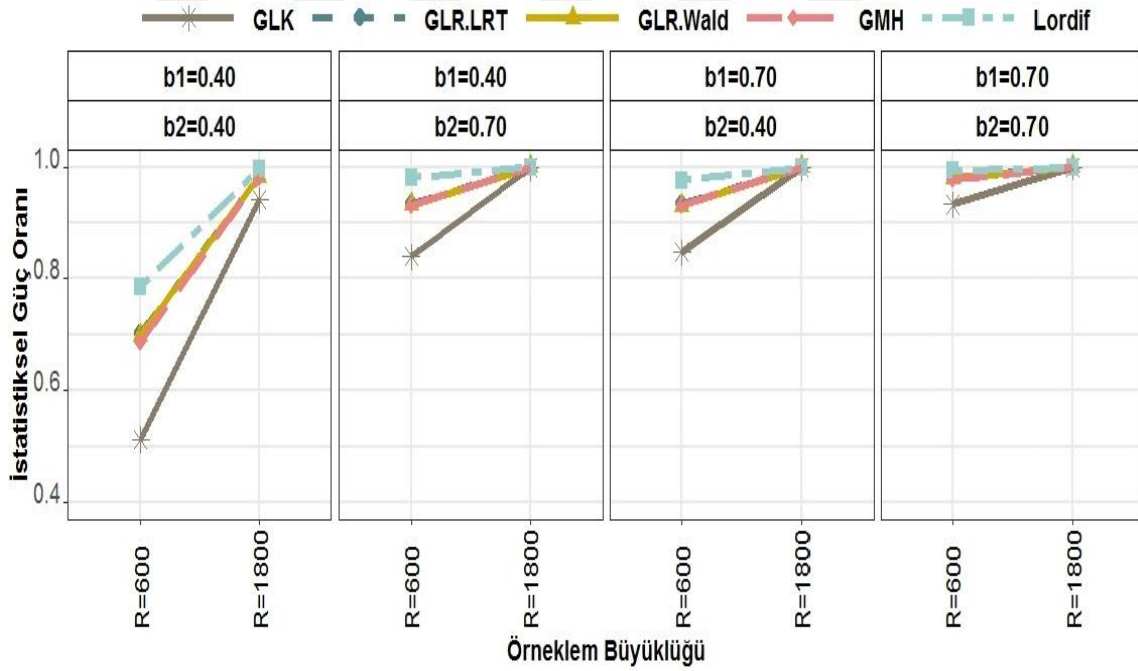
Şekil 16.2. Dört Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TB DMF Miktarı Etkileşimi İçin Tip I Hata Oranları

EK 17. Üç ve Dört Grupta TB DMF Türü İçin İstatistiksel Güç Oranları

Tablo 17.1

Üç Grupta TB DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama İstatistiksel Güç Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı	Odak2 DMF Miktarı	GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
			Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata
$N_R=600$	0.40	0.40	.513	.017	.702	.013	.700	.014	.686	.013	.785	.014
	0.40	0.70	.839	.017	.934	.011	.932	.011	.931	.011	.979	.005
	0.70	0.40	.847	.018	.933	.013	.931	.013	.932	.012	.975	.007
	0.70	0.70	.933	.010	.981	.004	.981	.004	.979	.004	.993	.001
$N_R=1800$	0.40	0.40	.941	.009	.983	.004	.983	.004	.981	.004	.996	.001
	0.40	0.70	.999	.001	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.40	.998	.001	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000

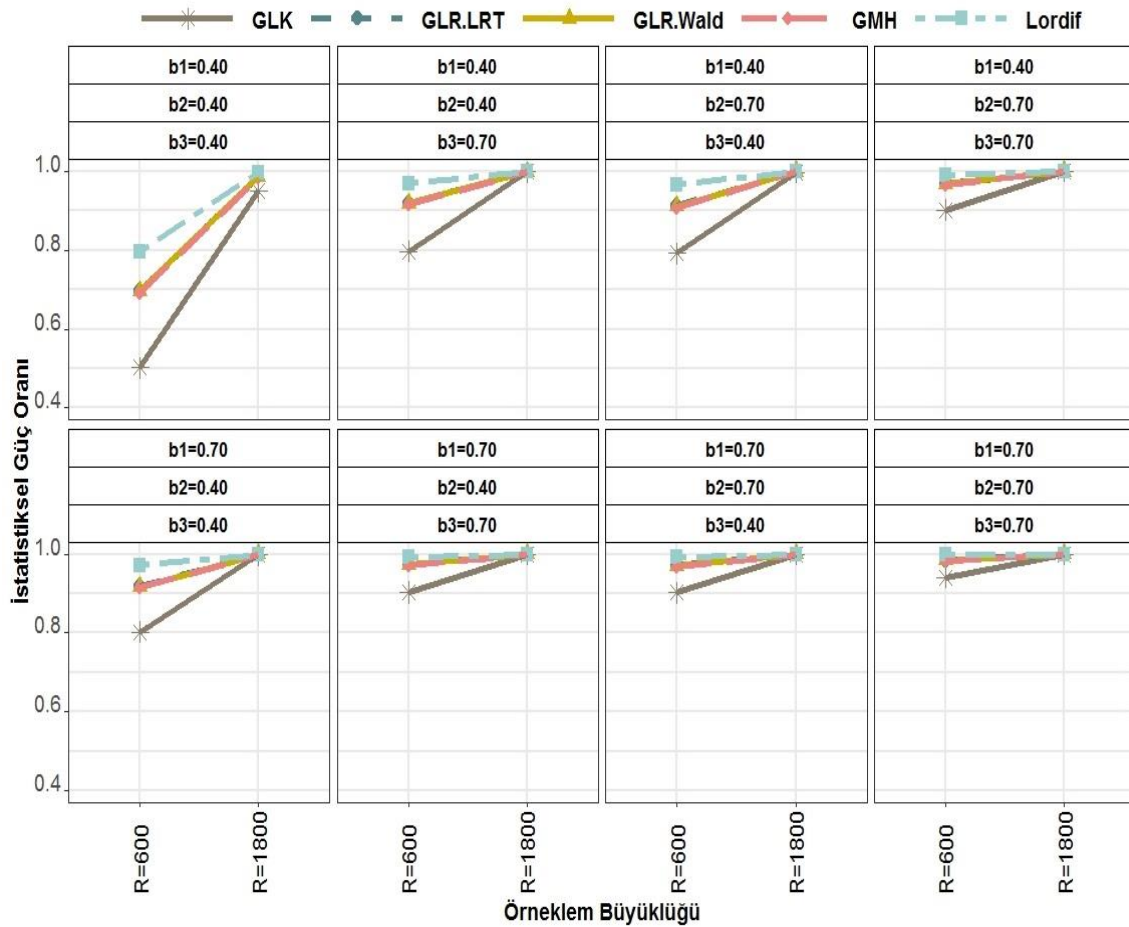


Şekil 17.1. Üç Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TB DMF Miktarı Etkileşimi İçin İstatistiksel Güç Oranları

Tablo 17.2

Dört Grupta TB DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama İstatistiksel Güç Oranları ve Standart Hatalar

Referans	Odak1	Odak2	Odak3	GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
Örneklem	DMF	DMF	DMF	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata
Büyüklük	Miktarı	Miktarı	Miktarı										
N _R =600	0.40	0.40	0.40	.502	.008	.699	.006	.697	.006	.692	.006	.798	.007
	0.40	0.40	0.70	.801	.010	.921	.007	.919	.007	.915	.007	.969	.004
	0.40	0.70	0.40	.800	.011	.912	.006	.911	.006	.907	.007	.964	.004
	0.40	0.70	0.70	.902	.006	.969	.003	.969	.004	.967	.004	.992	.001
	0.70	0.40	0.40	.802	.010	.919	.005	.917	.005	.915	.005	.971	.003
	0.70	0.40	0.70	.903	.005	.973	.002	.973	.002	.971	.002	.994	.001
	0.70	0.70	0.40	.902	.008	.970	.003	.970	.003	.967	.003	.992	.002
	0.70	0.70	0.70	.941	.004	.985	.001	.984	.001	.984	.001	.997	.001
N _R =1800	0.40	0.40	0.40	.950	.004	.987	.002	.987	.002	.985	.002	.998	.001
	0.40	0.40	0.70	.998	.001	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40	0.70	0.40	.998	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40	0.70	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.40	0.40	.997	.001	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.40	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.70	0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.70	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000



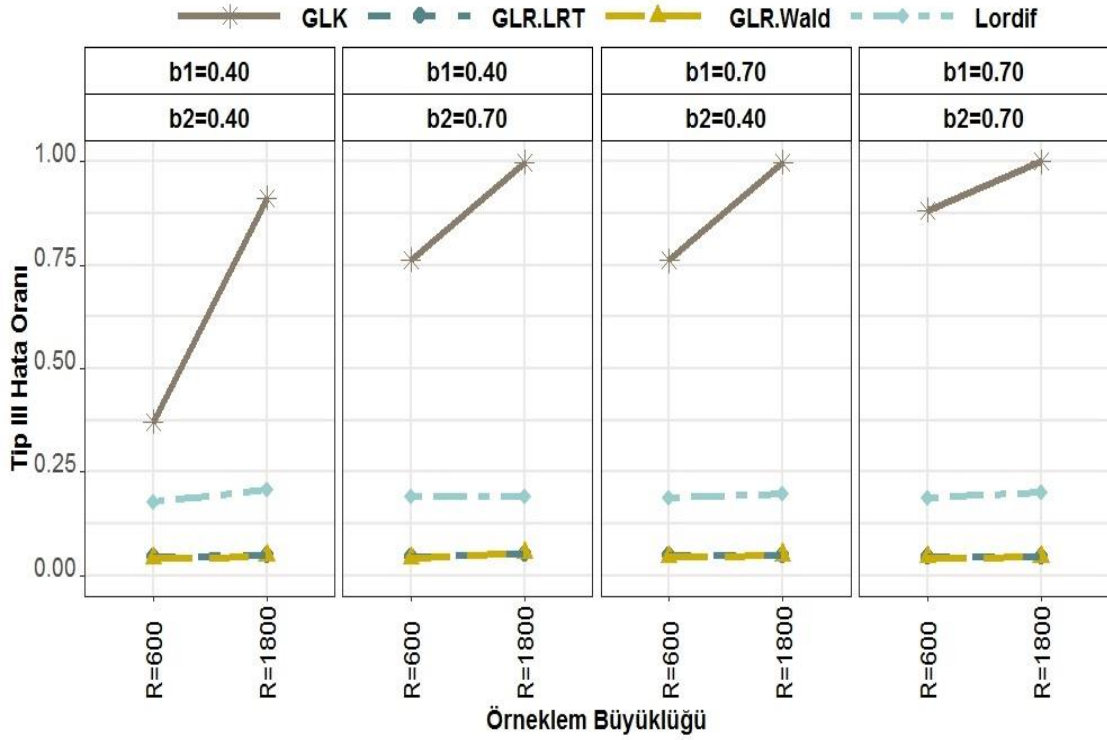
Şekil 17.2. Dört Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TB DMF Miktarı Etkileşimi İçin İstatistiksel Güç Oranları

EK 18. Üç ve Dört Grupta TB DMF Türü İçin Tip III Hata Oranları

Tablo 18.1

Üç Grupta TB DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip III Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı	Odak2 DMF Miktarı	GLK		GLR-LRT		GLR-Wald		Lordif	
			Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata
$N_R=600$	0.40	0.40	.372	.010	.045	.004	.043	.004	.179	.006
	0.40	0.70	.763	.024	.045	.002	.042	.002	.193	.009
	0.70	0.40	.760	.025	.048	.002	.044	.003	.189	.004
	0.70	0.70	.882	.012	.047	.003	.043	.003	.188	.009
$N_R=1800$	0.40	0.40	.911	.013	.048	.003	.047	.003	.209	.009
	0.40	0.70	.998	.001	.054	.003	.054	.003	.192	.004
	0.70	0.40	.997	.001	.051	.002	.050	.002	.200	.004
	0.70	0.70	1.000	.000	.046	.004	.046	.004	.203	.008

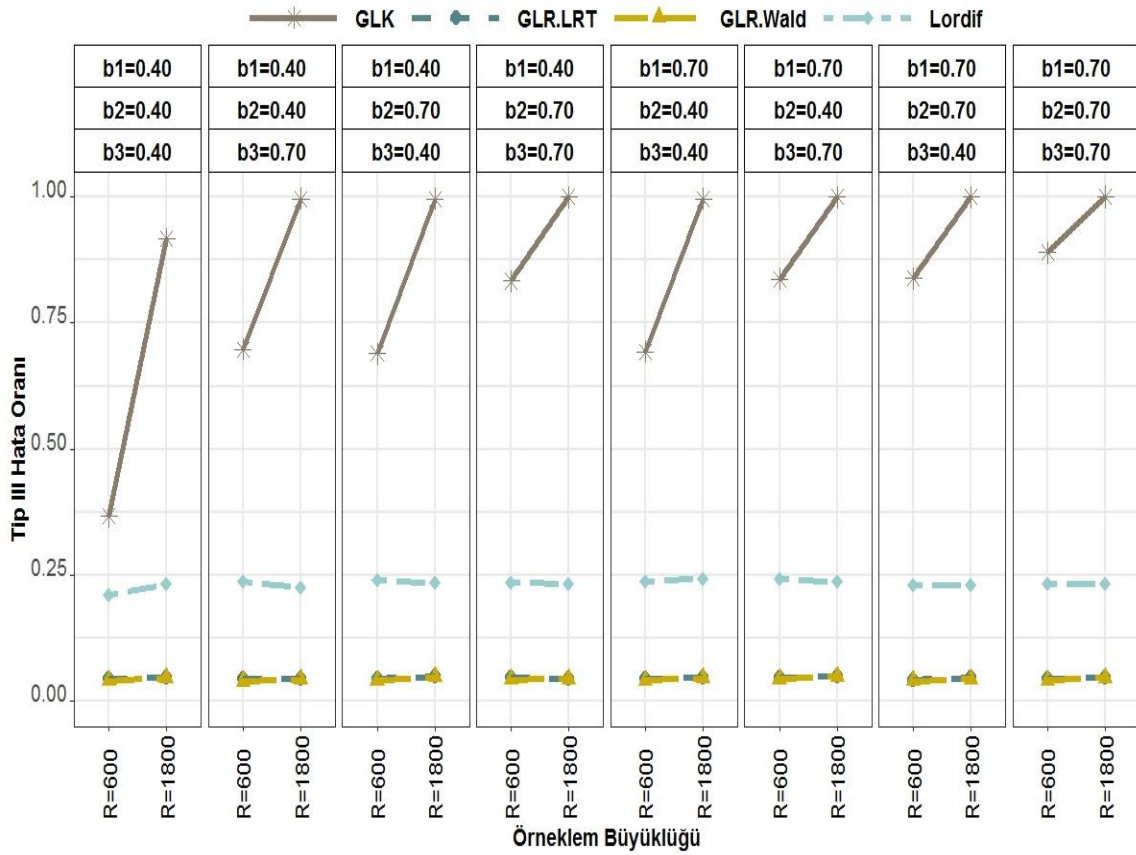


Şekil 18.1. Üç Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TB DMF Miktarı Etkileşimi İçin Tip III Hata Oranları

Tablo 18.2

Dört Grupta TB DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip III Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı	Odak2 DMF Miktarı	Odak3 DMF Miktarı	GLK		GLR-LRT		GLR-Wald		Lordif	
				Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata
N _R =600	0.40	0.40	0.40	.367	.008	.045	.002	.041	.002	.211	.004
	0.40	0.40	0.70	.698	.013	.045	.001	.041	.002	.238	.004
	0.40	0.70	0.40	.689	.013	.045	.002	.042	.002	.240	.005
	0.40	0.70	0.70	.833	.010	.047	.002	.043	.002	.236	.004
	0.70	0.40	0.40	.692	.012	.046	.002	.042	.002	.237	.004
	0.70	0.40	0.70	.837	.007	.048	.003	.044	.002	.243	.004
	0.70	0.70	0.40	.838	.009	.042	.002	.041	.002	.232	.004
	0.70	0.70	0.70	.890	.006	.045	.002	.043	.002	.233	.004
N _R =1800	0.40	0.40	0.40	.917	.005	.048	.002	.048	.002	.232	.005
	0.40	0.40	0.70	.996	.001	.044	.002	.044	.002	.227	.003
	0.40	0.70	0.40	.996	.001	.050	.002	.048	.002	.235	.005
	0.40	0.70	0.70	1.000	.000	.046	.002	.044	.001	.232	.004
	0.70	0.40	0.40	.996	.001	.047	.001	.046	.001	.242	.004
	0.70	0.40	0.70	1.000	.000	.050	.002	.050	.002	.238	.004
	0.70	0.70	0.40	1.000	.000	.046	.002	.045	.002	.231	.004
	0.70	0.70	0.70	1.000	.000	.048	.002	.047	.002	.233	.004



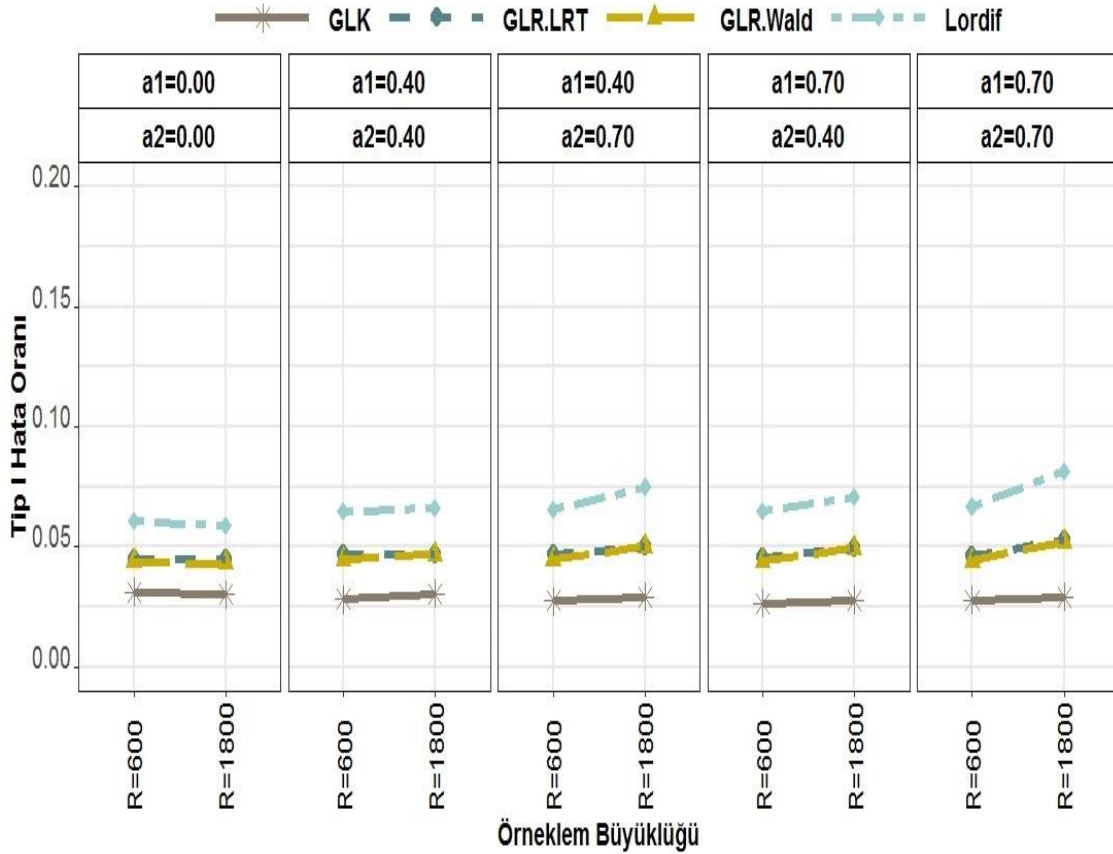
Şekil 18.2. Dört Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TB DMF Miktarı Etkileşimi İçin Tip III Hata Oranları

EK 19. Üç ve Dört Grupta TBO(a) DMF Türü İçin Tip I Hata Oranları

Tablo 19.1

Üç Grupta TBO(a) DMF İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip I Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı	Odak2 DMF Miktarı	GLK		GLR-LRT		GLR-Wald		Lordif	
			Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata
$N_R=600$	0.00	0.00	.030	.002	.045	.002	.044	.002	.061	.002
	0.40	0.40	.028	.001	.047	.002	.045	.002	.065	.002
	0.40	0.70	.027	.002	.047	.003	.045	.003	.066	.003
	0.70	0.40	.027	.002	.045	.002	.044	.002	.065	.002
	0.70	0.70	.028	.001	.046	.002	.044	.002	.067	.002
$N_R=1800$	0.00	0.00	.030	.001	.045	.002	.043	.002	.059	.002
	0.40	0.40	.030	.001	.047	.002	.047	.002	.066	.001
	0.40	0.70	.029	.001	.050	.002	.050	.002	.075	.002
	0.70	0.40	.028	.002	.050	.002	.049	.002	.071	.003
	0.70	0.70	.029	.002	.053	.002	.052	.002	.081	.002

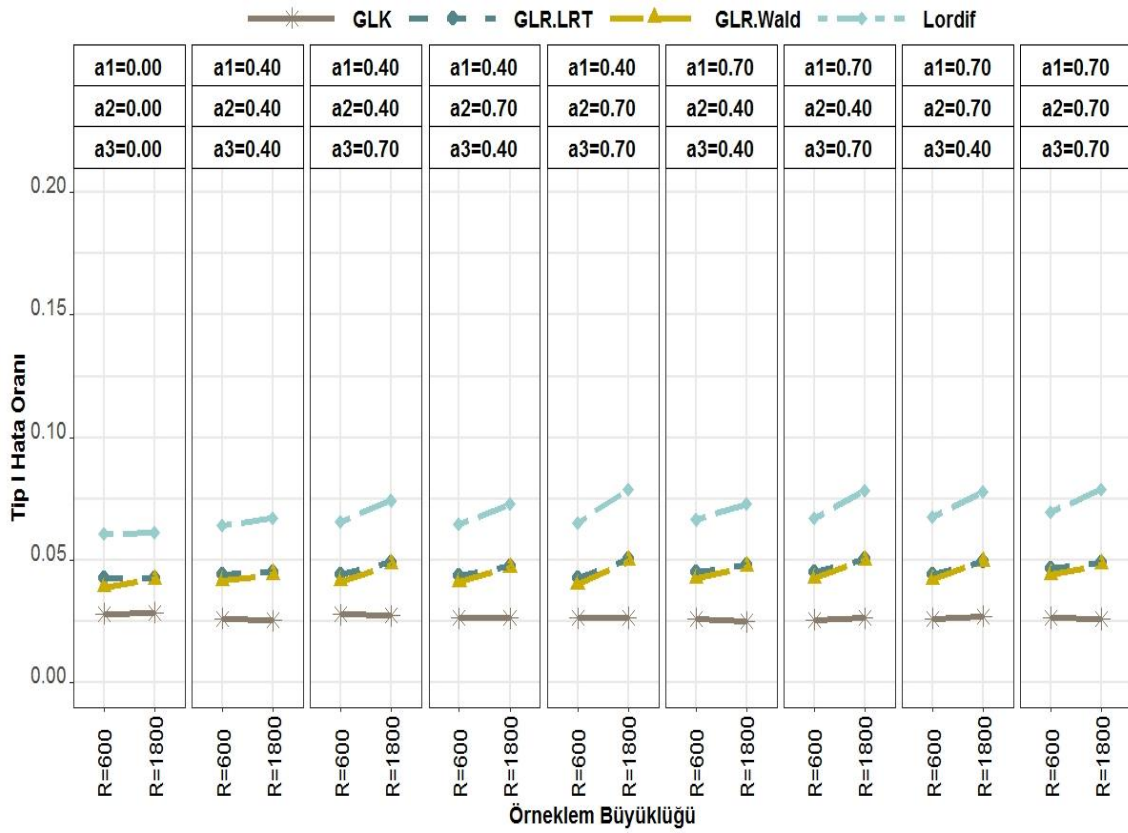


Şekil 19.1. Üç Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TBO(a) DMF Miktarı Etkileşimi İçin Tip I Hata Oranları

