



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

TEST BOYUTLULUĞUNUN DEĞİŞEN MADDE FONKSİYONUNA ETKİSİNİN
FARKLI KOŞULLAR ALTINDA İNCELENMESİ

Bulut YILDIZTEKİN

Doktora Tezi

Ankara, 2020

Liderlik, arařtırma, inovasyon, kaliteli eęitim ve deęiřim ile

Daha ileriye ... En İyiyeye ...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

TEST BOYUTLULUĞUNUN DEĞİŞEN MADDE FONKSİYONUNA ETKİSİNİN
FARKLI KOŞULLAR ALTINDA İNCELENMESİ

THE INVESTIGATION OF TEST DIMENSION EFFECT TO DIFFERENTIAL ITEM
FUNCTIONING UNDER DIFFERENT CONDITIONS

Bulut YILDIZTEKİN

Doktora Tezi

Ankara, 2020

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Bulut YILDIZTEKİN'in hazırladıđı "Test Boyutluluđunun Deđişen Madde Fonksiyonuna Etkisinin Farklı Koşullar Altında İncelenmesi" başlıklı bu çalıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiřtir.

J¼ri Başkanı	Prof. Dr. Selahattin GELBAL	İmza
J¼ri Üyesi (Danıřman)	Prof. Dr. H¼lya KELECİOđLU	İmza
J¼ri Üyesi	Prof. Dr. Nuri DOđAN	İmza
J¼ri Üyesi	Doç. Dr. Kaan Z¼lfikar DENİZ	İmza
J¼ri Üyesi	Doç. Dr. Dilara BAKAN KALAYCIOđLU	İmza

Bu tez Hacettepe Üniversitesi Lisansüstü Eđitim, Öğretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 16 / 07 / 2020 tarihinde uygun gör¼lm¼ř ve Enstitü Yönetim Kurulunca / / tarihinde kabul edilmiřtir.

Prof. Dr. Selahattin GELBAL
Eđitim Bilimleri Enstitüsü M¼d¼r¼

Öz

Bu araştırmanın amacı iki boyutlu ve üç boyutlu yapay ve iki boyutlu gerçek veri setlerinde DMF'nin farklı koşullardan nasıl etkilendiğini belirlemektir. Testlerin basit ya da karmaşık yapıda olmaları, orta veya yüksek düzeyde DMF etki büyüklüğü içermeleri, 600-1200-2400-3600-4800-6000 olarak değişen örneklem büyüklükleri, 1:1, 1:2 olarak değişen odak ve referans grup oranına sahip olmaları araştırmanın koşullarındandır. İki boyutlu ve üç boyutlu olarak üretilen iki kategorili veri setlerine geleneksel SIBTEST ile çok boyutlu çoklu nedenler çoklu göstergeler (Multi-MIMIC) yöntemleri kullanılarak gerçekleştirilen DMF analizleri neticesinde ulaşılan 1. tip hata ve güç değerleri karşılaştırılmıştır. Gerçek veri seti ile gerçekleştirilen analizlerde ise DMF yöntemlerine bağlı olarak belirlenen DMF'li madde sayıları karşılaştırılmıştır. Yapay veri setlerini elde etmek için madde parametreleri ITEMGEN programı kullanılarak üretilmiştir. Gerçek veri seti PISA 2012 Türkiye örnekleminde sırasıyla 600,1200 ve 3600; odak (kız) ve referans (erkek) oranı 1:1 ile 1:2 olarak seçilmiştir. Veri setlerinin çalışmaya uygunluğunun kontrol edilmesi için FACTOR 10.8, yapay ve gerçek veri setleri için gerçekleştirilen analizlerde SAS 9.4 ve MPLUS 7 programları kullanılmıştır. Yapay veri setleri için 100 tekrar yapılmıştır. Araştırma neticesinde iki boyutlu verilerde, örneklem büyüklüğü arttıkça, DMF etki düzeyi yükseldikçe ve testin yapısı basitten karmaşığa gittikçe 1. tip hata azalmaktadır. Güç oranları ise örneklem büyüklüğüne bağlı olarak anlamlı farklılık göstermemekte, DMF etki düzeyinin artmasına bağlı azalmakta ve testin yapısının karmaşıklaşmasıyla yükselmektedir. Üç boyutlu verilerde de paralel sonuçlar çıkmıştır. İki ve üç boyutlu verilerde DMF yöntemi olarak klasik yöntem yerine çok boyutlu model olan Multi-MIMIC yönteminde daha düşük 1. tip hata ve daha yüksek güç oranları elde edilmiştir. Gerçek veri setinde ise DMF'li madde sayısı örneklem büyüklüğüne ve odak-referans grup oranına bağlı olarak değişmezken, DMF yöntemi olarak Multi-MIMIC yöntemi seçildiğinde DMF gösteren madde sayısında azalma olduğu belirlenmiştir. Sonuç olarak yeterli örneklem büyüklüklerine uygulanmış testlerin basit yapıları olması ve DMF analizleri için Multi-MIMIC yönteminin kullanılması önerilmektedir.