Tablo 19.2

Dört Grupta TBO(a) DMF Türü için Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip I Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı	Odak2 DMF Miktarı	Odak3 DMF Miktarı	GLK		GLR-LRT		GLR-Wald		Lordif	
				Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata
N _R =600	0.00	0.00	0.00	.028	.001	.043	.001	.042	.001	.061	.001
	0.40	0.40	0.40	.026	.001	.044	.001	.042	.001	.064	.002
	0.40	0.40	0.70	.028	.001	.044	.001	.041	.001	.066	.001
	0.40	0.70	0.40	.026	.001	.043	.001	.041	.001	.065	.001
	0.40	0.70	0.70	.026	.001	.043	.001	.040	.001	.065	.001
	0.70	0.40	0.40	.026	.001	.045	.001	.043	.001	.066	.001
	0.70	0.40	0.70	.026	.001	.045	.001	.042	.001	.067	.001
	0.70	0.70	0.40	.026	.001	.044	.001	.042	.001	.068	.001
	0.70	0.70	0.70	.027	.001	.047	.001	.044	.001	.070	.001
N _R =1800	0.00	0.00	0.00	.028	.001	.043	.001	.042	.001	.061	.001
	0.40	0.40	0.40	.026	.001	.045	.001	.044	.001	.067	.001
	0.40	0.40	0.70	.027	.001	.049	.001	.048	.001	.074	.001
	0.40	0.70	0.40	.026	.001	.048	.001	.047	.001	.073	.001
	0.40	0.70	0.70	.027	.001	.051	.001	.050	.001	.079	.001
	0.70	0.40	0.40	.025	.001	.048	.001	.047	.001	.073	.001
	0.70	0.40	0.70	.026	.001	.050	.001	.050	.001	.078	.001
	0.70	0.70	0.40	.027	.001	.050	.001	.049	.001	.078	.001
	0.70	0.70	0.70	.026	.001	.049	.001	.048	.001	.079	.001



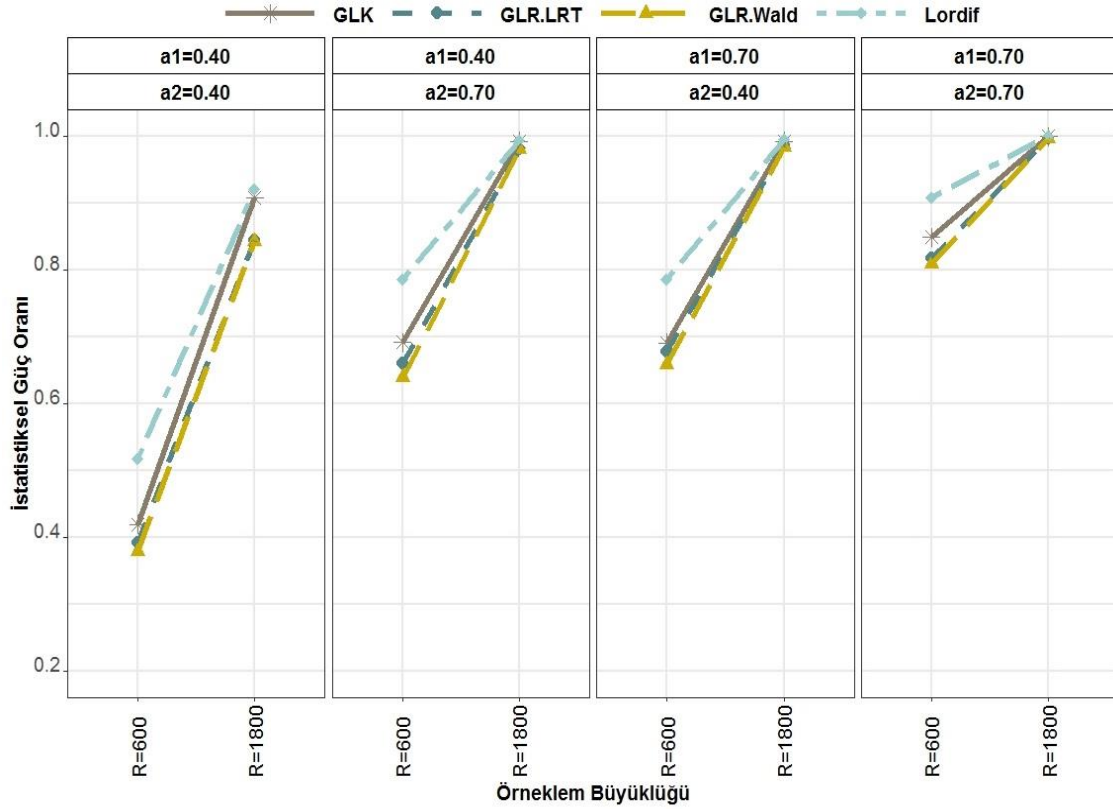
Şekil 19.2. Dört Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TBO(a) DMF Miktarı Etkileşimi İçin Tip I Hata Oranları

EK 20. Üç ve Dört Grupta TBO(a) DMF Türü İçin İstatistiksel Güç Oranları

Tablo 20.1

Üç Grupta TBO(a) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama İstatistiksel Güç Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı	Odak2 DMF Miktarı	GLK		GLR-LRT		GLR-Wald		Lordif	
			Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata
$N_R=600$	0.40	0.40	.419	.018	.391	.013	.378	.016	.517	.015
	0.40	0.70	.692	.029	.661	.029	.639	.032	.786	.021
	0.70	0.40	.691	.028	.678	.019	.658	.022	.785	.018
	0.70	0.70	.849	.014	.818	.010	.810	.011	.907	.006
$N_R=1800$	0.40	0.40	.909	.008	.845	.013	.843	.013	.921	.006
	0.40	0.70	.993	.002	.981	.005	.980	.005	.995	.002
	0.70	0.40	.993	.002	.986	.003	.984	.004	.996	.002
	0.70	0.70	1.000	.000	.997	.001	.997	.001	1.000	.000

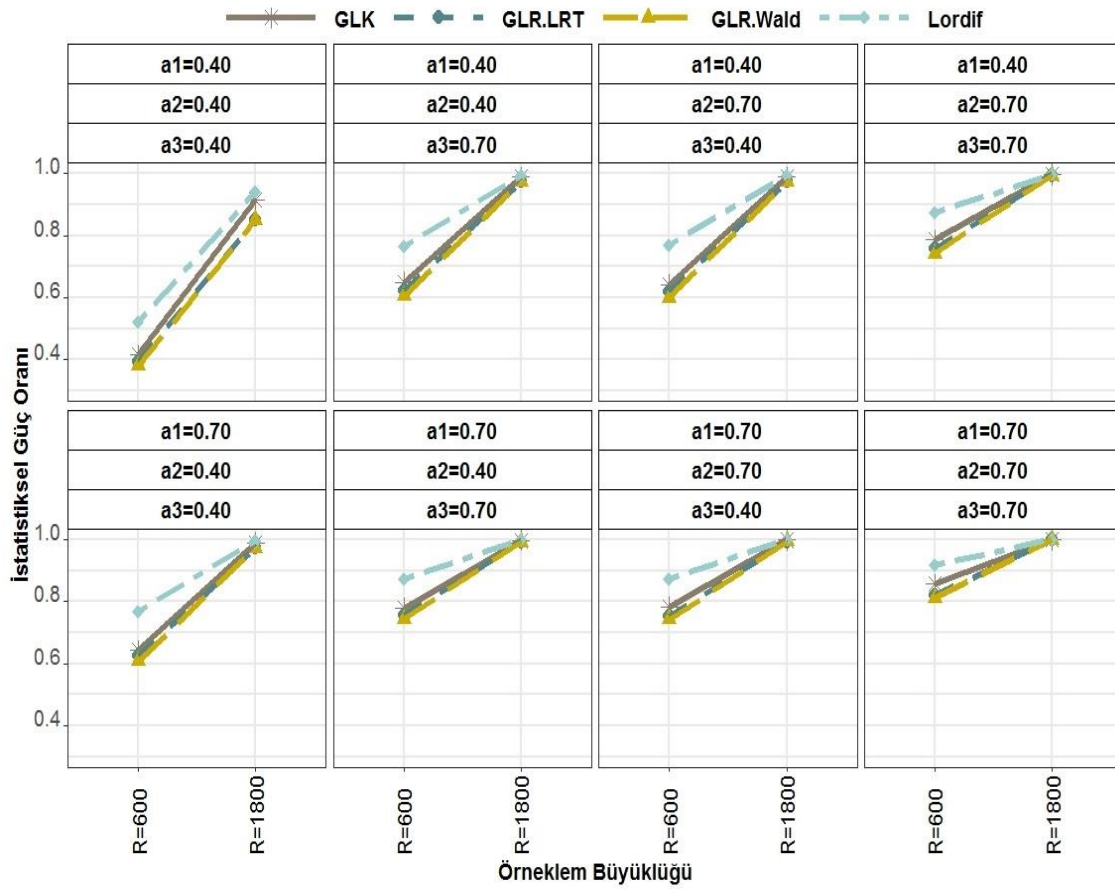


Şekil 20.1. Üç Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TBO(a) DMF Miktarı Etkileşimi İçin İstatistiksel Güç Oranları

Tablo 20.2

Dört Grupta TBO(a) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama İstatistiksel Güç Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı	Odak2 DMF Miktarı	Odak3 DMF Miktarı	GLK		GLR-LRT		GLR-Wald		Lordif	
				Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata
N _R =600	0.40	0.40	0.40	.414	.008	.392	.007	.380	.008	.521	.007
	0.40	0.40	0.70	.649	.014	.623	.010	.605	.012	.764	.010
	0.40	0.70	0.40	.642	.013	.617	.012	.599	.013	.770	.010
	0.40	0.70	0.70	.787	.010	.756	.008	.742	.009	.872	.006
	0.70	0.40	0.40	.642	.015	.625	.012	.608	.013	.769	.010
	0.70	0.40	0.70	.780	.010	.755	.008	.744	.009	.872	.006
	0.70	0.70	0.40	.784	.009	.753	.008	.743	.009	.874	.006
	0.70	0.70	0.70	.857	.006	.819	.006	.813	.006	.919	.005
N _R =1800	0.40	0.40	0.40	.917	.004	.852	.006	.852	.006	.942	.004
	0.40	0.40	0.70	.989	.001	.975	.003	.974	.003	.996	.001
	0.40	0.70	0.40	.989	.002	.975	.003	.974	.003	.995	.001
	0.40	0.70	0.70	.998	.001	.994	.001	.994	.001	.999	.000
	0.70	0.40	0.40	.990	.002	.974	.002	.973	.002	.995	.001
	0.70	0.40	0.70	.998	.001	.993	.001	.993	.001	.999	.000
	0.70	0.70	0.40	.999	.000	.995	.001	.994	.001	1.000	.000
	0.70	0.70	0.70	1.000	.000	.999	.000	.999	.000	1.000	.000



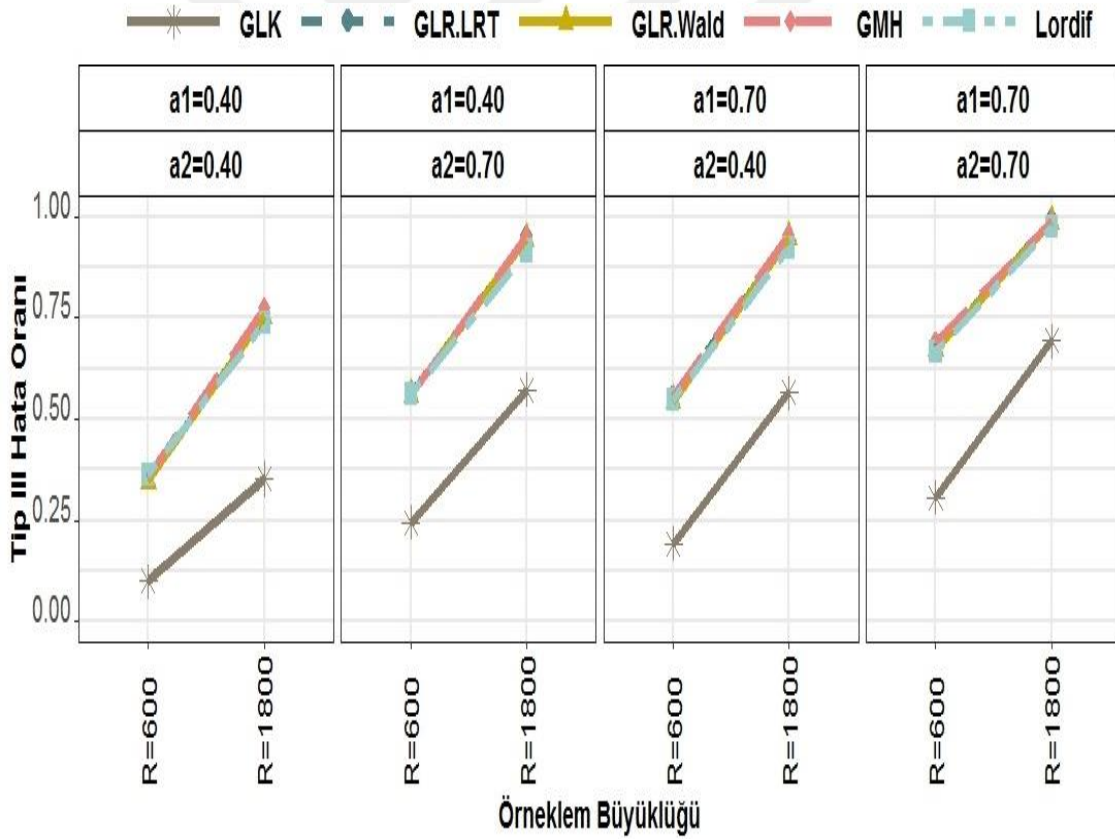
Şekil 20.2. Dört Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TBO(a) DMF Miktarı Etkileşimi İçin İstatistiksel Güç Oranları

EK 21. Üç ve Dört Grupta TBO(a) DMF Türü İçin Tip III Hata Oranları

Tablo 21.1

Üç Grupta TBO(a) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip III Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı	Odak2 DMF Miktarı	GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
			Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata
$N_R=600$	0.40	0.40	.101	.016	.351	.013	.347	.013	.357	.013	.364	.013
	0.40	0.70	.246	.039	.564	.020	.559	.020	.564	.021	.563	.019
	0.70	0.40	.192	.016	.551	.021	.545	.021	.557	.024	.545	.019
	0.70	0.70	.307	.051	.676	.024	.672	.025	.691	.021	.667	.022
$N_R=1800$	0.40	0.40	.353	.031	.756	.017	.754	.018	.777	.016	.736	.017
	0.40	0.70	.572	.018	.944	.011	.944	.011	.952	.009	.915	.014
	0.70	0.40	.565	.054	.947	.011	.947	.011	.954	.010	.920	.013
	0.70	0.70	.694	.031	.987	.003	.987	.003	.989	.003	.973	.004

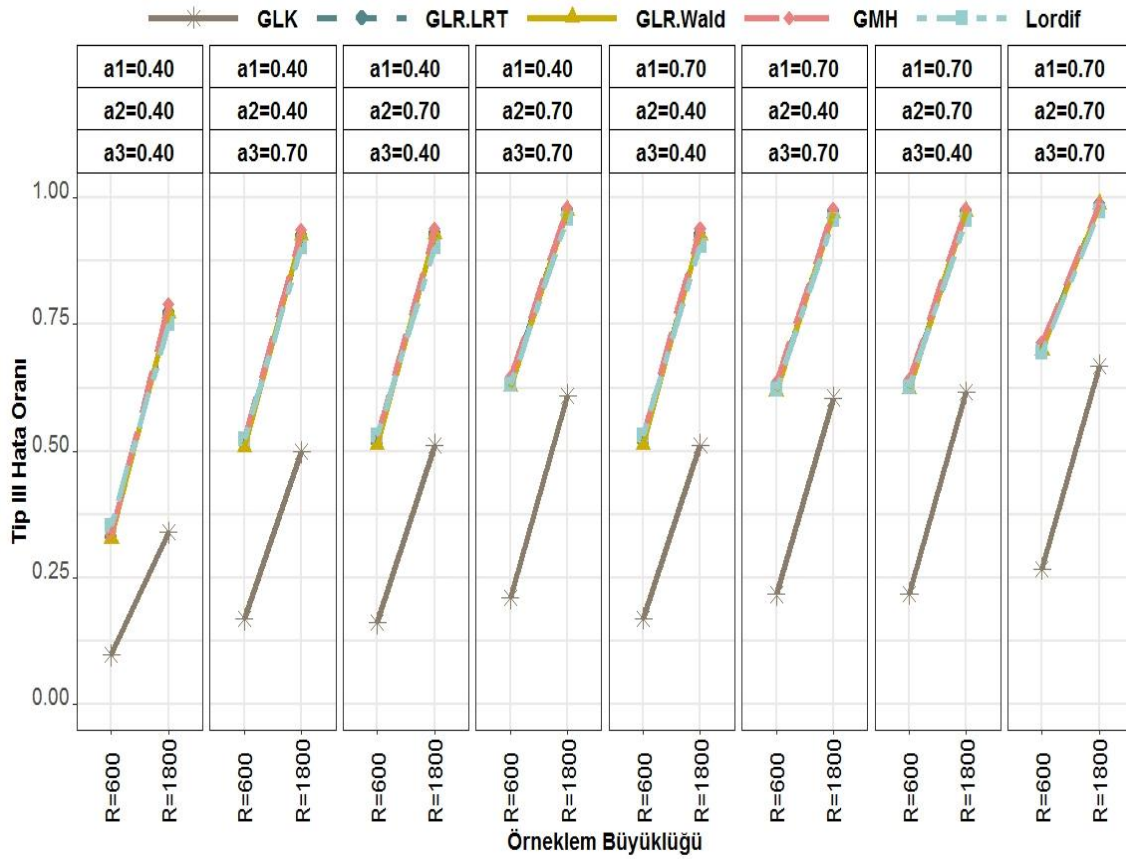


Şekil 21.1. Üç Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TBO(a) DMF Miktarı Etkileşimi İçin Tip III Hata Oranları

Tablo 21.2

Dört Grupta TBO(a) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip III Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı	Odak2 DMF Miktarı	Odak3 DMF Miktarı	GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
				Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata
N _R =600	0.40	0.40	0.40	.098	.004	.331	.006	.327	.007	.337	.007	.353	.007
	0.40	0.40	0.70	.169	.006	.513	.012	.508	.012	.524	.012	.524	.011
	0.40	0.70	0.40	.161	.006	.516	.011	.512	.011	.529	.011	.531	.010
	0.40	0.70	0.70	.211	.008	.630	.010	.627	.011	.646	.010	.632	.009
	0.70	0.40	0.40	.170	.007	.519	.012	.513	.012	.530	.012	.531	.011
	0.70	0.40	0.70	.217	.008	.622	.010	.619	.010	.636	.010	.622	.008
	0.70	0.70	0.40	.219	.008	.626	.009	.623	.009	.642	.010	.626	.008
	0.70	0.70	0.70	.267	.008	.701	.006	.698	.006	.715	.006	.696	.005
N _R =1800	0.40	0.40	0.40	.340	.008	.774	.006	.773	.006	.790	.005	.750	.005
	0.40	0.40	0.70	.498	.012	.925	.005	.925	.005	.935	.005	.902	.005
	0.40	0.70	0.40	.512	.010	.929	.005	.929	.005	.939	.005	.901	.006
	0.40	0.70	0.70	.610	.011	.975	.002	.975	.002	.980	.002	.958	.003
	0.70	0.40	0.40	.512	.010	.927	.005	.926	.005	.938	.004	.904	.005
	0.70	0.40	0.70	.606	.008	.971	.002	.970	.002	.977	.002	.955	.002
	0.70	0.70	0.40	.616	.009	.972	.002	.972	.002	.978	.002	.956	.003
	0.70	0.70	0.70	.669	.008	.986	.001	.986	.001	.989	.001	.973	.002



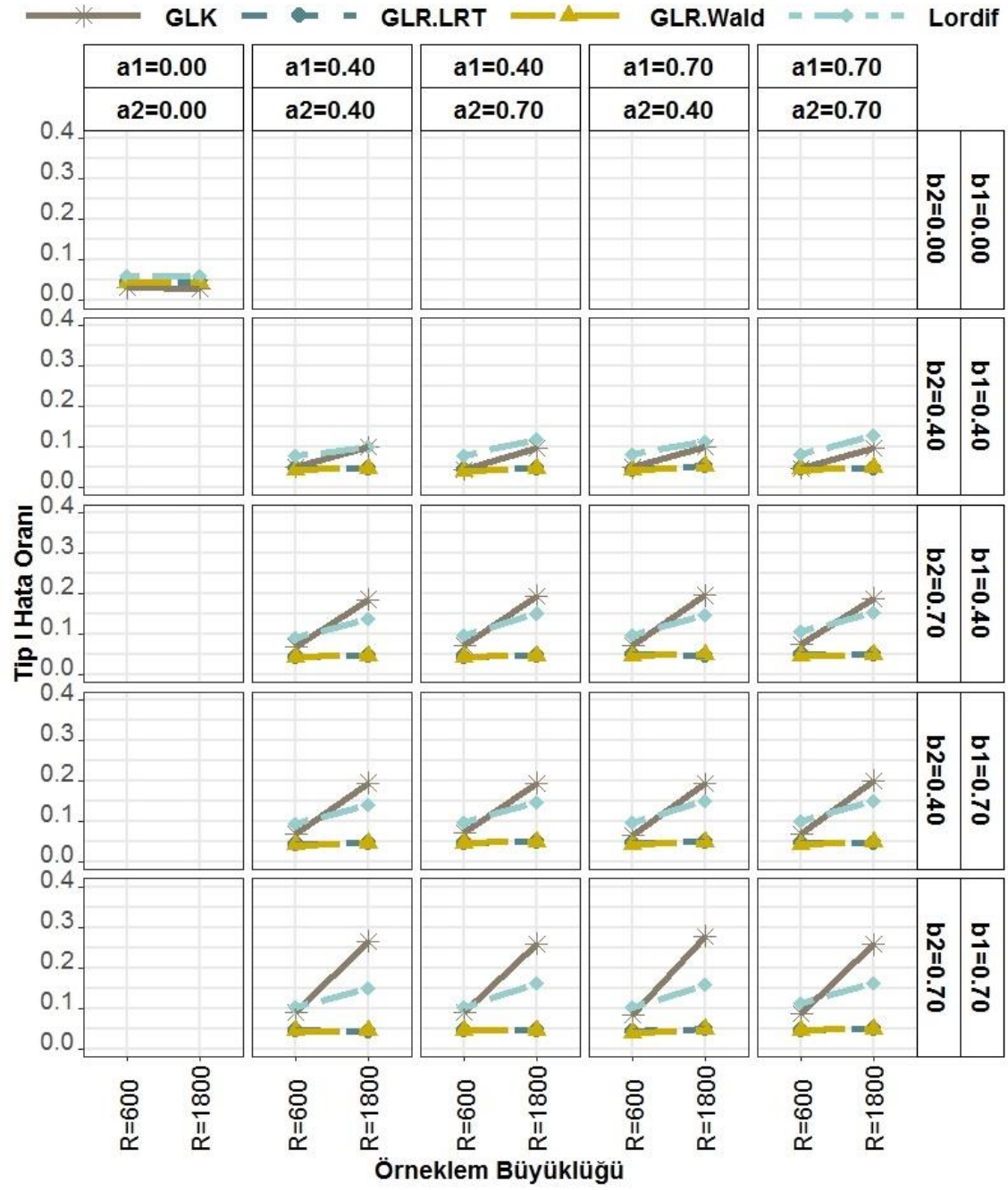
Şekil 21.2. Dört Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TBO(a) DMF Miktarı Etkileşimi İçin Tip III Hata Oranları

EK 22. Üç ve Dört Grupta TBO(ab) DMF Türü İçin Tip I Hata Oranları

Tablo 22.1

Üç Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip I Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı		Odak2 DMF Miktarı		GLK		GLR-LRT		GLR-Wald		Lordif	
					Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata
	a	b	a	b								
N _R =600	0.00	0.00	0.00	0.00	.030	.002	.045	.002	.044	.002	.061	.002
	0.40	0.40	0.40	0.40	.050	.002	.046	.001	.043	.001	.077	.002
	0.40	0.40	0.40	0.70	.069	.002	.044	.002	.042	.001	.089	.002
	0.40	0.40	0.70	0.40	.045	.003	.044	.002	.041	.002	.079	.002
	0.40	0.40	0.70	0.70	.071	.004	.045	.003	.042	.003	.095	.004
	0.40	0.70	0.40	0.40	.069	.003	.043	.002	.041	.002	.092	.002
	0.40	0.70	0.40	0.70	.089	.002	.045	.001	.043	.002	.102	.003
	0.40	0.70	0.70	0.40	.072	.003	.048	.002	.046	.002	.095	.003
	0.40	0.70	0.70	0.70	.088	.003	.047	.002	.044	.002	.101	.002
	0.70	0.40	0.40	0.40	.049	.002	.045	.003	.043	.003	.082	.003
	0.70	0.40	0.40	0.70	.070	.003	.048	.002	.046	.002	.097	.004
	0.70	0.40	0.70	0.40	.047	.002	.047	.002	.045	.002	.082	.002
	0.70	0.40	0.70	0.70	.074	.004	.049	.002	.046	.002	.106	.003
	0.70	0.70	0.40	0.40	.065	.003	.045	.002	.043	.002	.095	.004
	0.70	0.70	0.40	0.70	.084	.003	.043	.002	.040	.002	.103	.004
	0.70	0.70	0.70	0.40	.066	.002	.045	.002	.043	.002	.099	.003
	0.70	0.70	0.70	0.70	.087	.003	.046	.002	.045	.003	.112	.003
N _R =1800	0.00	0.00	0.00	0.00	.030	.001	.045	.002	.043	.002	.059	.002
	0.40	0.40	0.40	0.40	.099	.003	.047	.001	.045	.002	.100	.003
	0.40	0.40	0.40	0.70	.186	.011	.046	.002	.046	.002	.139	.005
	0.40	0.40	0.70	0.40	.098	.003	.047	.001	.047	.002	.118	.004
	0.40	0.40	0.70	0.70	.194	.010	.047	.002	.046	.002	.151	.005
	0.40	0.70	0.40	0.40	.195	.013	.047	.001	.046	.001	.139	.006
	0.40	0.70	0.40	0.70	.266	.009	.044	.002	.043	.002	.148	.005
	0.40	0.70	0.70	0.40	.195	.014	.049	.002	.049	.002	.146	.005
	0.40	0.70	0.70	0.70	.259	.008	.046	.001	.045	.001	.161	.003
	0.70	0.40	0.40	0.40	.099	.003	.052	.001	.052	.001	.114	.003
	0.70	0.40	0.40	0.70	.195	.010	.048	.002	.048	.002	.147	.004
	0.70	0.40	0.70	0.40	.096	.004	.048	.002	.048	.002	.129	.005
	0.70	0.40	0.70	0.70	.189	.011	.049	.001	.048	.001	.153	.004
	0.70	0.70	0.40	0.40	.192	.010	.048	.001	.047	.001	.148	.006
	0.70	0.70	0.40	0.70	.279	.012	.048	.002	.047	.002	.160	.004
	0.70	0.70	0.70	0.40	.200	.009	.047	.002	.047	.002	.151	.005
	0.70	0.70	0.70	0.70	.257	.010	.049	.001	.048	.001	.163	.005



Şekil 22.1. Üç Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TBO(ab) DMF Miktarı Etkileşimi İçin Tip I Hata Oranları

Tablo 22.2

Dört Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip I Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1	Odak2	Odak3	GLK		GLR-LRT		GLR-Wald		Lordif	
	DMF	DMF	DMF	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata
	Miktarı a ve b	Miktarı a ve b	Miktarı a ve b								
N _R =600	0.00;0.00	0.00;0.00	0.00;0.00	.028	.001	.043	.001	.042	.001	.061	.001
	0.40;0.40	0.40;0.40	0.40;0.40	.046	.001	.045	.001	.042	.001	.082	.001
	0.40;0.40	0.40;0.40	0.40;0.70	.061	.002	.044	.001	.042	.001	.096	.002
	0.40;0.40	0.40;0.40	0.70;0.40	.044	.001	.044	.001	.041	.001	.084	.001
	0.40;0.40	0.40;0.40	0.70;0.70	.060	.002	.045	.001	.042	.001	.102	.002
	0.40;0.40	0.40;0.70	0.40;0.40	.059	.001	.043	.001	.041	.001	.093	.002
	0.40;0.40	0.40;0.70	0.40;0.70	.073	.002	.044	.001	.042	.001	.103	.002
	0.40;0.40	0.40;0.70	0.70;0.40	.059	.001	.044	.001	.041	.001	.098	.001
	0.40;0.40	0.40;0.70	0.70;0.70	.072	.001	.045	.001	.042	.001	.112	.002
	0.40;0.40	0.70;0.40	0.40;0.40	.045	.001	.045	.001	.042	.001	.085	.002
	0.40;0.40	0.70;0.40	0.40;0.70	.058	.001	.044	.001	.041	.001	.096	.002
	0.40;0.40	0.70;0.40	0.70;0.40	.044	.001	.046	.001	.043	.001	.087	.002
	0.40;0.40	0.70;0.40	0.70;0.70	.059	.001	.046	.001	.043	.001	.103	.002
	0.40;0.40	0.70;0.70	0.40;0.40	.058	.001	.045	.001	.042	.001	.101	.002
	0.40;0.40	0.70;0.70	0.40;0.70	.070	.001	.045	.001	.042	.001	.110	.001
	0.40;0.40	0.70;0.70	0.70;0.40	.056	.001	.046	.001	.043	.001	.103	.002
	0.40;0.40	0.70;0.70	0.70;0.70	.072	.001	.044	.001	.041	.001	.115	.002
	0.40;0.70	0.40;0.40	0.40;0.40	.060	.001	.045	.001	.042	.001	.096	.002
	0.40;0.70	0.40;0.40	0.40;0.70	.070	.001	.044	.001	.041	.001	.104	.001
	0.40;0.70	0.40;0.40	0.70;0.40	.058	.002	.045	.001	.042	.001	.098	.002
	0.40;0.70	0.40;0.40	0.70;0.70	.071	.001	.044	.001	.041	.001	.109	.002
	0.40;0.70	0.40;0.70	0.40;0.40	.072	.001	.045	.001	.041	.001	.107	.002
	0.40;0.70	0.40;0.70	0.40;0.70	.082	.001	.044	.001	.041	.001	.112	.002
	0.40;0.70	0.40;0.70	0.70;0.40	.071	.001	.045	.001	.043	.001	.109	.002
	0.40;0.70	0.40;0.70	0.70;0.70	.079	.002	.045	.001	.042	.001	.117	.002
	0.40;0.70	0.70;0.40	0.40;0.40	.058	.001	.043	.001	.041	.001	.097	.001
	0.40;0.70	0.70;0.40	0.40;0.70	.071	.001	.044	.001	.041	.001	.108	.002
	0.40;0.70	0.70;0.40	0.70;0.40	.059	.001	.045	.001	.042	.001	.102	.001
	0.40;0.70	0.70;0.40	0.70;0.70	.072	.001	.045	.001	.043	.001	.112	.002
	0.40;0.70	0.70;0.70	0.40;0.40	.071	.001	.045	.001	.042	.001	.112	.002
	0.40;0.70	0.70;0.70	0.40;0.70	.081	.002	.045	.001	.042	.001	.116	.002
	0.40;0.70	0.70;0.70	0.70;0.40	.071	.002	.047	.001	.044	.001	.114	.002
	0.40;0.70	0.70;0.70	0.70;0.70	.079	.001	.045	.001	.042	.001	.121	.002
	0.70;0.40	0.40;0.40	0.40;0.40	.045	.001	.046	.001	.043	.001	.086	.001
	0.70;0.40	0.40;0.40	0.40;0.70	.057	.001	.045	.001	.041	.001	.099	.001
	0.70;0.40	0.40;0.40	0.70;0.40	.045	.001	.043	.001	.041	.001	.087	.001
	0.70;0.40	0.40;0.40	0.70;0.70	.060	.001	.046	.001	.043	.001	.106	.002
	0.70;0.40	0.40;0.70	0.40;0.40	.057	.001	.044	.001	.042	.001	.098	.002
	0.70;0.40	0.40;0.70	0.40;0.70	.072	.001	.046	.001	.043	.001	.108	.002
	0.70;0.40	0.40;0.70	0.70;0.40	.059	.001	.046	.001	.043	.001	.100	.002
	0.70;0.40	0.40;0.70	0.70;0.70	.071	.002	.046	.001	.042	.001	.112	.002
	0.70;0.40	0.70;0.40	0.40;0.40	.042	.001	.045	.001	.042	.001	.087	.002
	0.70;0.40	0.70;0.40	0.40;0.70	.057	.001	.044	.001	.041	.001	.099	.002
	0.70;0.40	0.70;0.40	0.70;0.40	.044	.001	.046	.001	.043	.001	.090	.002
	0.70;0.40	0.70;0.40	0.70;0.70	.058	.002	.045	.001	.042	.001	.106	.001
	0.70;0.40	0.70;0.70	0.40;0.40	.059	.001	.045	.001	.041	.001	.102	.002
	0.70;0.40	0.70;0.70	0.40;0.70	.070	.001	.042	.001	.039	.001	.109	.002
	0.70;0.40	0.70;0.70	0.70;0.40	.058	.001	.044	.001	.041	.001	.105	.002
	0.70;0.40	0.70;0.70	0.70;0.70	.071	.002	.046	.001	.043	.001	.120	.002
	0.70;0.70	0.40;0.40	0.40;0.40	.060	.001	.045	.001	.042	.001	.101	.002
	0.70;0.70	0.40;0.40	0.40;0.70	.071	.001	.045	.001	.043	.001	.111	.002

(Devam ediyor)

Tablo 22.2 (Devam)

Dört Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip I Hata Oranları ve Standart Hatalar

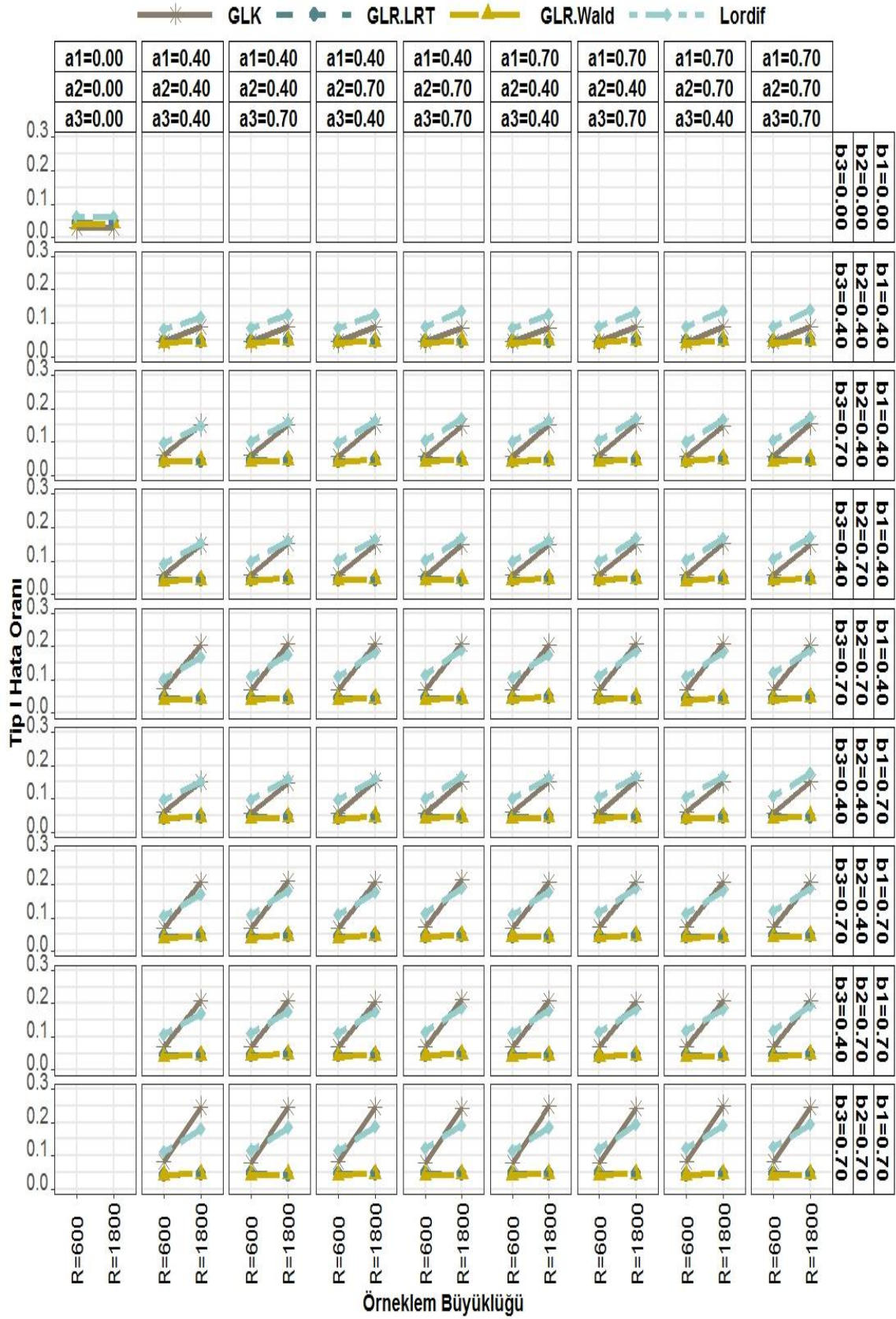
Referans Örneklem Büyüklük	Odak1	Odak2	Odak3	GLK		GLR-LRT		GLR-Wald		Lordif	
	DMF	DMF	DMF	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata	Tip I Hata	St. Hata
	Miktarı	Miktarı	Miktarı								
	a ve b	a ve b	a ve b								
	0.70;0.70	0.40;0.40	0.70;0.40	.057	.001	.046	.001	.043	.001	.105	.002
	0.70;0.70	0.40;0.40	0.70;0.70	.072	.002	.044	.001	.041	.001	.117	.002
	0.70;0.70	0.40;0.70	0.40;0.40	.070	.002	.043	.001	.040	.001	.110	.002
	0.70;0.70	0.40;0.70	0.40;0.70	.079	.002	.045	.001	.042	.001	.117	.001
	0.70;0.70	0.40;0.70	0.70;0.40	.071	.001	.043	.001	.040	.001	.113	.001
	0.70;0.70	0.40;0.70	0.70;0.70	.080	.001	.045	.001	.042	.001	.120	.002
	0.70;0.70	0.70;0.40	0.40;0.40	.060	.001	.045	.001	.042	.001	.104	.002
	0.70;0.70	0.70;0.40	0.40;0.70	.072	.002	.045	.001	.042	.001	.112	.002
	0.70;0.70	0.70;0.40	0.70;0.40	.059	.001	.046	.001	.043	.001	.108	.002
	0.70;0.70	0.70;0.40	0.70;0.70	.072	.002	.047	.001	.044	.001	.119	.002
	0.70;0.70	0.70;0.70	0.40;0.40	.070	.002	.044	.001	.042	.001	.117	.002
	0.70;0.70	0.70;0.70	0.40;0.70	.081	.002	.044	.001	.041	.001	.122	.002
	0.70;0.70	0.70;0.70	0.70;0.40	.071	.002	.044	.001	.041	.001	.117	.002
	0.70;0.70	0.70;0.70	0.70;0.70	.082	.002	.045	.001	.042	.001	.126	.002
N _R =1800	0.00;0.00	0.00;0.00	0.00;0.00	.028	.001	.043	.001	.042	.001	.061	.001
	0.40;0.40	0.40;0.40	0.40;0.40	.089	.001	.046	.001	.045	.001	.119	.002
	0.40;0.40	0.40;0.40	0.40;0.70	.152	.003	.044	.001	.044	.001	.149	.002
	0.40;0.40	0.40;0.40	0.70;0.40	.088	.002	.047	.001	.047	.001	.126	.002
	0.40;0.40	0.40;0.40	0.70;0.70	.153	.004	.045	.001	.044	.001	.159	.002
	0.40;0.40	0.40;0.70	0.40;0.40	.150	.003	.045	.001	.044	.001	.152	.002
	0.40;0.40	0.40;0.70	0.40;0.70	.207	.005	.044	.001	.043	.001	.168	.002
	0.40;0.40	0.40;0.70	0.70;0.40	.153	.004	.049	.001	.048	.001	.161	.003
	0.40;0.40	0.40;0.70	0.70;0.70	.208	.004	.045	.001	.044	.001	.177	.002
	0.40;0.40	0.70;0.40	0.40;0.40	.088	.002	.046	.001	.045	.001	.125	.002
	0.40;0.40	0.70;0.40	0.40;0.70	.153	.004	.048	.001	.047	.001	.162	.002
	0.40;0.40	0.70;0.40	0.70;0.40	.086	.001	.047	.001	.046	.001	.134	.002
	0.40;0.40	0.70;0.40	0.70;0.70	.150	.004	.046	.001	.046	.001	.168	.002
	0.40;0.40	0.70;0.70	0.40;0.40	.151	.004	.046	.001	.045	.001	.162	.003
	0.40;0.40	0.70;0.70	0.40;0.70	.209	.004	.047	.001	.045	.001	.183	.002
	0.40;0.40	0.70;0.70	0.70;0.40	.151	.004	.046	.001	.045	.001	.166	.002
	0.40;0.40	0.70;0.70	0.70;0.70	.208	.004	.047	.001	.046	.001	.189	.003
	0.40;0.70	0.40;0.40	0.40;0.40	.152	.005	.046	.001	.045	.001	.151	.002
	0.40;0.70	0.40;0.40	0.40;0.70	.207	.005	.046	.001	.045	.001	.170	.003
	0.40;0.70	0.40;0.40	0.70;0.40	.150	.004	.046	.001	.045	.001	.158	.002
	0.40;0.70	0.40;0.40	0.70;0.70	.210	.004	.047	.001	.046	.001	.183	.002
	0.40;0.70	0.40;0.70	0.40;0.40	.209	.004	.045	.001	.045	.001	.170	.002
	0.40;0.70	0.40;0.70	0.40;0.70	.246	.003	.045	.001	.045	.001	.182	.002
	0.40;0.70	0.40;0.70	0.70;0.40	.207	.004	.049	.001	.047	.001	.175	.002
	0.40;0.70	0.40;0.70	0.70;0.70	.245	.004	.044	.001	.044	.001	.186	.003
	0.40;0.70	0.70;0.40	0.40;0.40	.154	.004	.048	.001	.047	.001	.160	.002
	0.40;0.70	0.70;0.40	0.40;0.70	.208	.004	.047	.001	.047	.001	.177	.002
	0.40;0.70	0.70;0.40	0.70;0.40	.152	.004	.047	.001	.047	.001	.166	.002
	0.40;0.70	0.70;0.40	0.70;0.70	.213	.004	.048	.001	.047	.001	.188	.002
	0.40;0.70	0.70;0.70	0.40;0.40	.205	.004	.044	.001	.044	.001	.175	.003
	0.40;0.70	0.70;0.70	0.40;0.70	.245	.004	.046	.001	.046	.001	.187	.003
	0.40;0.70	0.70;0.70	0.70;0.40	.210	.004	.049	.001	.048	.001	.185	.002
	0.40;0.70	0.70;0.70	0.70;0.70	.243	.003	.045	.001	.044	.001	.193	.002
	0.70;0.40	0.40;0.40	0.40;0.40	.086	.002	.047	.001	.046	.001	.127	.002
	0.70;0.40	0.40;0.40	0.40;0.70	.152	.004	.047	.001	.046	.001	.162	.002
	0.70;0.40	0.40;0.40	0.70;0.40	.088	.001	.048	.001	.047	.001	.132	.002

(Devam ediyor)

Tablo 22.2 (Devam)

Dört Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip I Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1	Odak2	Odak3	GLK		GLR-LRT		GLR-Wald		Lordif	
	DMF	DMF	DMF	Tip I	St.	Tip I	St.	Tip I	St.	Tip I	St.
	a ve b	a ve b	a ve b	Hata	Hata	Hata	Hata	Hata	Hata	Hata	Hata
	0.70;0.40	0.40;0.40	0.70;0.70	.155	.005	.048	.001	.047	.001	.171	.003
	0.70;0.40	0.40;0.70	0.40;0.40	.150	.004	.048	.001	.047	.001	.158	.003
	0.70;0.40	0.40;0.70	0.40;0.70	.207	.004	.048	.001	.047	.001	.174	.002
	0.70;0.40	0.40;0.70	0.70;0.40	.150	.004	.048	.001	.047	.001	.167	.002
	0.70;0.40	0.40;0.70	0.70;0.70	.209	.004	.046	.001	.046	.001	.186	.002
	0.70;0.40	0.70;0.40	0.40;0.40	.087	.001	.048	.001	.047	.001	.135	.002
	0.70;0.40	0.70;0.40	0.40;0.70	.149	.003	.051	.001	.050	.001	.166	.002
	0.70;0.40	0.70;0.40	0.70;0.40	.087	.002	.047	.001	.047	.001	.139	.002
	0.70;0.40	0.70;0.40	0.70;0.70	.155	.004	.048	.001	.047	.001	.173	.003
	0.70;0.40	0.70;0.70	0.40;0.40	.153	.004	.047	.001	.046	.001	.168	.003
	0.70;0.40	0.70;0.70	0.40;0.70	.207	.004	.046	.001	.045	.001	.183	.003
	0.70;0.40	0.70;0.70	0.70;0.40	.151	.004	.048	.001	.047	.001	.171	.002
	0.70;0.40	0.70;0.70	0.70;0.70	.206	.004	.047	.001	.047	.001	.190	.002
	0.70;0.70	0.40;0.40	0.40;0.40	.152	.004	.045	.001	.044	.001	.162	.003
	0.70;0.70	0.40;0.40	0.40;0.70	.207	.004	.045	.001	.044	.001	.178	.002
	0.70;0.70	0.40;0.40	0.70;0.40	.154	.004	.047	.001	.046	.001	.166	.002
	0.70;0.70	0.40;0.40	0.70;0.70	.208	.004	.047	.001	.046	.001	.188	.002
	0.70;0.70	0.40;0.70	0.40;0.40	.208	.005	.045	.001	.044	.001	.179	.003
	0.70;0.70	0.40;0.70	0.40;0.70	.250	.003	.046	.001	.045	.001	.186	.002
	0.70;0.70	0.40;0.70	0.70;0.40	.205	.004	.046	.001	.045	.001	.184	.003
	0.70;0.70	0.40;0.70	0.70;0.70	.243	.004	.046	.001	.046	.001	.196	.003
	0.70;0.70	0.70;0.40	0.40;0.40	.150	.004	.045	.001	.044	.001	.165	.002
	0.70;0.70	0.70;0.40	0.40;0.70	.205	.004	.045	.001	.044	.001	.183	.003
	0.70;0.70	0.70;0.40	0.70;0.40	.152	.004	.048	.001	.048	.001	.178	.003
	0.70;0.70	0.70;0.40	0.70;0.70	.206	.004	.046	.001	.045	.001	.188	.002
	0.70;0.70	0.70;0.70	0.40;0.40	.209	.004	.044	.001	.043	.001	.181	.002
	0.70;0.70	0.70;0.70	0.40;0.70	.251	.004	.047	.001	.046	.001	.191	.002
	0.70;0.70	0.70;0.70	0.70;0.40	.210	.004	.047	.001	.046	.001	.193	.002
	0.70;0.70	0.70;0.70	0.70;0.70	.245	.004	.045	.001	.044	.001	.195	.002



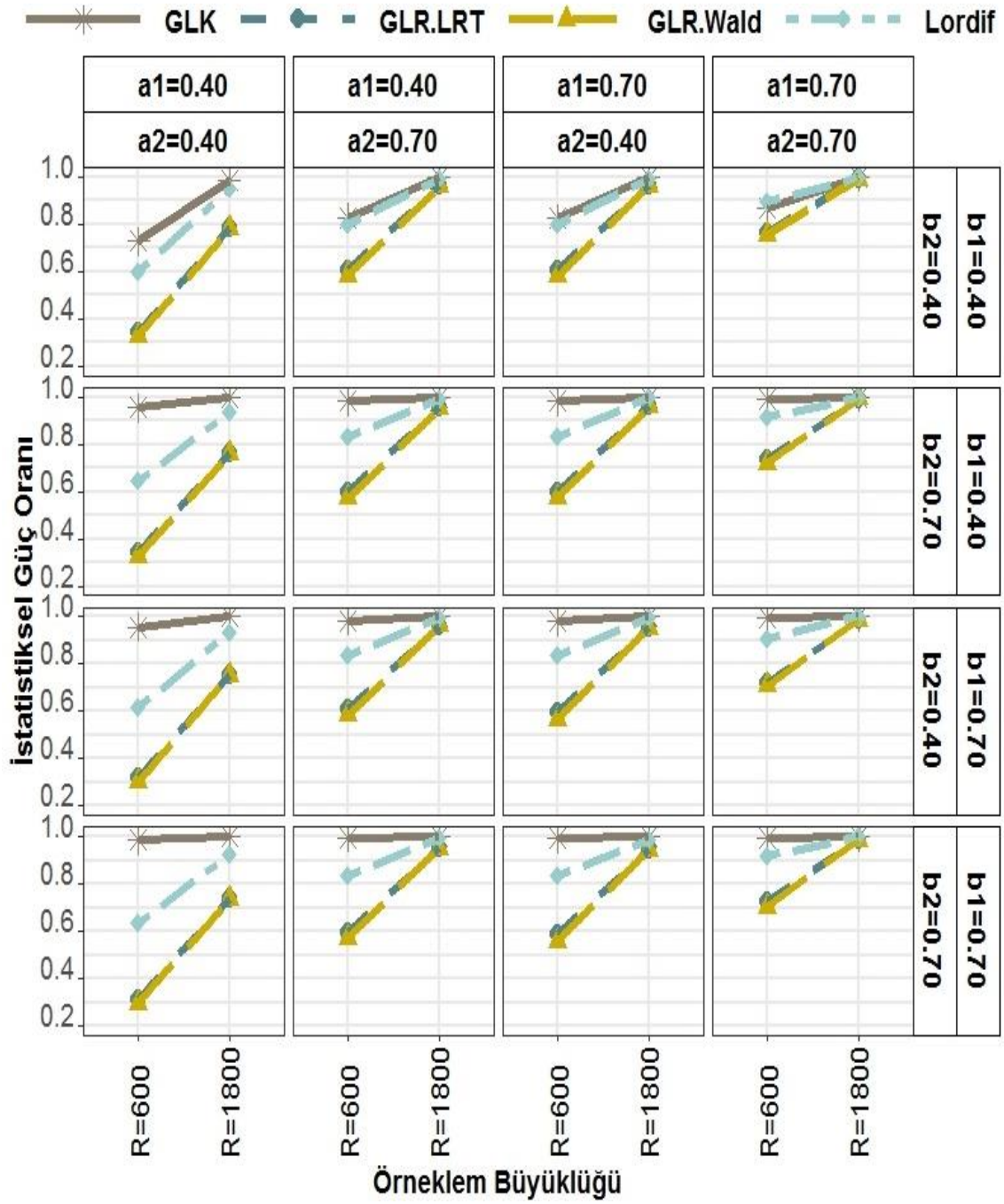
Şekil 22.2. Dört Grupta DMF Belirleme Teknikleri, Referans Grubun Örnekleme Büyüklüğü ve TBO(ab) DMF Miktarı Etkileşimi İçin Tip I Hata Oranları

EK 23. Üç ve Dört Grupta TBO(ab) DMF Türü İçin İstatistiksel Güç Oranları

Tablo 23.1

Üç Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama İstatistiksel Güç Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1 DMF Miktarı		Odak2 DMF Miktarı		GLK		GLR-LRT		GLR-Wald		Lordif	
					Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata
	a	b	a	b								
N _R =600	0.40	0.40	0.40	0.40	.728	.008	.346	.017	.329	.018	.601	.014
	0.40	0.40	0.40	0.70	.955	.005	.341	.014	.327	.016	.646	.011
	0.40	0.40	0.70	0.40	.827	.011	.604	.024	.586	.025	.796	.019
	0.40	0.40	0.70	0.70	.981	.004	.596	.022	.572	.025	.832	.015
	0.40	0.70	0.40	0.40	.955	.007	.317	.015	.302	.016	.618	.014
	0.40	0.70	0.40	0.70	.985	.003	.314	.016	.297	.018	.633	.014
	0.40	0.70	0.70	0.40	.984	.003	.610	.018	.585	.023	.838	.016
	0.40	0.70	0.70	0.70	.995	.001	.594	.025	.571	.029	.838	.016
	0.70	0.40	0.40	0.40	.825	.010	.602	.024	.580	.027	.796	.019
	0.70	0.40	0.40	0.70	.983	.003	.597	.031	.575	.035	.834	.018
	0.70	0.40	0.70	0.40	.868	.008	.765	.020	.754	.022	.894	.015
	0.70	0.40	0.70	0.70	.992	.002	.735	.011	.719	.014	.915	.005
	0.70	0.70	0.40	0.40	.981	.006	.598	.021	.569	.024	.837	.016
	0.70	0.70	0.40	0.70	.993	.002	.587	.021	.559	.024	.836	.015
	0.70	0.70	0.70	0.40	.995	.001	.723	.012	.709	.013	.908	.008
	0.70	0.70	0.70	0.70	.998	.001	.724	.012	.703	.015	.916	.006
N _R =1800	0.40	0.40	0.40	0.40	.986	.003	.784	.013	.780	.014	.947	.004
	0.40	0.40	0.40	0.70	1.000	.000	.759	.018	.757	.018	.934	.009
	0.40	0.40	0.70	0.40	.999	.001	.961	.007	.960	.008	.993	.002
	0.40	0.40	0.70	0.70	1.000	.000	.953	.008	.951	.008	.993	.002
	0.40	0.70	0.40	0.40	1.000	.000	.754	.018	.751	.019	.932	.009
	0.40	0.70	0.40	0.70	1.000	.000	.738	.011	.735	.012	.926	.005
	0.40	0.70	0.70	0.40	1.000	.000	.964	.006	.963	.006	.994	.001
	0.40	0.70	0.70	0.70	1.000	.000	.954	.007	.952	.008	.994	.002
	0.70	0.40	0.40	0.40	.999	.001	.963	.007	.962	.007	.994	.002
	0.70	0.40	0.40	0.70	1.000	.000	.960	.007	.958	.008	.994	.002
	0.70	0.40	0.70	0.40	1.000	.000	.992	.002	.992	.003	.999	.000
	0.70	0.40	0.70	0.70	1.000	.000	.992	.003	.992	.003	1.000	.000
	0.70	0.70	0.40	0.40	1.000	.000	.955	.008	.954	.008	.995	.001
	0.70	0.70	0.40	0.70	1.000	.000	.949	.009	.947	.010	.990	.002
	0.70	0.70	0.70	0.40	1.000	.000	.990	.003	.990	.003	1.000	.000
	0.70	0.70	0.70	0.70	1.000	.000	.989	.002	.988	.003	1.000	.000



Şekil 23.1. Üç Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TBO(ab) DMF Miktarı Etkileşimi İçin İstatistiksel Güç Oranları

Tablo 23.2

Dört Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama İstatistiksel Güç Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1	Odak2	Odak3	GLK		GLR-LRT		GLR-Wald		Lordif	
	DMF	DMF	DMF	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata
	Miktarı	Miktarı	Miktarı								
	a ve b	a ve b	a ve b								
N _R =600	0.40;0.40	0.40;0.40	0.40;0.40	.697	.003	.350	.005	.335	.006	.641	.006
	0.40;0.40	0.40;0.40	0.40;0.70	.932	.004	.333	.005	.320	.007	.695	.005
	0.40;0.40	0.40;0.40	0.70;0.40	.794	.006	.565	.011	.545	.013	.806	.008
	0.40;0.40	0.40;0.40	0.70;0.70	.967	.003	.534	.010	.513	.011	.842	.007
	0.40;0.40	0.40;0.70	0.40;0.40	.934	.005	.338	.005	.326	.005	.693	.005
	0.40;0.40	0.40;0.70	0.40;0.70	.978	.002	.333	.006	.319	.007	.688	.006
	0.40;0.40	0.40;0.70	0.70;0.40	.966	.003	.570	.012	.550	.013	.856	.007
	0.40;0.40	0.40;0.70	0.70;0.70	.991	.001	.536	.011	.511	.012	.835	.007
	0.40;0.40	0.70;0.40	0.40;0.40	.794	.005	.556	.011	.534	.013	.799	.008
	0.40;0.40	0.70;0.40	0.40;0.70	.966	.004	.557	.011	.535	.012	.850	.007
	0.40;0.40	0.70;0.40	0.70;0.40	.850	.004	.695	.009	.683	.010	.879	.005
	0.40;0.40	0.70;0.40	0.70;0.70	.982	.002	.681	.009	.665	.010	.911	.004
	0.40;0.40	0.70;0.70	0.40;0.40	.968	.003	.526	.010	.505	.010	.840	.006
	0.40;0.40	0.70;0.70	0.40;0.70	.989	.001	.536	.010	.513	.012	.836	.006
	0.40;0.40	0.70;0.70	0.70;0.40	.981	.002	.693	.010	.675	.012	.914	.005
	0.40;0.40	0.70;0.70	0.70;0.70	.995	.001	.660	.009	.641	.010	.905	.004
	0.40;0.70	0.40;0.40	0.40;0.40	.938	.005	.339	.005	.329	.006	.698	.006
	0.40;0.70	0.40;0.40	0.40;0.70	.976	.002	.325	.005	.314	.006	.685	.006
	0.40;0.70	0.40;0.40	0.70;0.40	.966	.003	.562	.009	.541	.011	.852	.006
	0.40;0.70	0.40;0.40	0.70;0.70	.990	.001	.539	.011	.515	.012	.838	.007
	0.40;0.70	0.40;0.70	0.40;0.40	.976	.002	.335	.005	.321	.006	.692	.004
	0.40;0.70	0.40;0.70	0.40;0.70	.982	.001	.320	.004	.303	.005	.686	.006
	0.40;0.70	0.40;0.70	0.70;0.40	.990	.001	.556	.013	.533	.014	.849	.008
	0.40;0.70	0.40;0.70	0.70;0.70	.992	.001	.525	.011	.504	.012	.834	.008
	0.40;0.70	0.70;0.40	0.40;0.40	.967	.003	.567	.010	.544	.012	.851	.007
	0.40;0.70	0.70;0.40	0.40;0.70	.993	.001	.555	.012	.533	.013	.854	.007
	0.40;0.70	0.70;0.40	0.70;0.40	.981	.002	.698	.008	.685	.009	.917	.004
	0.40;0.70	0.70;0.40	0.70;0.70	.996	.001	.685	.009	.668	.010	.911	.004
	0.40;0.70	0.70;0.70	0.40;0.40	.990	.001	.529	.010	.506	.011	.830	.007
	0.40;0.70	0.70;0.70	0.40;0.70	.992	.001	.527	.012	.502	.013	.830	.007
	0.40;0.70	0.70;0.70	0.70;0.40	.995	.001	.681	.009	.661	.011	.907	.005
	0.40;0.70	0.70;0.70	0.70;0.70	.997	.001	.667	.009	.646	.010	.908	.004
	0.70;0.40	0.40;0.40	0.40;0.40	.789	.007	.566	.012	.546	.013	.804	.010
	0.70;0.40	0.40;0.40	0.40;0.70	.969	.003	.569	.013	.546	.015	.857	.008
	0.70;0.40	0.40;0.40	0.70;0.40	.854	.005	.699	.009	.684	.010	.882	.006
	0.70;0.40	0.40;0.40	0.70;0.70	.982	.002	.687	.008	.672	.009	.911	.004
	0.70;0.40	0.40;0.70	0.40;0.40	.964	.004	.559	.012	.537	.013	.851	.007
	0.70;0.40	0.40;0.70	0.40;0.70	.992	.001	.554	.010	.531	.011	.843	.006
	0.70;0.40	0.40;0.70	0.70;0.40	.980	.003	.697	.010	.679	.011	.917	.005
	0.70;0.40	0.40;0.70	0.70;0.70	.995	.001	.684	.009	.666	.011	.912	.004
	0.70;0.40	0.70;0.40	0.40;0.40	.846	.004	.697	.009	.682	.011	.881	.006
	0.70;0.40	0.70;0.40	0.40;0.70	.979	.002	.698	.009	.683	.010	.914	.004
	0.70;0.40	0.70;0.40	0.70;0.40	.873	.004	.771	.006	.763	.007	.914	.004
	0.70;0.40	0.70;0.40	0.70;0.70	.987	.002	.760	.005	.751	.006	.942	.003
	0.70;0.40	0.70;0.70	0.40;0.40	.982	.002	.684	.008	.666	.009	.913	.005
	0.70;0.40	0.70;0.70	0.40;0.70	.995	.001	.681	.009	.662	.011	.913	.005
	0.70;0.40	0.70;0.70	0.70;0.40	.986	.002	.763	.007	.754	.008	.948	.003
	0.70;0.40	0.70;0.70	0.70;0.70	.998	.000	.751	.007	.740	.008	.941	.003
	0.70;0.70	0.40;0.40	0.40;0.40	.967	.004	.527	.010	.508	.011	.837	.007
	0.70;0.70	0.40;0.40	0.40;0.70	.990	.001	.537	.010	.514	.011	.843	.006
	0.70;0.70	0.40;0.40	0.70;0.40	.980	.002	.681	.008	.665	.010	.910	.005

(Devam ediyor)

Tablo 23.2 (Devam)

Dört Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama İstatistiksel Güç Oranları ve Standart Hatalar

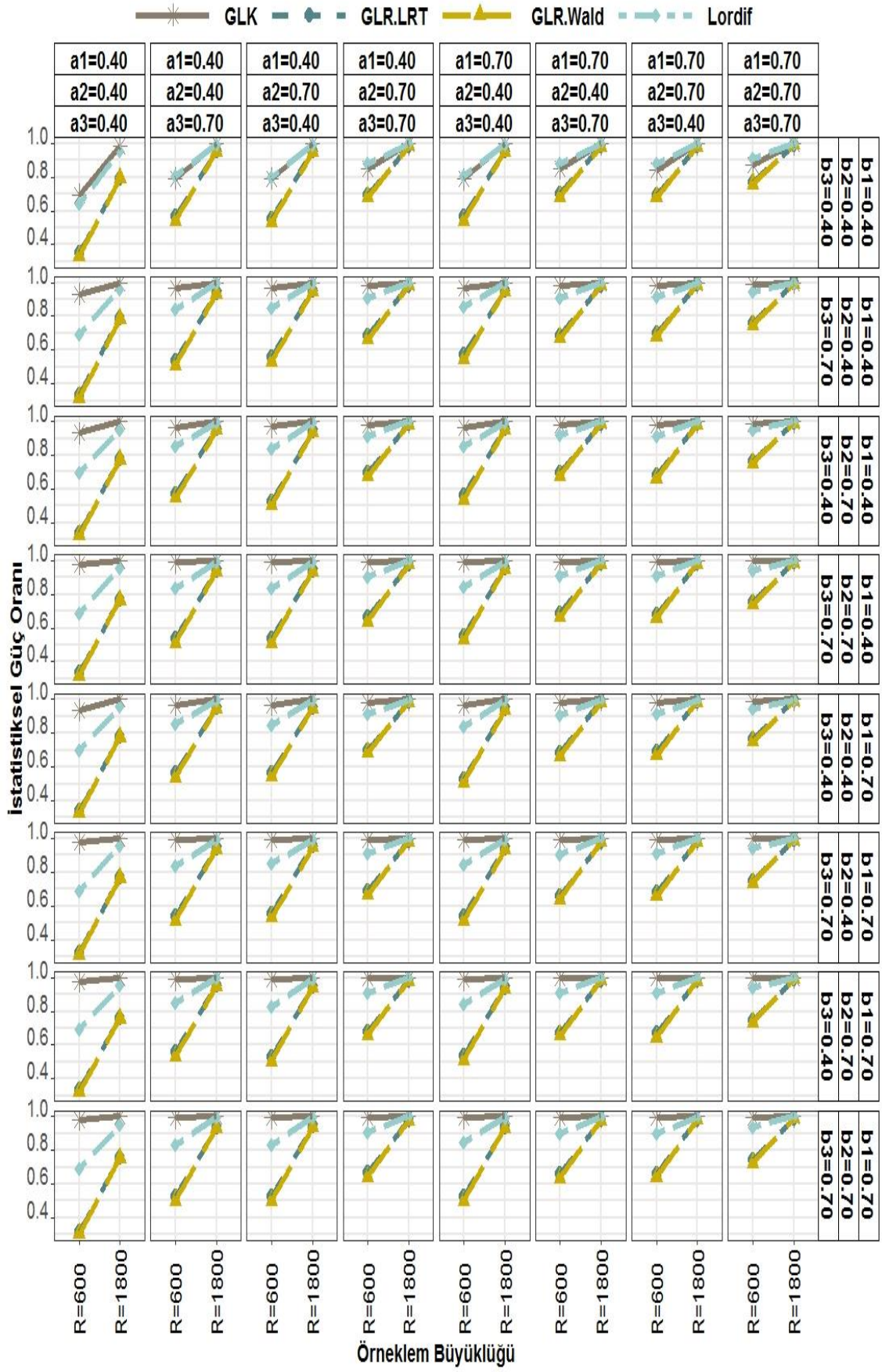
Referans Örneklem Büyüklük	Odak1	Odak2	Odak3	GLK		GLR-LRT		GLR-Wald		Lordif	
	DMF	DMF	DMF	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata
	Miktarı	Miktarı	Miktarı								
	a ve b	a ve b	a ve b								
	0.70;0.70	0.40;0.40	0.70;0.70	.994	.001	.660	.009	.642	.010	.906	.005
	0.70;0.70	0.40;0.70	0.40;0.40	.990	.001	.536	.011	.513	.012	.838	.008
	0.70;0.70	0.40;0.70	0.40;0.70	.992	.001	.528	.010	.504	.011	.844	.007
	0.70;0.70	0.40;0.70	0.70;0.40	.995	.001	.673	.008	.657	.010	.911	.004
	0.70;0.70	0.40;0.70	0.70;0.70	.994	.001	.657	.009	.638	.011	.897	.005
	0.70;0.70	0.70;0.40	0.40;0.40	.981	.002	.688	.008	.670	.009	.912	.004
	0.70;0.70	0.70;0.40	0.40;0.70	.995	.001	.678	.009	.661	.010	.913	.004
	0.70;0.70	0.70;0.40	0.70;0.40	.988	.002	.763	.006	.753	.007	.946	.003
	0.70;0.70	0.70;0.40	0.70;0.70	.998	.000	.748	.007	.739	.008	.939	.003
	0.70;0.70	0.70;0.70	0.40;0.40	.994	.001	.667	.008	.644	.010	.905	.004
	0.70;0.70	0.70;0.70	0.40;0.70	.995	.001	.663	.009	.642	.011	.901	.005
	0.70;0.70	0.70;0.70	0.70;0.40	.997	.001	.746	.006	.735	.008	.939	.003
	0.70;0.70	0.70;0.70	0.70;0.70	.998	.000	.739	.006	.726	.007	.938	.003
N _R =1800	0.40;0.40	0.40;0.40	0.40;0.40	.987	.001	.794	.005	.793	.005	.959	.002
	0.40;0.40	0.40;0.40	0.40;0.70	1.000	.000	.786	.006	.783	.007	.957	.003
	0.40;0.40	0.40;0.40	0.70;0.40	.998	.000	.956	.004	.954	.004	.994	.001
	0.40;0.40	0.40;0.40	0.70;0.70	1.000	.000	.941	.004	.938	.005	.994	.001
	0.40;0.40	0.40;0.70	0.40;0.40	1.000	.000	.775	.007	.774	.007	.951	.003
	0.40;0.40	0.40;0.70	0.40;0.70	1.000	.000	.767	.006	.765	.006	.952	.003
	0.40;0.40	0.40;0.70	0.70;0.40	1.000	.000	.955	.004	.952	.005	.995	.001
	0.40;0.40	0.40;0.70	0.70;0.70	1.000	.000	.940	.006	.938	.006	.993	.001
	0.40;0.40	0.70;0.40	0.40;0.40	.999	.000	.953	.004	.951	.005	.996	.001
	0.40;0.40	0.70;0.40	0.40;0.70	1.000	.000	.954	.004	.953	.005	.994	.001
	0.40;0.40	0.70;0.40	0.70;0.40	.999	.000	.988	.001	.988	.001	.999	.000
	0.40;0.40	0.70;0.40	0.70;0.70	1.000	.000	.985	.002	.984	.002	.998	.001
	0.40;0.40	0.70;0.70	0.40;0.40	1.000	.000	.942	.004	.940	.005	.993	.001
	0.40;0.40	0.70;0.70	0.40;0.70	1.000	.000	.939	.005	.937	.005	.993	.001
	0.40;0.40	0.70;0.70	0.70;0.40	1.000	.000	.986	.001	.986	.002	.999	.000
	0.40;0.40	0.70;0.70	0.70;0.70	1.000	.000	.979	.002	.979	.002	.998	.001
	0.40;0.70	0.40;0.40	0.40;0.40	1.000	.000	.781	.007	.779	.007	.958	.003
	0.40;0.70	0.40;0.40	0.40;0.70	1.000	.000	.767	.006	.766	.007	.955	.003
	0.40;0.70	0.40;0.40	0.70;0.40	1.000	.000	.953	.004	.951	.005	.995	.001
	0.40;0.70	0.40;0.40	0.70;0.70	1.000	.000	.941	.004	.938	.005	.993	.001
	0.40;0.70	0.40;0.70	0.40;0.40	1.000	.000	.761	.006	.759	.007	.954	.003
	0.40;0.70	0.40;0.70	0.40;0.70	1.000	.000	.755	.006	.753	.007	.948	.003
	0.40;0.70	0.40;0.70	0.70;0.40	1.000	.000	.952	.004	.950	.005	.995	.001
	0.40;0.70	0.40;0.70	0.70;0.70	1.000	.000	.936	.004	.933	.005	.993	.001
	0.40;0.70	0.70;0.40	0.40;0.40	1.000	.000	.952	.005	.949	.005	.995	.001
	0.40;0.70	0.70;0.40	0.40;0.70	1.000	.000	.953	.004	.951	.005	.994	.001
	0.40;0.70	0.70;0.40	0.70;0.40	1.000	.000	.988	.001	.988	.002	1.000	.000
	0.40;0.70	0.70;0.40	0.70;0.70	1.000	.000	.984	.002	.983	.002	.999	.000
	0.40;0.70	0.70;0.70	0.40;0.40	1.000	.000	.946	.004	.942	.004	.993	.001
	0.40;0.70	0.70;0.70	0.40;0.70	1.000	.000	.939	.004	.937	.005	.993	.001
	0.40;0.70	0.70;0.70	0.70;0.40	1.000	.000	.983	.002	.982	.002	.999	.000
	0.40;0.70	0.70;0.70	0.70;0.70	1.000	.000	.980	.003	.979	.003	.999	.000
	0.70;0.40	0.40;0.40	0.40;0.40	.998	.000	.955	.004	.953	.004	.995	.001
	0.70;0.40	0.40;0.40	0.40;0.70	1.000	.000	.953	.004	.950	.004	.996	.001
	0.70;0.40	0.40;0.40	0.70;0.40	1.000	.000	.986	.001	.985	.002	.999	.000
	0.70;0.40	0.40;0.40	0.70;0.70	1.000	.000	.985	.002	.985	.002	.999	.000
	0.70;0.40	0.40;0.70	0.40;0.40	1.000	.000	.954	.004	.952	.005	.994	.001
	0.70;0.40	0.40;0.70	0.40;0.70	1.000	.000	.955	.005	.952	.005	.995	.001

(Devam ediyor)

Tablo 23.2 (Devam)

Dört Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama İstatistiksel Güç Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1	Odak2	Odak3	GLK		GLR-LRT		GLR-Wald		Lordif	
	DMF	DMF	DMF	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata	Güç	St. Hata
	Miktarı	Miktarı	Miktarı								
	a ve b	a ve b	a ve b								
	0.70;0.40	0.40;0.70	0.70;0.40	1.000	.000	.989	.001	.989	.001	1.000	.000
	0.70;0.40	0.40;0.70	0.70;0.70	1.000	.000	.985	.002	.984	.002	.999	.000
	0.70;0.40	0.70;0.40	0.40;0.40	.999	.000	.987	.002	.986	.002	.999	.000
	0.70;0.40	0.70;0.40	0.40;0.70	1.000	.000	.987	.002	.986	.002	.999	.000
	0.70;0.40	0.70;0.40	0.70;0.40	1.000	.000	.995	.001	.995	.001	1.000	.000
	0.70;0.40	0.70;0.40	0.70;0.70	1.000	.000	.996	.001	.996	.001	1.000	.000
	0.70;0.40	0.70;0.70	0.40;0.40	1.000	.000	.984	.001	.983	.002	.999	.000
	0.70;0.40	0.70;0.70	0.40;0.70	1.000	.000	.985	.002	.985	.002	.999	.000
	0.70;0.40	0.70;0.70	0.70;0.40	1.000	.000	.995	.001	.995	.001	1.000	.000
	0.70;0.40	0.70;0.70	0.70;0.70	1.000	.000	.993	.001	.993	.001	.999	.000
	0.70;0.70	0.40;0.40	0.40;0.40	1.000	.000	.941	.005	.939	.005	.994	.001
	0.70;0.70	0.40;0.40	0.40;0.70	1.000	.000	.940	.005	.937	.005	.993	.001
	0.70;0.70	0.40;0.40	0.70;0.40	1.000	.000	.985	.002	.985	.002	.999	.000
	0.70;0.70	0.40;0.40	0.70;0.70	1.000	.000	.981	.002	.980	.002	.999	.000
	0.70;0.70	0.40;0.70	0.40;0.40	1.000	.000	.937	.004	.935	.005	.993	.001
	0.70;0.70	0.40;0.70	0.40;0.70	1.000	.000	.938	.005	.934	.005	.991	.001
	0.70;0.70	0.40;0.70	0.70;0.40	1.000	.000	.985	.002	.985	.002	.999	.000
	0.70;0.70	0.40;0.70	0.70;0.70	1.000	.000	.981	.002	.980	.002	.999	.000
	0.70;0.70	0.70;0.40	0.40;0.40	1.000	.000	.986	.002	.985	.002	.999	.000
	0.70;0.70	0.70;0.40	0.40;0.70	1.000	.000	.982	.002	.982	.002	.998	.000
	0.70;0.70	0.70;0.40	0.70;0.40	1.000	.000	.995	.001	.995	.001	1.000	.000
	0.70;0.70	0.70;0.40	0.70;0.70	1.000	.000	.993	.001	.994	.001	1.000	.000
	0.70;0.70	0.70;0.70	0.40;0.40	1.000	.000	.982	.002	.982	.002	.999	.000
	0.70;0.70	0.70;0.70	0.40;0.70	1.000	.000	.983	.002	.982	.002	.998	.001
	0.70;0.70	0.70;0.70	0.70;0.40	1.000	.000	.993	.001	.993	.001	1.000	.000
	0.70;0.70	0.70;0.70	0.70;0.70	1.000	.000	.991	.001	.991	.001	1.000	.000



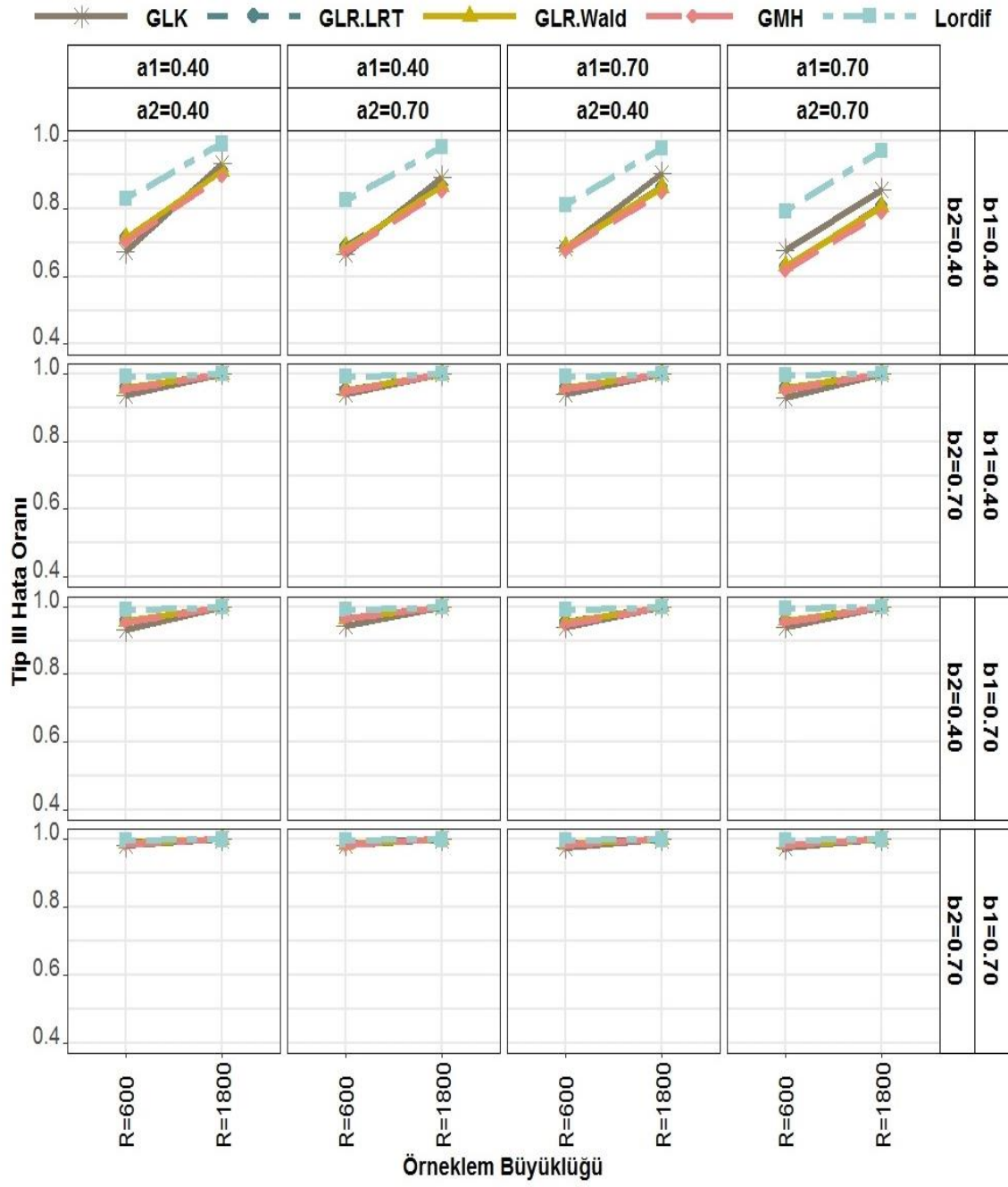
Şekil 23.2. Dört Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TBO(ab) DMF Miktarı Etkileşimi İçin İstatistiksel Güç Oranları

EK 24. Üç ve Dört Grupta TBO(ab) DMF Türü İçin Tip III Hata Oranları

Tablo 24.1

Üç Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip III Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1		Odak2		GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
	DMF Miktarı		DMF Miktarı		Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata
	a	b	a	b										
N _R =600	0.40	0.40	0.40	0.40	.673	.018	.713	.007	.712	.007	.704	.008	.831	.009
	0.40	0.40	0.40	0.70	.936	.011	.957	.007	.957	.007	.955	.006	.991	.002
	0.40	0.40	0.70	0.40	.665	.012	.686	.007	.685	.007	.672	.007	.824	.011
	0.40	0.40	0.70	0.70	.941	.011	.951	.007	.951	.007	.947	.007	.993	.002
	0.40	0.70	0.40	0.40	.931	.013	.958	.009	.958	.009	.956	.008	.990	.003
	0.40	0.70	0.40	0.70	.982	.003	.990	.001	.990	.001	.988	.002	.999	.001
	0.40	0.70	0.70	0.40	.945	.012	.966	.006	.966	.006	.964	.006	.992	.003
	0.40	0.70	0.70	0.70	.984	.002	.986	.003	.986	.003	.982	.003	1.000	.000
	0.70	0.40	0.40	0.40	.685	.021	.686	.008	.686	.008	.675	.009	.812	.010
	0.70	0.40	0.40	0.70	.940	.011	.959	.008	.959	.008	.957	.008	.994	.002
	0.70	0.40	0.70	0.40	.678	.022	.630	.007	.630	.007	.619	.004	.792	.010
	0.70	0.40	0.70	0.70	.930	.011	.958	.007	.958	.007	.953	.007	.996	.002
	0.70	0.70	0.40	0.40	.941	.014	.954	.010	.954	.010	.946	.010	.992	.004
	0.70	0.70	0.40	0.70	.976	.003	.986	.002	.986	.002	.984	.003	.999	.001
	0.70	0.70	0.70	0.40	.941	.009	.960	.007	.960	.007	.955	.007	.994	.002
	0.70	0.70	0.70	0.70	.975	.004	.985	.002	.985	.002	.982	.003	1.000	.000
N _R =1800	0.40	0.40	0.40	0.40	.934	.009	.911	.005	.911	.005	.896	.005	.991	.002
	0.40	0.40	0.40	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40	0.40	0.70	0.40	.893	.009	.866	.007	.866	.007	.850	.007	.981	.001
	0.40	0.40	0.70	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40	0.70	0.40	0.40	.999	.001	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40	0.70	0.40	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40	0.70	0.70	0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40	0.70	0.70	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.40	0.40	0.40	.905	.012	.863	.009	.863	.009	.849	.009	.978	.003
	0.70	0.40	0.40	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.40	0.70	0.40	.856	.009	.806	.008	.806	.008	.789	.008	.969	.005
	0.70	0.40	0.70	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.70	0.40	0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.70	0.40	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.70	0.70	0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70	0.70	0.70	0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000



Şekil 24.1. Üç Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TBO(ab) DMF Miktarı Etkileşimi İçin Tip III Hata Oranları

Tablo 24.2

Dört Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip III Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1	Odak2	Odak3	GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
	DMF	DMF	DMF	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata	Tip III Hata	St. Hata
	a ve b	a ve b	a ve b										
N _R =600	0.40;0.40	0.40;0.40	0.40;0.40	.676	.005	.714	.004	.712	.004	.704	.004	.835	.004
	0.40;0.40	0.40;0.40	0.40;0.70	.922	.005	.950	.004	.950	.004	.946	.004	.990	.002
	0.40;0.40	0.40;0.40	0.70;0.40	.667	.005	.690	.003	.689	.003	.680	.003	.829	.005
	0.40;0.40	0.40;0.40	0.70;0.70	.916	.005	.940	.005	.940	.005	.934	.005	.990	.002
	0.40;0.40	0.40;0.70	0.40;0.40	.916	.007	.947	.005	.946	.005	.942	.005	.990	.001
	0.40;0.40	0.40;0.70	0.40;0.70	.968	.002	.984	.002	.984	.002	.983	.002	.999	.000
	0.40;0.40	0.40;0.70	0.70;0.40	.919	.007	.952	.005	.952	.005	.950	.005	.991	.002
	0.40;0.40	0.40;0.70	0.70;0.70	.966	.003	.980	.002	.980	.002	.978	.002	.999	.000
	0.40;0.40	0.70;0.40	0.40;0.40	.675	.005	.694	.004	.693	.004	.686	.004	.831	.004
	0.40;0.40	0.70;0.40	0.40;0.70	.915	.007	.950	.005	.949	.005	.946	.005	.990	.002
	0.40;0.40	0.70;0.40	0.70;0.40	.663	.004	.667	.003	.666	.003	.659	.003	.823	.004
	0.40;0.40	0.70;0.40	0.70;0.70	.910	.006	.940	.005	.940	.005	.936	.005	.991	.001
	0.40;0.40	0.70;0.70	0.40;0.40	.911	.007	.942	.005	.941	.005	.936	.005	.989	.002
	0.40;0.40	0.70;0.70	0.40;0.70	.964	.003	.982	.002	.982	.002	.979	.002	.999	.000
	0.40;0.40	0.70;0.70	0.70;0.40	.916	.007	.941	.005	.941	.005	.937	.005	.992	.001
	0.40;0.40	0.70;0.70	0.70;0.70	.966	.003	.981	.002	.980	.002	.977	.002	.999	.000
	0.40;0.70	0.40;0.40	0.40;0.40	.919	.007	.953	.004	.953	.004	.949	.005	.990	.002
	0.40;0.70	0.40;0.40	0.40;0.70	.968	.003	.986	.001	.986	.001	.983	.001	.999	.000
	0.40;0.70	0.40;0.40	0.70;0.40	.916	.006	.952	.005	.951	.005	.948	.005	.991	.001
	0.40;0.70	0.40;0.40	0.70;0.70	.968	.003	.984	.002	.983	.002	.981	.002	.999	.000
	0.40;0.70	0.40;0.70	0.40;0.40	.966	.003	.983	.002	.983	.002	.981	.002	.998	.000
	0.40;0.70	0.40;0.70	0.40;0.70	.980	.002	.992	.001	.991	.001	.990	.001	1.000	.000
	0.40;0.70	0.40;0.70	0.70;0.40	.966	.003	.987	.002	.987	.002	.986	.002	.999	.000
	0.40;0.70	0.40;0.70	0.70;0.70	.977	.002	.989	.001	.989	.001	.987	.001	1.000	.000
	0.40;0.70	0.70;0.40	0.40;0.40	.920	.006	.952	.005	.952	.005	.949	.005	.990	.002
	0.40;0.70	0.70;0.40	0.40;0.70	.974	.003	.989	.002	.989	.002	.989	.002	.999	.000
	0.40;0.70	0.70;0.40	0.70;0.40	.920	.007	.952	.005	.951	.005	.949	.005	.991	.001
	0.40;0.70	0.70;0.40	0.70;0.70	.970	.002	.986	.002	.986	.002	.985	.002	1.000	.000
	0.40;0.70	0.70;0.70	0.40;0.40	.967	.003	.984	.002	.984	.002	.982	.002	.999	.000
	0.40;0.70	0.70;0.70	0.40;0.70	.975	.002	.988	.001	.988	.001	.986	.001	1.000	.000
	0.40;0.70	0.70;0.70	0.70;0.40	.971	.003	.988	.001	.988	.002	.986	.002	.999	.000
	0.40;0.70	0.70;0.70	0.70;0.70	.973	.002	.986	.001	.986	.001	.985	.001	1.000	.000
	0.70;0.40	0.40;0.40	0.40;0.40	.669	.005	.691	.003	.690	.003	.682	.003	.831	.005
	0.70;0.40	0.40;0.40	0.40;0.70	.918	.006	.952	.005	.952	.005	.951	.005	.991	.002
	0.70;0.40	0.40;0.40	0.70;0.40	.665	.006	.667	.004	.665	.004	.655	.004	.822	.004
	0.70;0.40	0.40;0.40	0.70;0.70	.914	.006	.943	.005	.943	.005	.938	.006	.992	.001
	0.70;0.40	0.40;0.70	0.40;0.40	.924	.006	.950	.005	.950	.005	.948	.005	.989	.002
	0.70;0.40	0.40;0.70	0.40;0.70	.972	.003	.988	.002	.988	.002	.986	.002	.999	.000
	0.70;0.40	0.40;0.70	0.70;0.40	.920	.006	.952	.005	.951	.005	.949	.005	.992	.002
	0.70;0.40	0.40;0.70	0.70;0.70	.973	.002	.987	.002	.987	.002	.986	.002	1.000	.000
	0.70;0.40	0.70;0.40	0.40;0.40	.658	.003	.664	.004	.662	.004	.653	.004	.820	.004
	0.70;0.40	0.70;0.40	0.40;0.70	.920	.006	.951	.005	.950	.005	.947	.006	.990	.002
	0.70;0.40	0.70;0.40	0.70;0.40	.652	.005	.634	.003	.633	.003	.625	.003	.803	.004
	0.70;0.40	0.70;0.40	0.70;0.70	.918	.007	.940	.006	.940	.006	.936	.006	.991	.002
	0.70;0.40	0.70;0.70	0.40;0.40	.918	.006	.944	.005	.943	.005	.939	.005	.990	.001
	0.70;0.40	0.70;0.70	0.40;0.70	.968	.003	.987	.001	.987	.001	.985	.002	.999	.000
	0.70;0.40	0.70;0.70	0.70;0.40	.918	.006	.942	.006	.942	.006	.940	.006	.993	.001
	0.70;0.40	0.70;0.70	0.70;0.70	.967	.003	.983	.002	.983	.002	.981	.002	.999	.000
	0.70;0.70	0.40;0.40	0.40;0.40	.909	.007	.937	.005	.936	.005	.932	.005	.989	.002
	0.70;0.70	0.40;0.40	0.40;0.70	.967	.003	.982	.002	.982	.002	.981	.002	.999	.000
	0.70;0.70	0.40;0.40	0.70;0.40	.912	.007	.942	.004	.942	.004	.936	.005	.992	.002

(Devam ediyor)

Tablo 24.2 (Devam)

Dört Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip III Hata Oranları ve Standart Hatalar

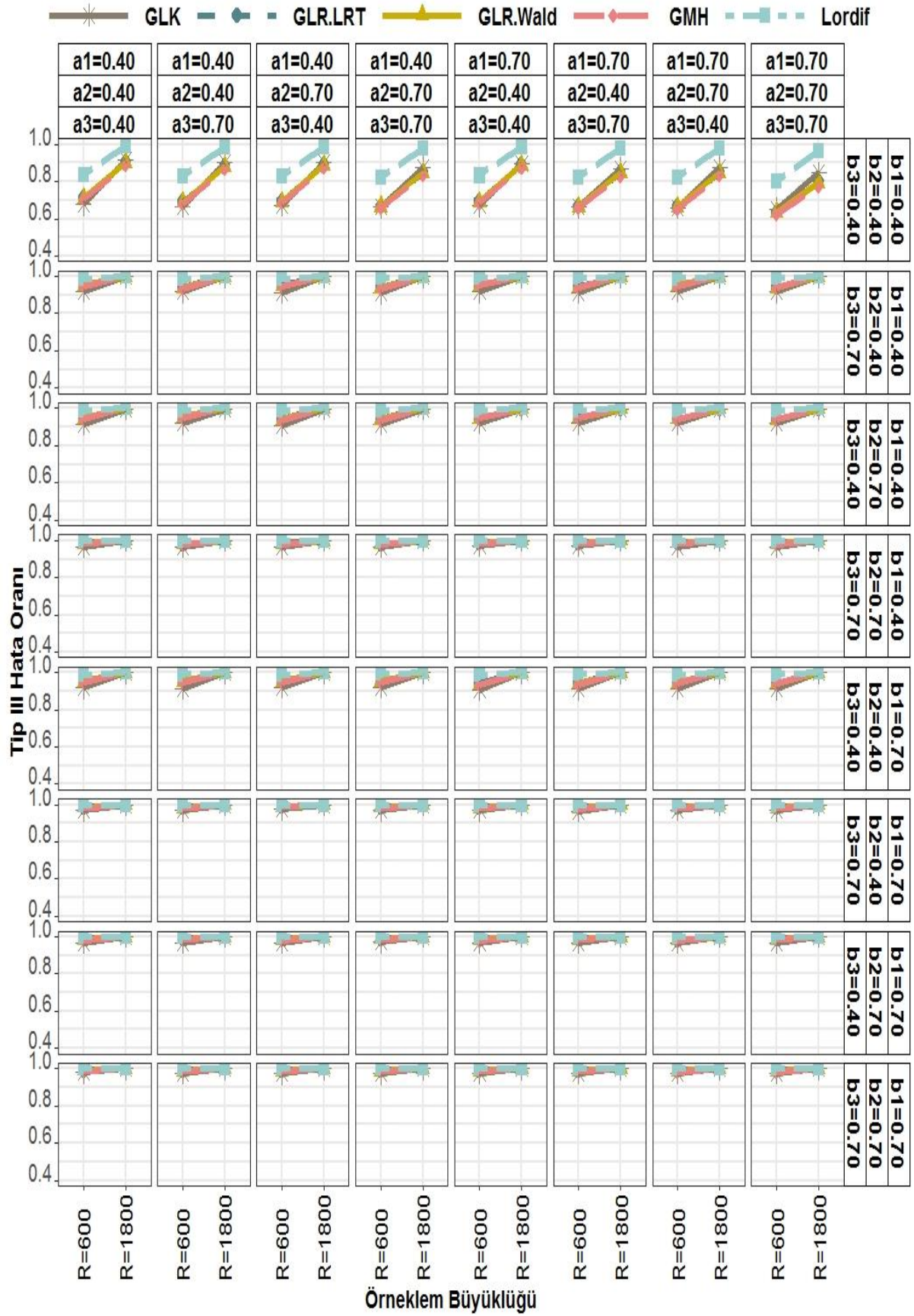
Referans Örneklem Büyüklük	Odak1	Odak2	Odak3	GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
	DMF	DMF	DMF	Tip	St.	Tip	St.	Tip	St.	Tip	St.	Tip	St.
	Miktarı	Miktarı	Miktarı	III	Hata	III	Hata	III	Hata	III	Hata	III	Hata
	a ve b	a ve b	a ve b	Hata		Hata		Hata		Hata		Hata	
	0.70;0.70	0.40;0.40	0.70;0.70	.965	.003	.980	.002	.980	.002	.978	.002	.999	.000
	0.70;0.70	0.40;0.70	0.40;0.40	.966	.002	.983	.001	.982	.001	.980	.002	.999	.000
	0.70;0.70	0.40;0.70	0.40;0.70	.977	.002	.989	.001	.989	.001	.987	.001	1.000	.000
	0.70;0.70	0.40;0.70	0.70;0.40	.968	.003	.987	.001	.986	.001	.983	.002	.999	.000
	0.70;0.70	0.40;0.70	0.70;0.70	.977	.002	.989	.001	.989	.001	.987	.001	1.000	.000
	0.70;0.70	0.70;0.40	0.40;0.40	.914	.006	.944	.005	.944	.005	.938	.005	.991	.002
	0.70;0.70	0.70;0.40	0.40;0.70	.972	.002	.985	.001	.985	.001	.983	.001	1.000	.000
	0.70;0.70	0.70;0.40	0.70;0.40	.910	.007	.942	.006	.942	.006	.939	.006	.991	.002
	0.70;0.70	0.70;0.40	0.70;0.70	.966	.003	.983	.001	.983	.001	.982	.002	.999	.000
	0.70;0.70	0.70;0.70	0.40;0.40	.966	.003	.982	.002	.982	.002	.979	.002	.999	.000
	0.70;0.70	0.70;0.70	0.40;0.70	.975	.002	.986	.001	.986	.001	.984	.001	1.000	.000
	0.70;0.70	0.70;0.70	0.70;0.40	.968	.002	.985	.002	.985	.002	.983	.002	.999	.000
	0.70;0.70	0.70;0.70	0.70;0.70	.975	.002	.986	.001	.986	.001	.984	.001	1.000	.000
N _R =1800	0.40;0.40	0.40;0.40	0.40;0.40	.913	.003	.903	.003	.903	.003	.891	.003	.990	.001
	0.40;0.40	0.40;0.40	0.40;0.70	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.40	0.40;0.40	0.70;0.40	.900	.003	.883	.003	.883	.003	.870	.003	.988	.001
	0.40;0.40	0.40;0.40	0.70;0.70	.998	.001	.999	.000	.999	.000	.999	.000	1.000	.000
	0.40;0.40	0.40;0.70	0.40;0.40	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.40	0.40;0.70	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.40	0.40;0.70	0.70;0.40	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.40	0.40;0.70	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.40	0.70;0.40	0.40;0.40	.900	.003	.888	.003	.888	.003	.876	.003	.986	.001
	0.40;0.40	0.70;0.40	0.40;0.70	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.40	0.70;0.40	0.70;0.40	.878	.004	.848	.004	.847	.004	.833	.004	.980	.002
	0.40;0.40	0.70;0.40	0.70;0.70	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.40	0.70;0.70	0.40;0.40	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.40	0.70;0.70	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.40	0.70;0.70	0.70;0.40	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.40	0.70;0.70	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.40;0.40	0.40;0.40	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.40;0.40	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.40;0.40	0.70;0.40	.998	.001	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.40;0.40	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.40;0.70	0.40;0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.40;0.70	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.40;0.70	0.70;0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.40;0.70	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.70;0.40	0.40;0.40	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.70;0.40	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.70;0.40	0.70;0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.70;0.40	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.70;0.70	0.40;0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.70;0.70	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.70;0.70	0.70;0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.40;0.70	0.70;0.70	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.40	0.40;0.40	0.40;0.40	.897	.004	.887	.003	.886	.003	.874	.004	.987	.001
	0.70;0.40	0.40;0.40	0.40;0.70	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.40	0.40;0.40	0.70;0.40	.872	.005	.845	.004	.845	.004	.831	.004	.976	.002
	0.70;0.40	0.40;0.40	0.70;0.70	.998	.001	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.40	0.40;0.70	0.40;0.40	.998	.001	.999	.000	.999	.000	.999	.000	1.000	.000
	0.70;0.40	0.40;0.70	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000

(Devam ediyor)

Tablo 24.2 (Devam)

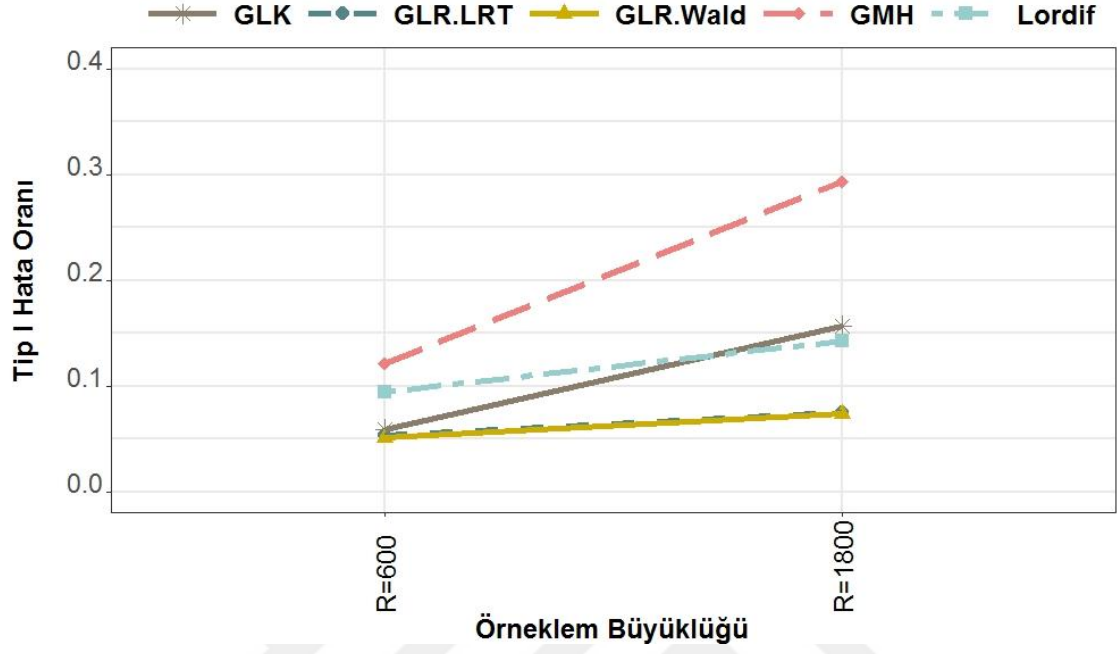
Dört Grupta TBO(ab) DMF Türü İçin Referans Grubun Örneklem Büyüklüğü ve DMF Miktarına Göre Hesaplanan Ortalama Tip III Hata Oranları ve Standart Hatalar

Referans Örneklem Büyüklük	Odak1	Odak2	Odak3	GLK		GLR-LRT		GLR-Wald		GMH		Lordif	
	DMF	DMF	DMF	Tip	St.	Tip	St.	Tip	St.	Tip	St.	Tip	St.
	Miktarı	Miktarı	Miktarı	III	Hata	III	Hata	III	Hata	III	Hata	III	Hata
	a ve b	a ve b	a ve b	Hata		Hata		Hata		Hata		Hata	
	0.70;0.40	0.40;0.70	0.70;0.40	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.40	0.40;0.70	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.40	0.70;0.40	0.40;0.40	.877	.003	.847	.004	.846	.004	.832	.004	.977	.001
	0.70;0.40	0.70;0.40	0.40;0.70	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.40	0.70;0.40	0.70;0.40	.846	.003	.791	.003	.791	.003	.774	.003	.967	.001
	0.70;0.40	0.70;0.40	0.70;0.70	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.40	0.70;0.70	0.40;0.40	.999	.000	.999	.000	.999	.000	.999	.000	1.000	.000
	0.70;0.40	0.70;0.70	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.40	0.70;0.70	0.70;0.40	.999	.000	1.000	.000	1.000	.000	.999	.000	1.000	.000
	0.70;0.40	0.70;0.70	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.40;0.40	0.40;0.40	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.40;0.40	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.40;0.40	0.70;0.40	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.40;0.40	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.40;0.70	0.40;0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.40;0.70	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.40;0.70	0.70;0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.40;0.70	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.70;0.40	0.40;0.40	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.70;0.40	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.70;0.40	0.70;0.40	.999	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.70;0.40	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.70;0.70	0.40;0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.70;0.70	0.40;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.70;0.70	0.70;0.40	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000
	0.70;0.70	0.70;0.70	0.70;0.70	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000

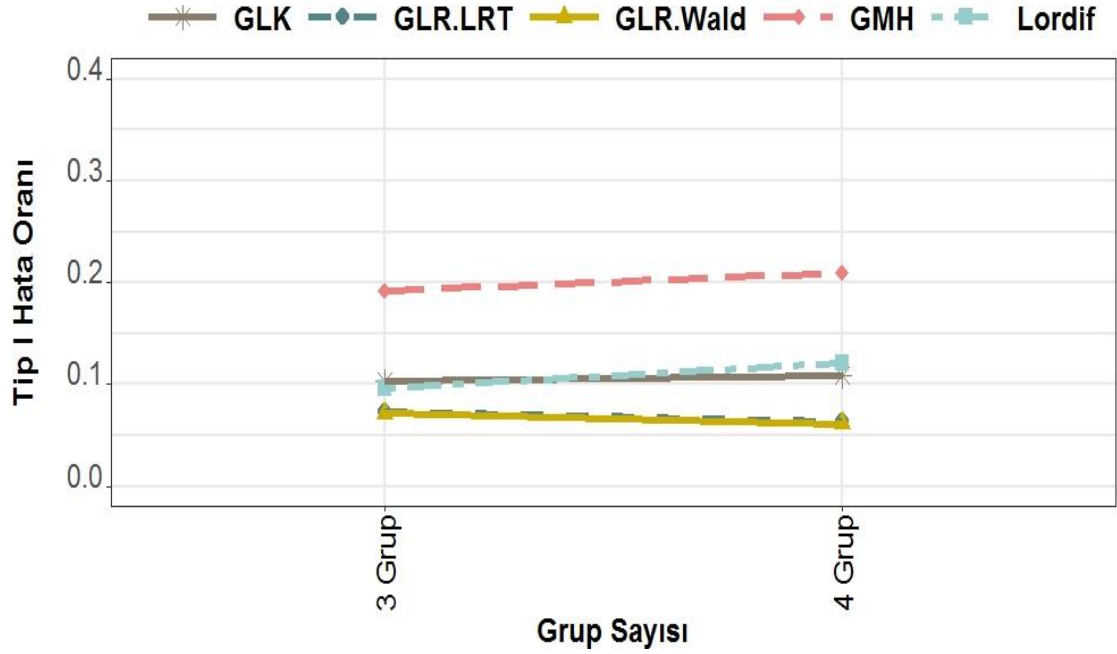


Şekil 24.2. Dört Grupta DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ve TBO(ab) DMF Miktarı Etkileşimi İçin Tip III Hata Oranları

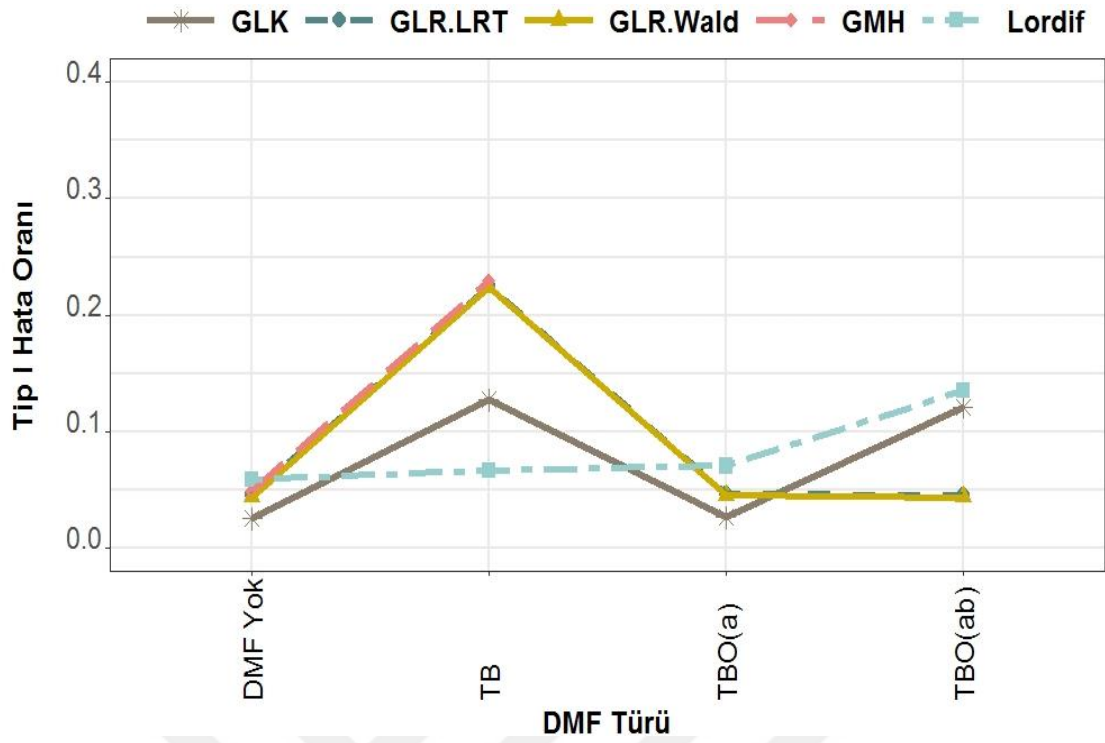
EK 25. Üç ve Dört Grup Birlikte Dikkate Alındığında Hesaplanan Tip I Hata Oranları



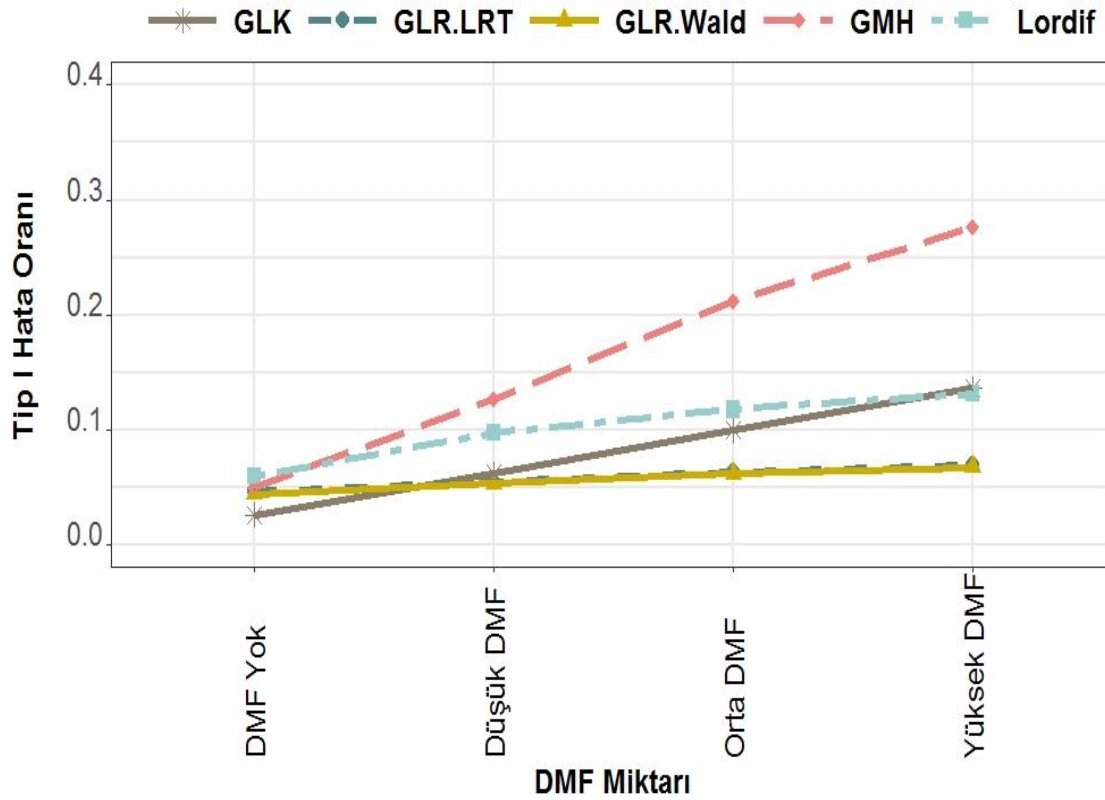
Şekil 25.1. DMF Belirleme Teknikleri ile Referans Grubun Örneklem Büyüklüğü Etkileşimi İçin Tip I Hata Oranları



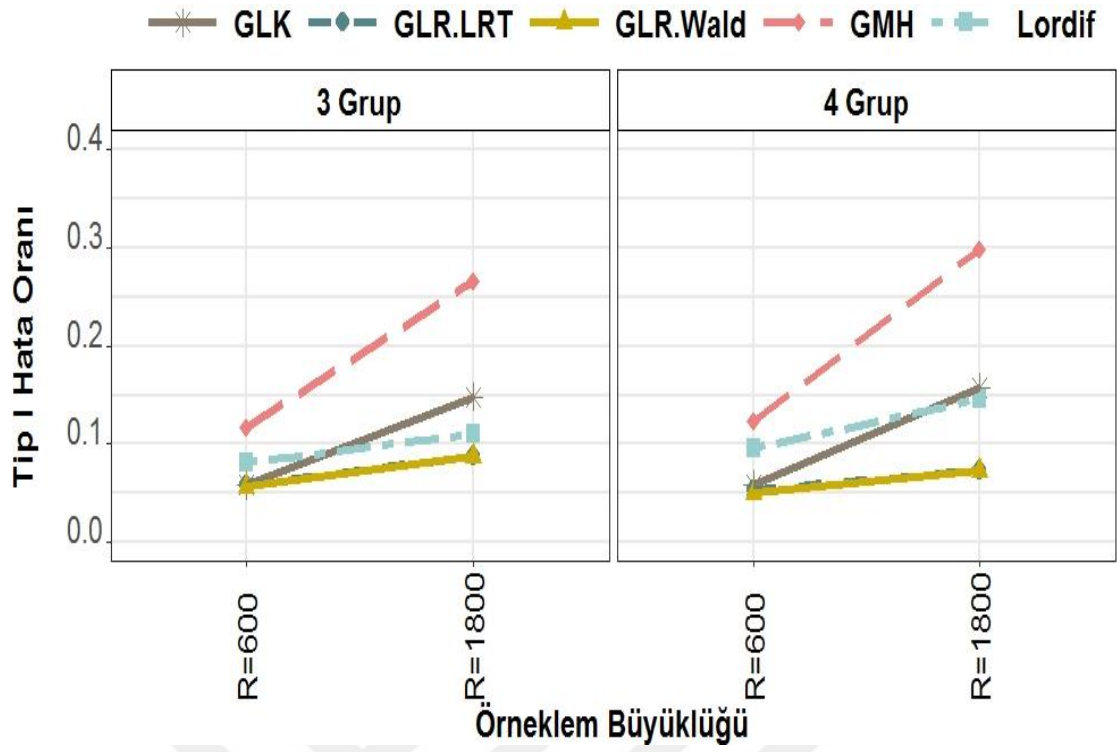
Şekil 25.2. DMF Belirleme Teknikleri ile Grup Sayısı Etkileşimi İçin Tip I Hata Oranları



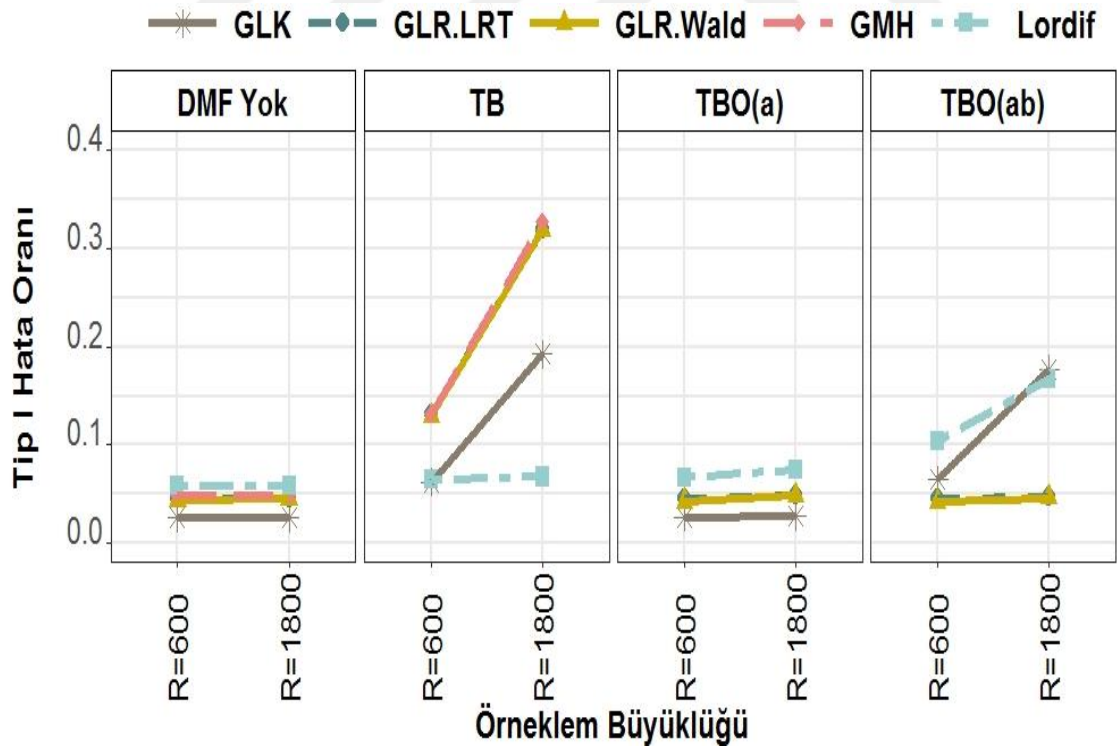
Şekil 25.3. DMF Belirleme Teknikleri ile DMF Türü Etkileşimi İçin Tip I Hata Oranları



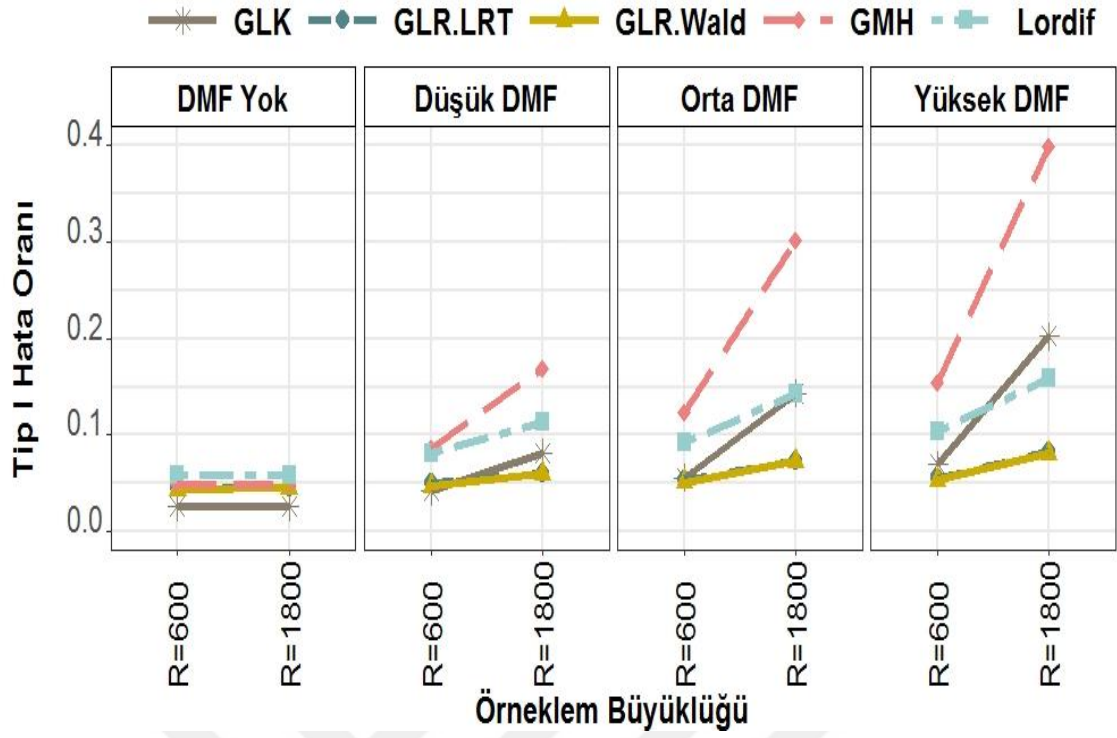
Şekil 25.4. DMF Belirleme Teknikleri ile DMF Miktarı Etkileşimi İçin Tip I Hata Oranları



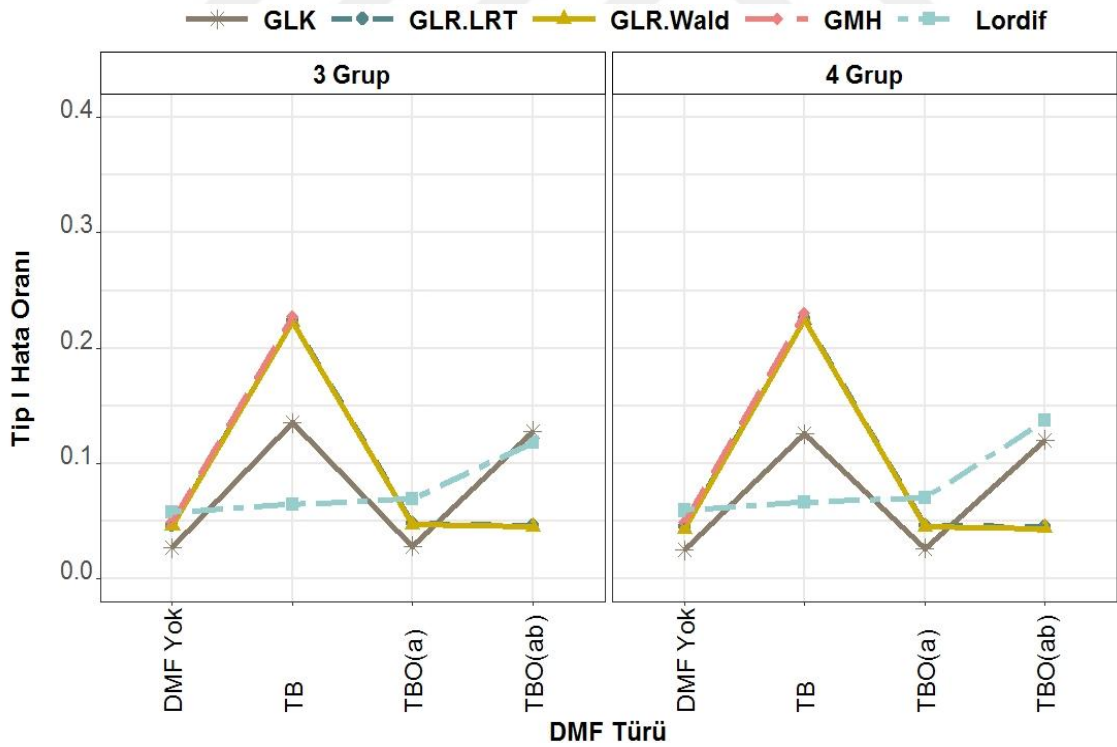
Şekil 25.5. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ile DMF Türü Etkileşimi İçin Tip I Hata Oranları



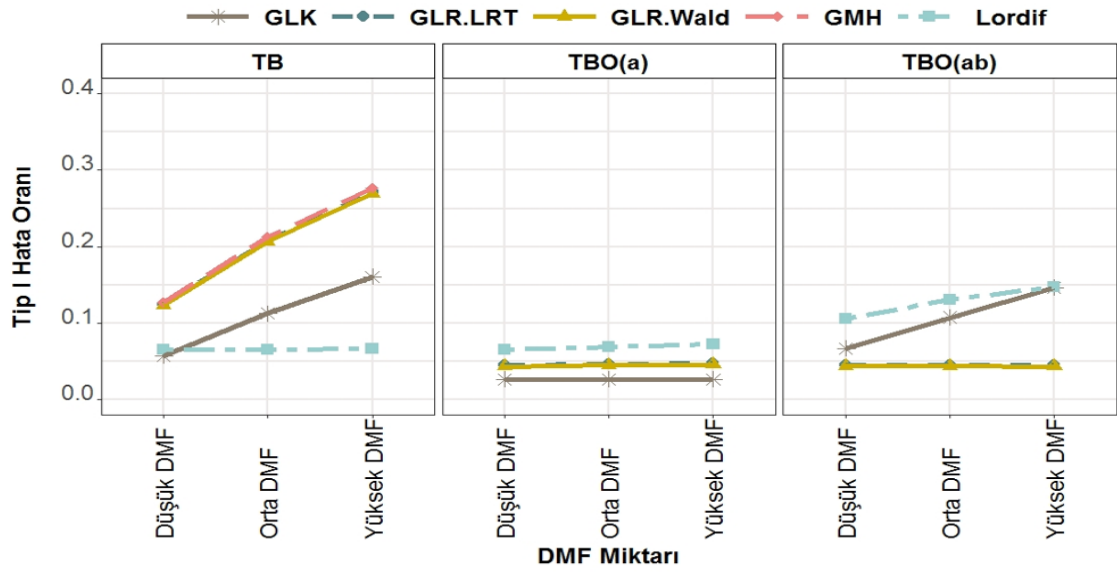
Şekil 25.6. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ile DMF Türü Etkileşimi İçin Tip I Hata Oranları



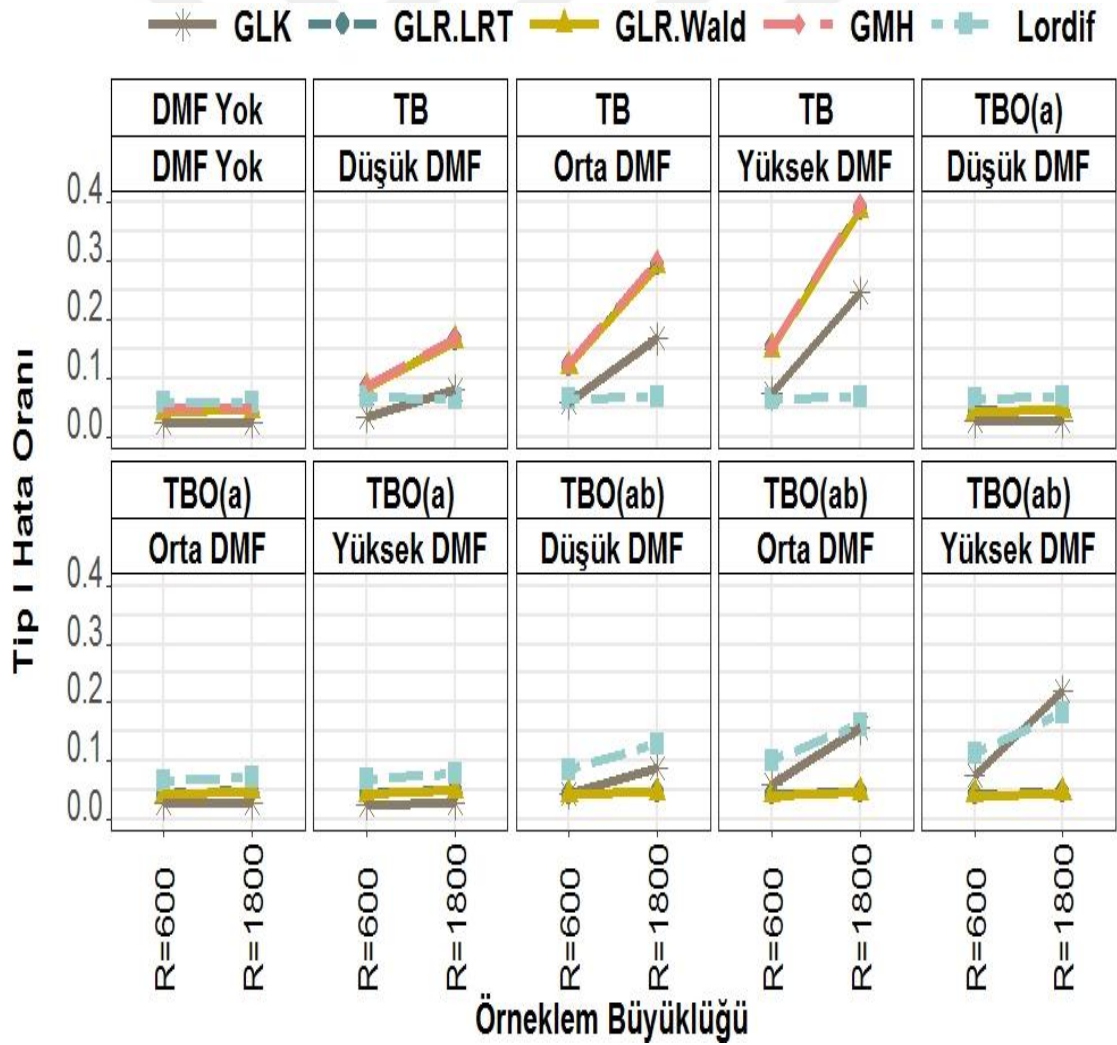
Şekil 25.7. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ile DMF Miktarı Etkileşimi İçin Tip I Hata Oranları



Şekil 25.8. DMF Belirleme Teknikleri, Grup Sayısı ile DMF Türü Etkileşimi İçin Tip I Hata Oranları

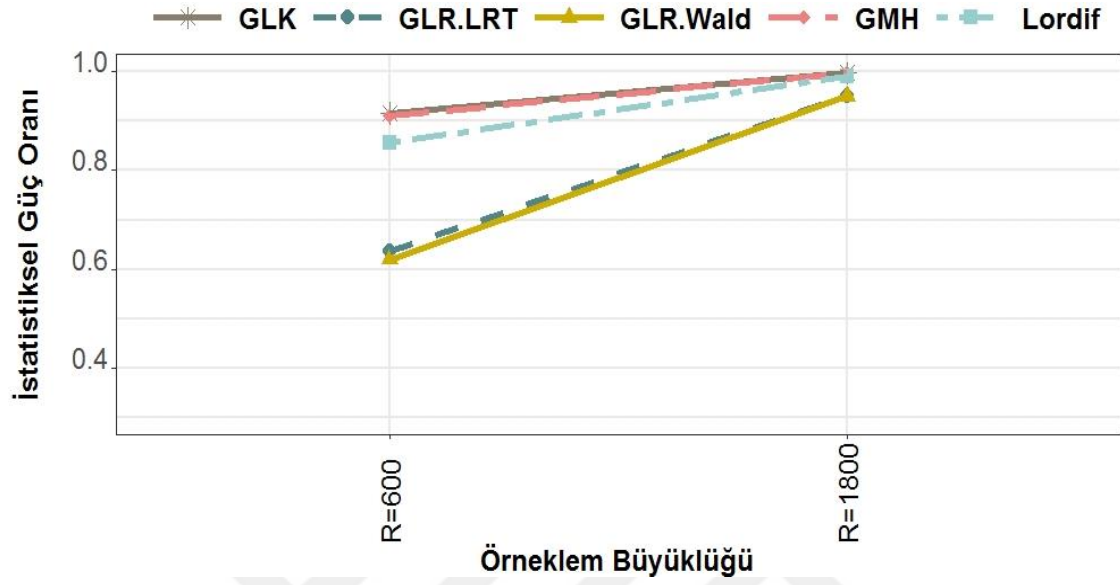


Şekil 25.9. DMF Belirleme Teknikleri, DMF Türü ile DMF Miktarı Etkileşimi İçin Tip I Hata Oranları

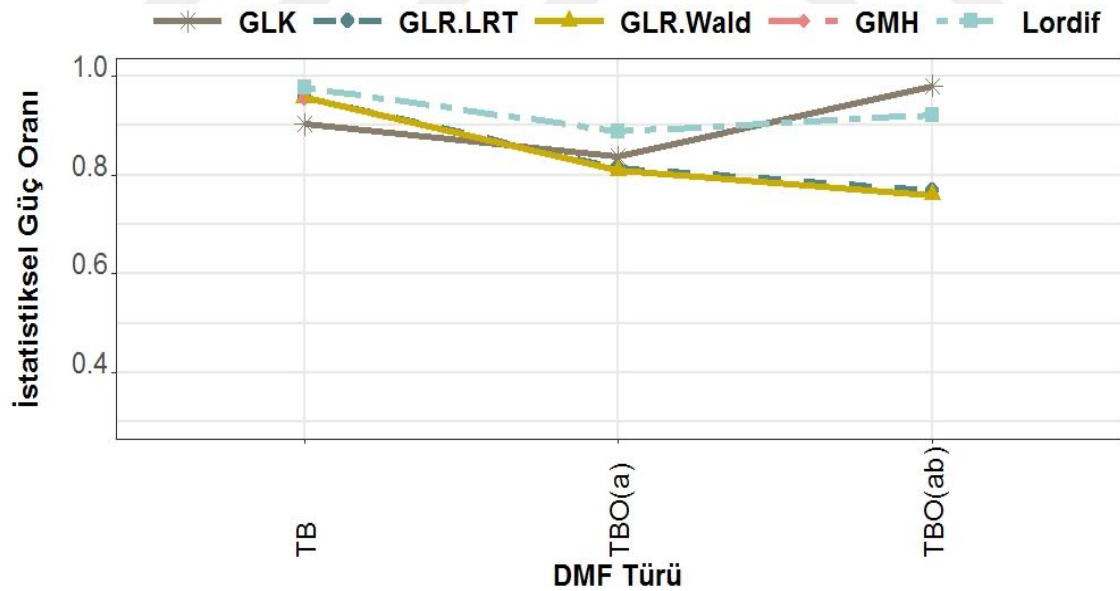


Şekil 25.10. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü, DMF Türü ile DMF Miktarı Etkileşimi İçin Tip I Hata Oranları

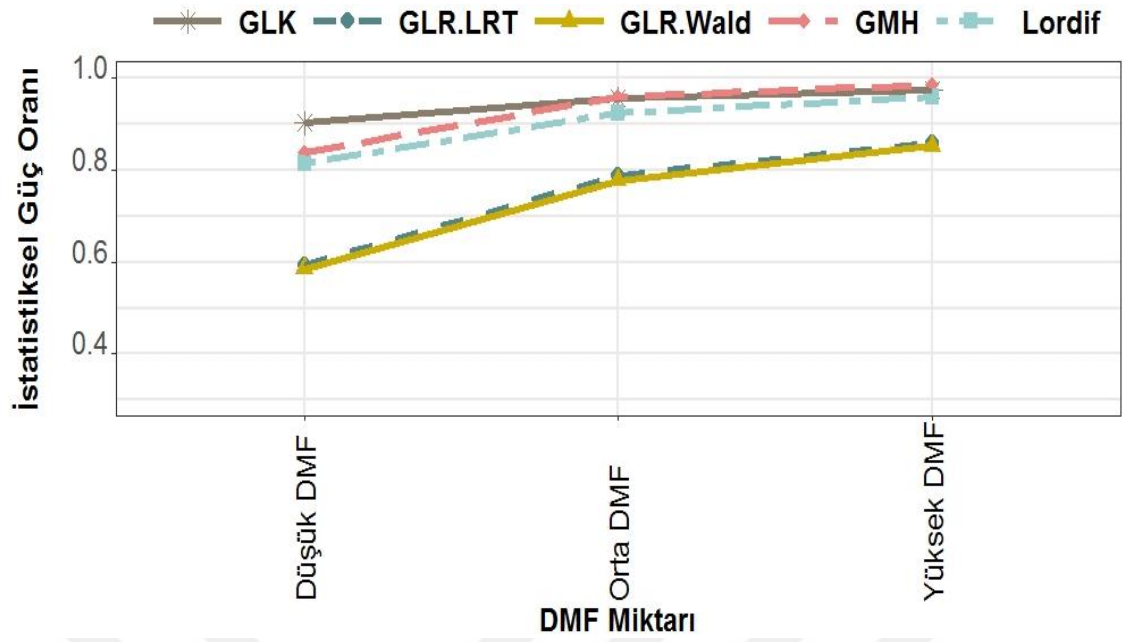
EK 26. Üç ve Dört Grup Birlikte Dikkate Alındığında Hesaplanan İstatistiksel Güç Oranları



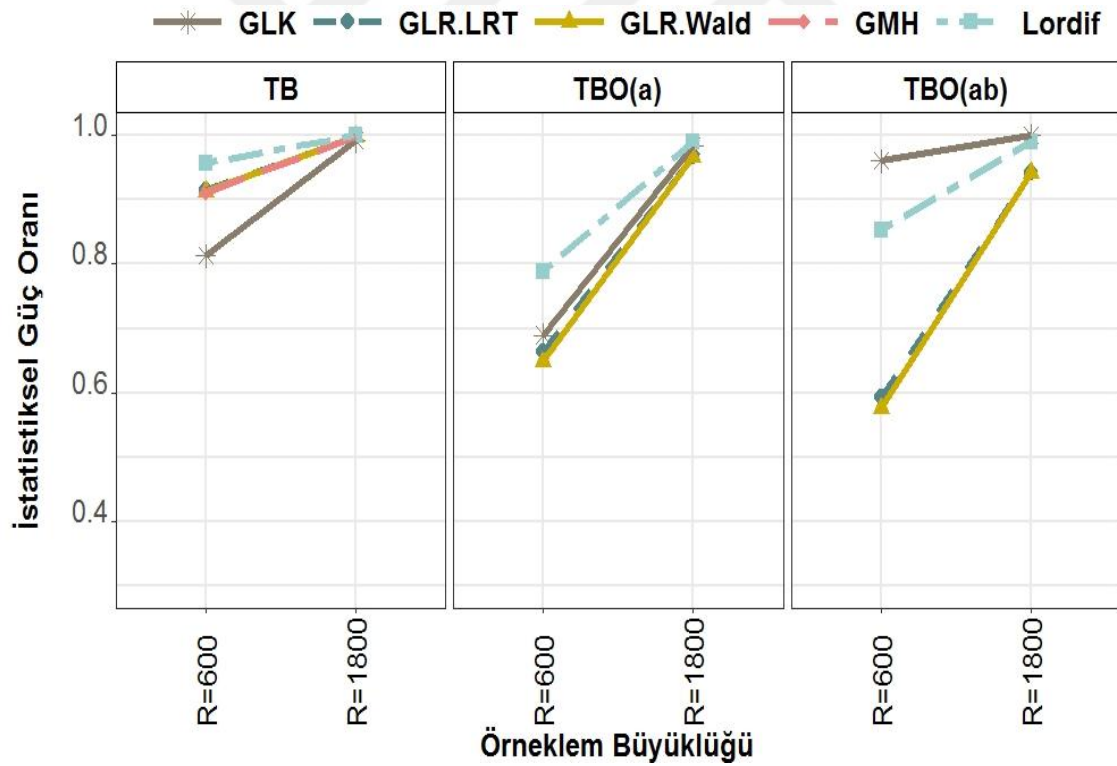
Şekil 26.1. DMF Belirleme Teknikleri ile Referans Grubun Örneklem Büyüklüğü Etkileşimi İçin İstatistiksel Güç Oranları



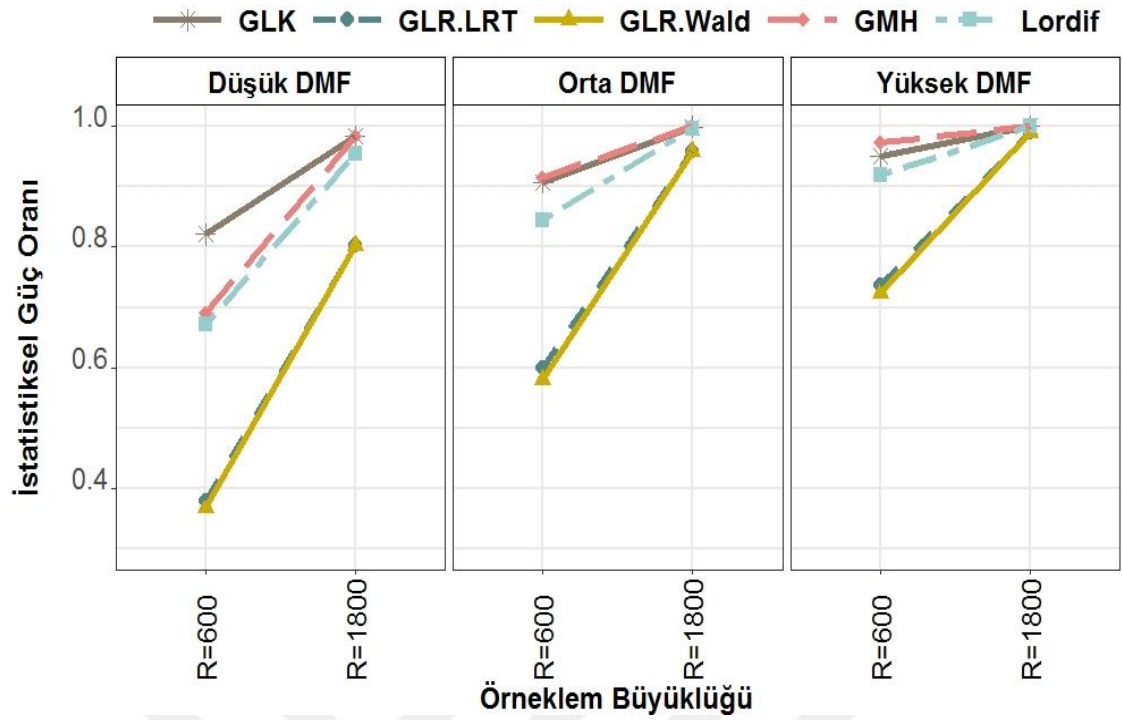
Şekil 26.2. DMF Belirleme Teknikleri ile DMF Türü Etkileşimi İçin İstatistiksel Güç Oranları



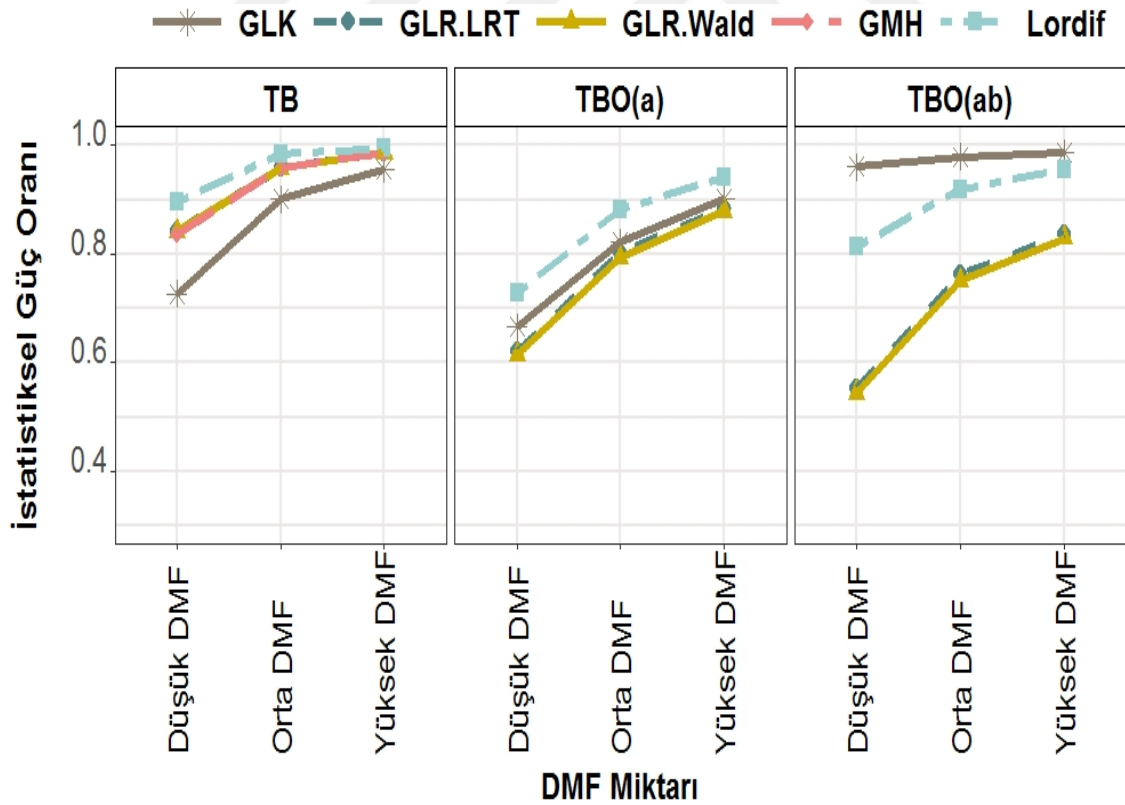
Şekil 26.3. DMF Belirleme Teknikleri ile DMF Miktarı Etkileşimi İçin İstatistiksel Güç Oranları



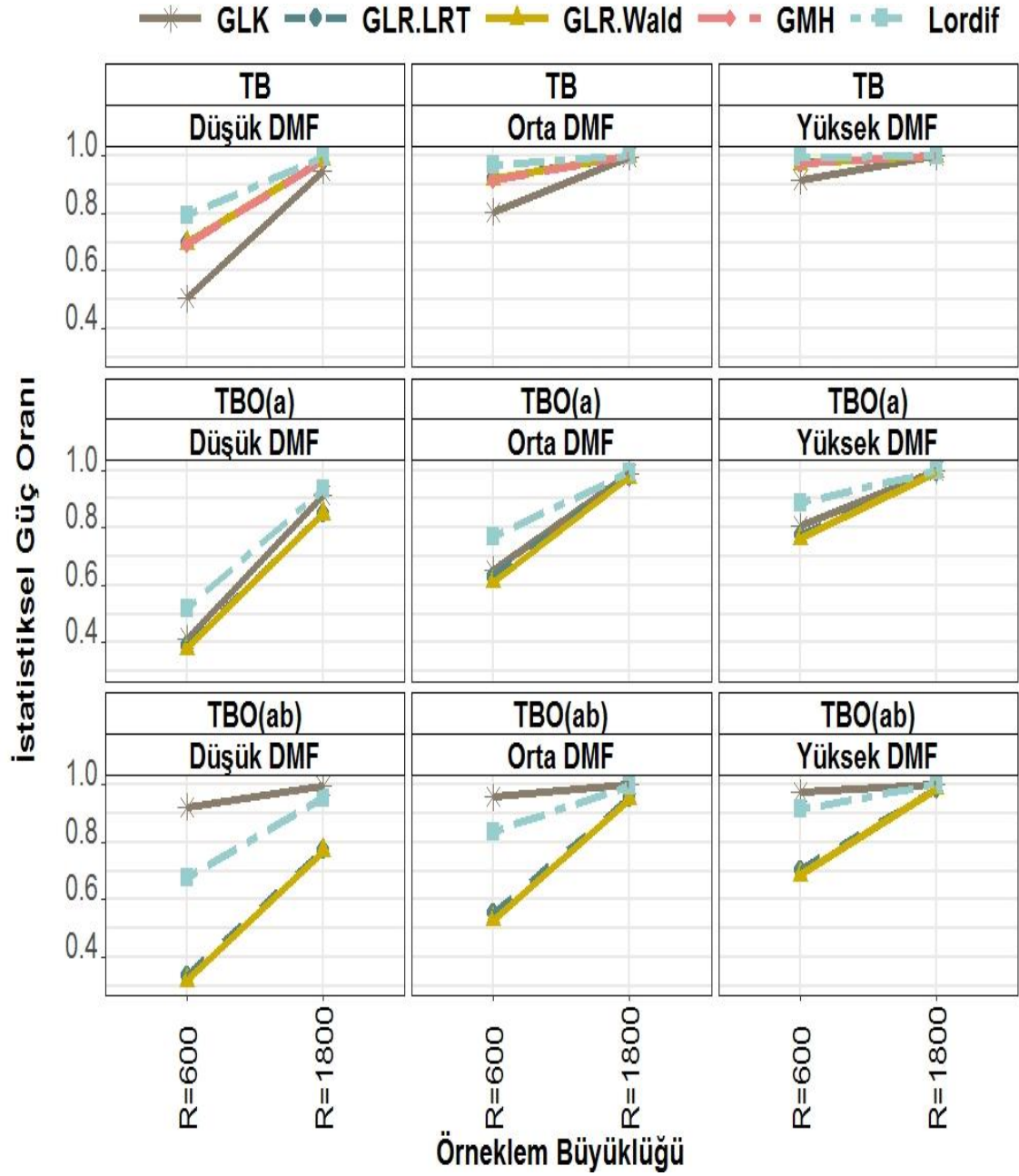
Şekil 26.4. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ile DMF Türü Etkileşimi İçin İstatistiksel Güç Oranları



Şekil 26.5. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ile DMF Miktarı Etkileşimi İçin İstatistiksel Güç Oranları

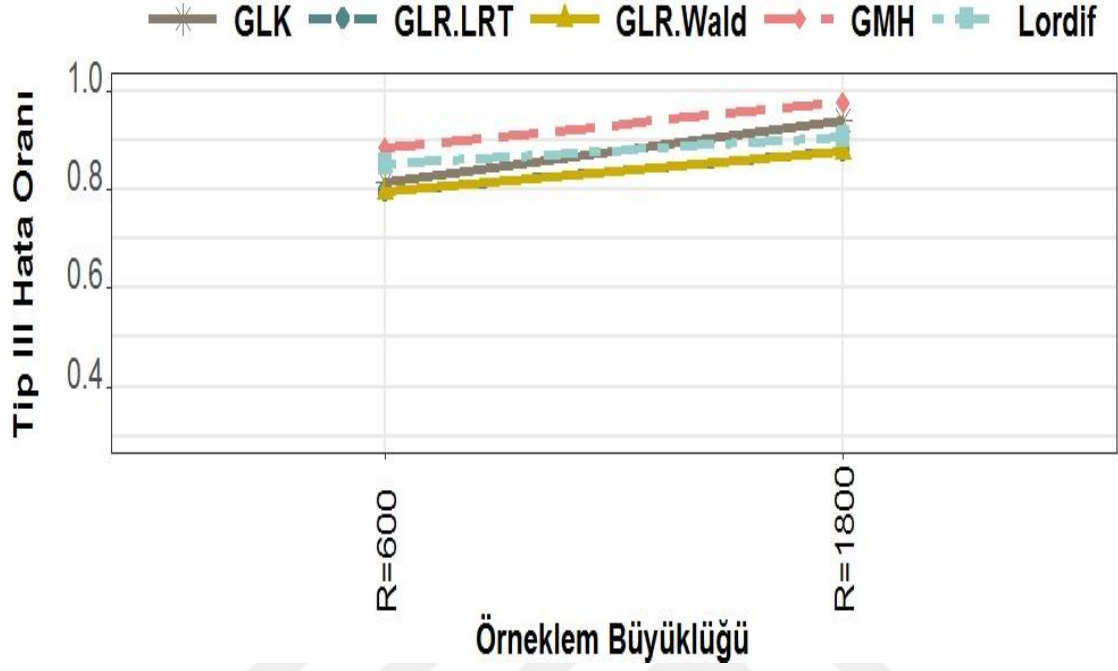


Şekil 26.6. DMF Belirleme Teknikleri, DMF Türü ile DMF Miktarı Etkileşimi İçin İstatistiksel Güç Oranları

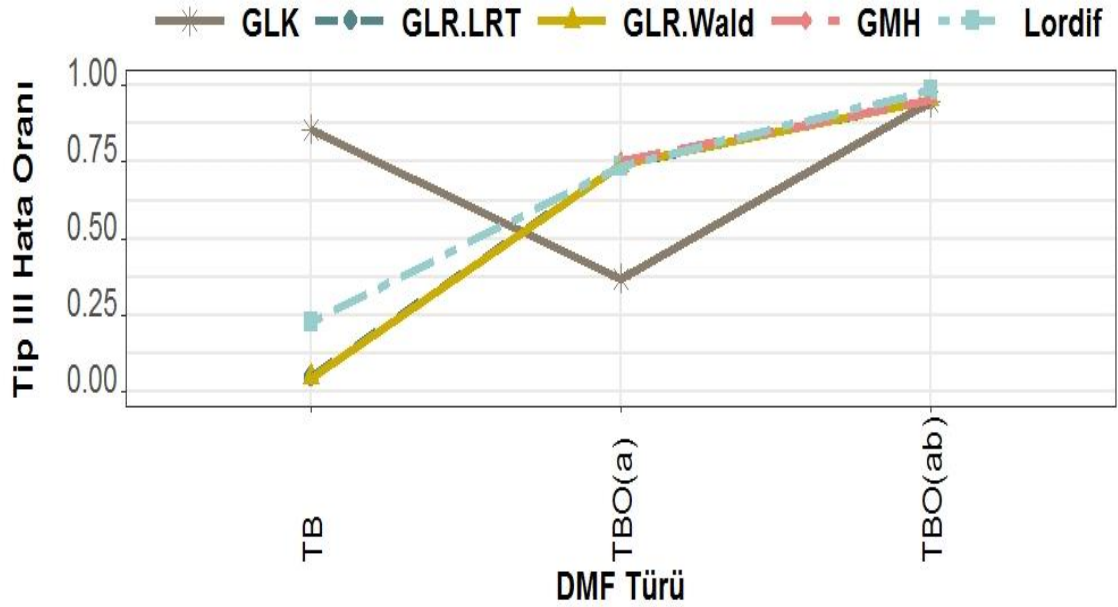


Şekil 26.7. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü, DMF Türü ile DMF Miktarı Etkileşimi İçin İstatistiksel Güç Oranları

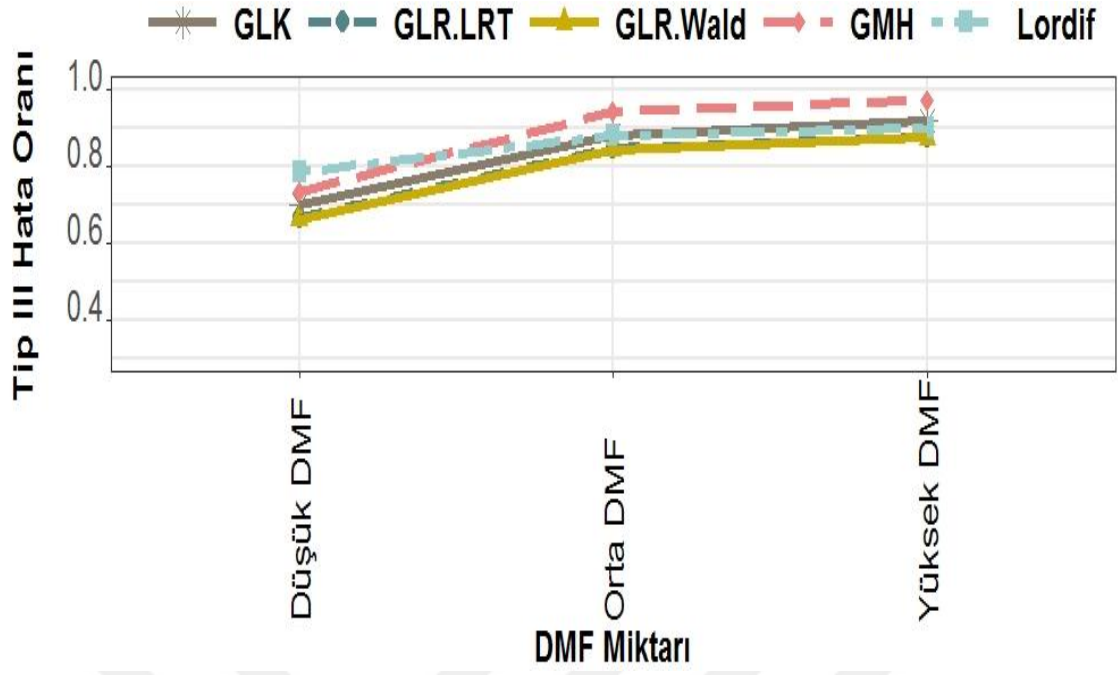
EK 27. Üç ve Dört Grup Birlikte Dikkate Alındığında Hesaplanan Tip III Hata Oranları



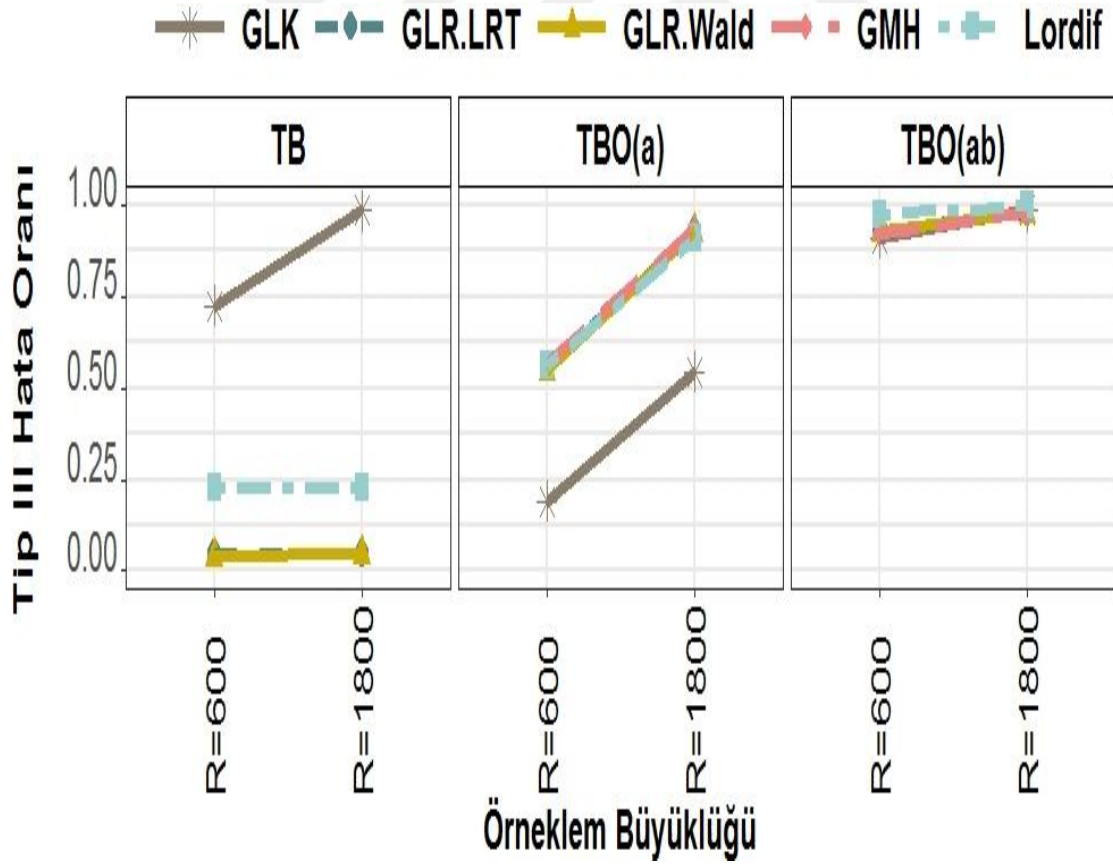
Şekil 27.1. DMF Belirleme Teknikleri ile Referans Grubun Örneklem Büyüklüğü Etkileşimi İçin Tip III Hata Oranları



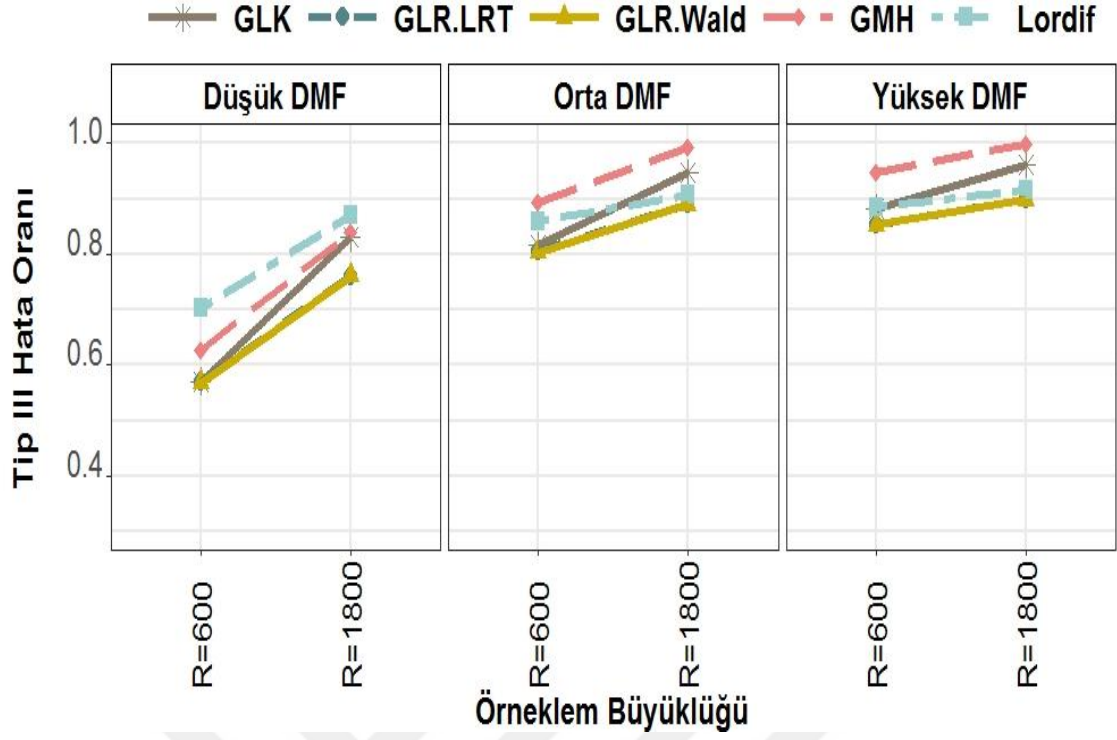
Şekil 27.2. DMF Belirleme Teknikleri ile DMF Türü Etkileşimi İçin Tip III Hata Oranları



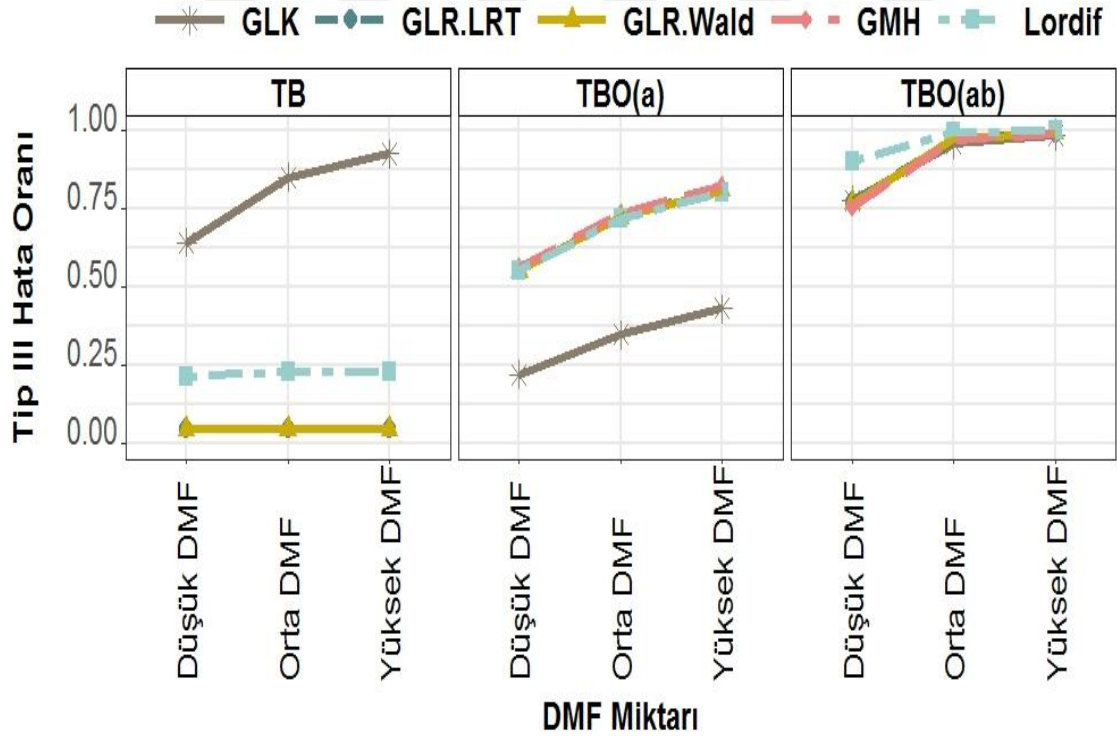
Şekil 27.3. DMF Belirleme Teknikleri ile DMF Miktarı Etkileşimi İçin Tip III Hata Oranları



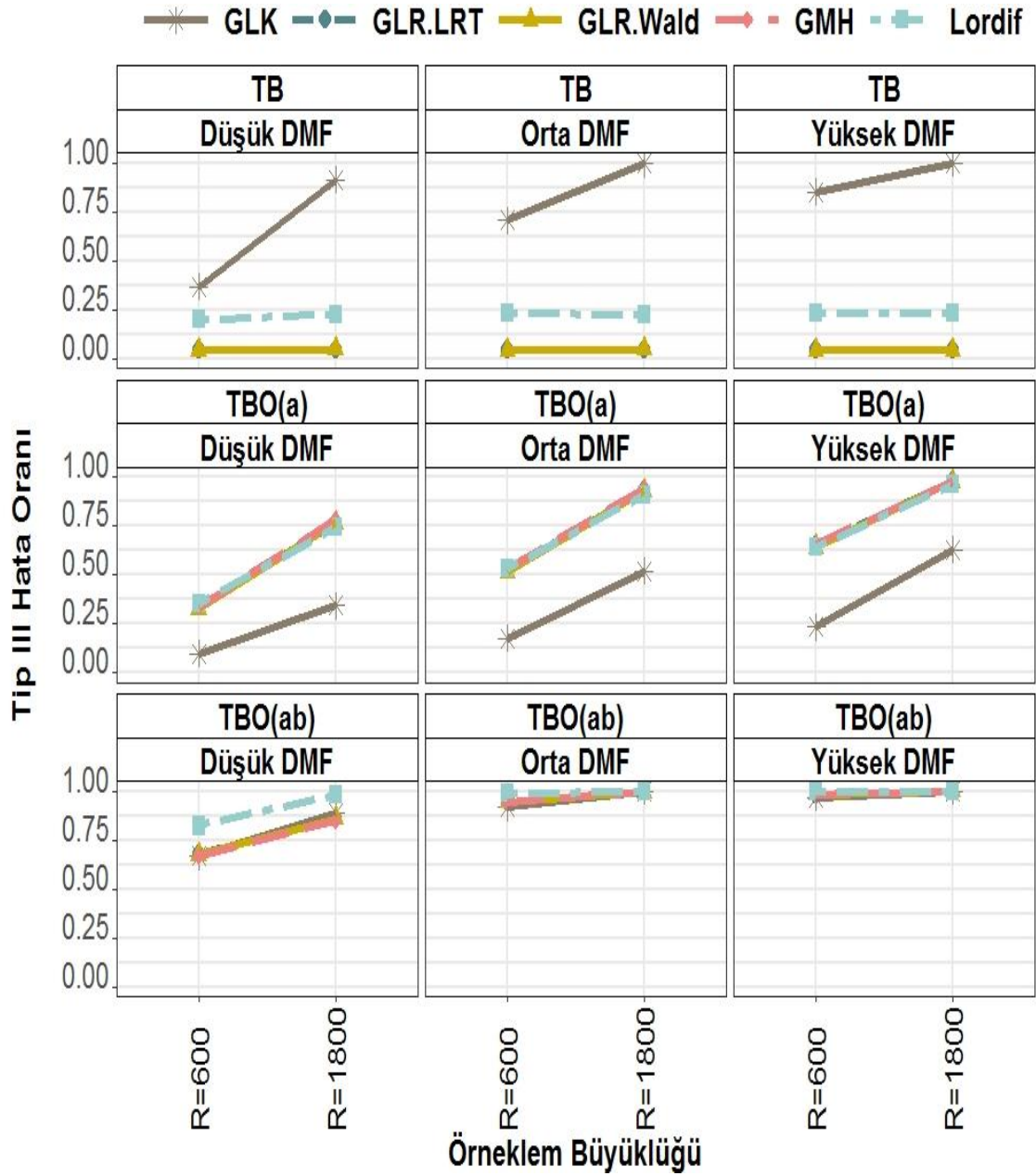
Şekil 27.4. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ile DMF Türü Etkileşimi İçin Tip III Hata Oranları



Şekil 27.5. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü ile DMF Miktarı Etkileşimi İçin Tip III Hata Oranları



Şekil 27.6. DMF Belirleme Teknikleri, DMF Türü ile DMF Miktarı Etkileşimi İçin Tip III Hata Oranları



Şekil 27.7. DMF Belirleme Teknikleri, Referans Grubun Örneklem Büyüklüğü, DMF Türü ile DMF Miktarı Etkileşimi İçin Tip III Hata Oranları

EK 28. Etik Kurul Belgesi

ANKARA ÜNİVERSİTESİ SOSYAL BİLİMLER ALT ETİK KURULU KARAR ÖRNEĞİ

Karar Tarihi : 22/04/2019

Toplantı Sayısı : 05

Karar Sayısı : 169

169- Üniversitemiz Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Anabilim Dalı doktora öğrencisi **Özkan Saatçioğlu**'nun "Çoklu Grup Değişen Madde Fonksiyonu Belirleme Tekniklerinin Farklı Simülasyon Koşullarında Karşılaştırılması" başlıklı tezi ile ilgili 28/03/2019 tarihli "İnsan Üzerinde Yapılan Klinik Dışı Araştırmalar Başvuru Formu" Etik Kurulumuzca incelendi.

Üniversitemiz Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Anabilim Dalı doktora öğrencisi **Özkan Saatçioğlu**'nun "Çoklu Grup Değişen Madde Fonksiyonu Belirleme Tekniklerinin Farklı Simülasyon Koşullarında Karşılaştırılması" başlıklı tezinin, insan üzerinde yapılan bir çalışma olmaması nedeniyle etik kurul onayına ihtiyaç duymadığına oy birliği ile karar verildi.



BENZERLİK BİLDİRİMİ

“Çoklu Grup Değişen Madde Fonksiyonu Belirleme Tekniklerinin Farklı Simülasyon Koşullarında Karşılaştırılması” başlıklı tezimin ana bölümü (ön bölüm, kaynaklar ve ekler hariç) Turnitin İntihali Engelleme Programı aracılığıyla incelenmiş ve ilgili rapor danışmanım tarafından da kontrol edilmiştir. Kontrol sırasında (1) “Beş sözcükten daha az olan benzeşmeler” (2) “Kaynaklar” (3) “Doğrudan Alıntılar” dışarıda tutulmuştur. Benzerlik kontrolüne ilişkin rapordan elde edilen sonuçlar aşağıda sunulmuştur.

Rapor Tarihi	: 08.05.2022
Gönderim Numarası	: 1831011484
Sayfa Sayısı	: 262
Sözcük Sayısı	: 90810
Karakter Sayısı	: 591933
Benzerlik Oranı	: %2
Savunma Tarihi	: 23/05/2022

Yukarıda belirtilen sonuçları gösteren Turnitin İntihali Engelleme Programı’na ilişkin orijinal raporu, sonuçlarda herhangi bir değişiklik yapmaksızın bu beyanım ekinde Enstitüye teslim ettiğimi, tezimin %10’dan fazla benzerlik oranı içerdiğinin belirlenmesi durumunda, bundan doğabilecek tüm yasal sorumluluğu kabul ettiğimi bildirir, saygılarımı sunarım.

Öğrencinin Adı Soyadı: Özkan SAATÇİOĞLU

Tarih: 23/05/2022

İmza:

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı ve Soyadı : Özkan SAATÇIOĞLU

E-Posta Adresi :

İş Deneyimi :

Unvan	Görev Yeri	Yıl
Araş. Gör.	Muş Alparslan Üniversitesi	2010-2011
Araş. Gör.	Ankara Üniversitesi	2011-2018
Araş. Gör.	Muş Alparslan Üniversitesi	2018-

Akademik Bilgiler

Öğrenim Durumu:

Derece	Bölüm/Program	Üniversite	Yıl
Lisans	İlköğretim Matematik	Atatürk Üniversitesi	2005-2009
Yüksek Lisans	Ölçme ve Değerlendirme	Ankara Üniversitesi	2011-2014
Doktora	Ölçme ve Değerlendirme	Ankara Üniversitesi	2014-2022

Yayınlar:

Çam, Z., Eşkisü, M., Kardaş, F., **Saatçioğlu, Ö.**, & Gelibolu, S. (2020). The mediating role of self-efficacy in the relationship between problem solving and hope. *Participatory Educational Research*, 7(1), 47-58. [ERIC]

Karademir, A., Yaman, F., & **Saatçioğlu, Ö.** (2020). Challenges of higher education institutions against COVID-19: The case of Turkey. *Journal of Pedagogical Research*, 4(4), 453-474. [ERIC]

Saatçioğlu, Özkan., & Gülleroğlu, H. D. (2017). Evaluation of reading skills subtest results of countries participating in the PISA 2009 application via profile analysis. *Education and Science*, 42(190), 401-422. [SSCI]

Demir, E., **Saatçioğlu, Ö.**, Ve İmrol, F. (2016). Uluslararası dergilerde yayımlanan eğitim araştırmalarının normallik varsayımları açısından incelenmesi. *Current Research in Education*, 2(3), 130-148. [Diğer İndeksler]