Anahtar sözcükler: çok boyutluluk, değişen madde fonksiyonu, SIBTEST, Multi-MIMIC

Abstract

The aim of this study is to determine how DIF is affected by different conditions in two dimensional and three dimensional simulation and two dimensional real data sets. Whether the tests are simple or complex, medium or high level of DIF effect size, sample sizes ranging from 600-1200-2400-3600-4800-6000, and having a focal and reference group ratio varying between 1: 1 and 1:2 are research conditions. Type one error and power values obtained from conventional SIBTEST and multi-dimensional multiple causes multiple indicators (Multi-MIMIC) methods were compared to two-category data sets produced in 2D and 3D models. In the analyzes performed with the real data set, the number of items with DIF determined according to DIF methods were compared. In order to obtain simulation data sets, item parameters were generated by using ITEMGEN program. Real data sets from PISA 2012 Turkey sample was selected as respectively 600, 1200 and 3600 and focus (girls) and reference (male) ratio of 1: 1 and 1: 2. The FACTOR 10.8 program was used to check the suitability of the data sets. SAS 9.4 and MPLUS 7 programs were used in the analyzes performed for simulation and real data sets. 100 replicates were made for simulation data sets. As a result of the research, type 1 error decreases as the sample size increases, DIF effect level increases and the structure of the test goes from simple to complex. Although power ratios do not differ significantly depending on sample size, decrease due to the increase in DIF effect level and increase with the complexity of the test structure. Parallel results were also found in the 3D data, but the effect of the test structure was not significant. In the 2D and 3D data, the Multi-MIMIC method, which is a multi-dimensional model instead of the classical method, has lower type 1 error and higher power ratios. In the real data set, the number of items with DIF did not change depending on sample size and focus-reference group ratio, but when the Multi-MIMIC method was chosen as the DIF method, it was found that there was a decrease in the number of items with DIF. As a result, the tests applied to sufficient sample sizes should be in simple structure and Multi-MIMIC method is recommended for DIF analysis.

Keywords: multi-dimensionality, differential item functioning, SIBTEST, Multi-MIMIC

Teşekkür

Doktora eğitimim boyunca emeğini ve desteğini esirgemeyen, bilgisinden ve deneyiminden istifade ettiğim için kendimi şanslı saydığım kıymetli hocam ve danışmanım Prof. Dr. Hülya KELEÇİOĞLUNA;

Hacettepe Üniversitesinde görev yaptığım süre boyunca bana ve arkadaşlarıma her zaman destek olan, bize zaman ayıran, bizi ailesi gibi gören değerli hocam Prof. Dr. Selahattin GELBAL'a;

İlk tanıdığım günden beri çalışkanlığına hayranlık duyduğum, örnek aldığım, akademik birikimini paylaşmaktan çekinmeyen, ihtiyaç duyan herkese yardıma koşan çok değerli hocam Prof. Dr. Nuri DOĞAN'a;

Tezime sundukları değerli katkılardan dolayı kıymetli hocalarım Doç. Dr. Kaan Zülfikar DENİZ ve Doç. Dr. Dilara BAKAN KALAYCIOĞLU'na;

Sorduğum sorulara içtenlikle yanıtlar aldığım Doç. Dr. Okan BULUT'a, arkadaşım Dr. Sakine Göçer ŞAHİN'e, her türlü teknik desteğini esirgemeyen arkadaşım Haydar KARAMAN'a, Dr. Başak ERDEM KARA'ya;

Hacettepe Üniversitesinde birlikte görev yaptığım değerli hocalarım ve araştırma görevlisi arkadaşlarıma;

Doktora öğrenimim sürecinde yurt içi doktora burs desteği veren TÜBİTAK'a;

Beni bu günlere getiren ve her başarımı borçlu olduğum anneme, rahmetli babama, kardeşlerime, ablama;

Sabrı, desteği, ilgisi ile başarı motivasyonumu arttıran can yoldaşım sevgili eşim Derya ARSLAN YILDIZTEKİN'e ve kızım Zeynep'e sonsuz teşekkürlerimi sunarım.

İçindekiler

Öz.....	ii
Abstract.....	iii
Teşekkür.....	iv
Tablolar Dizini.....	viii
Şekiller Dizini.....	ix
Simgeler ve Kısaltmalar Dizini.....	xi
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	3
Araştırma Problemi.....	5
Alt Problemler.....	5
Sınırlılıklar.....	6
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	7
Klasik Test Kuramı.....	7
Madde Tepki Kuramı.....	8
MTK Modelleri.....	9
Çok Boyutlu MTK Modelleri (ÇBMTK).....	13
Basit Yapı.....	15
Karmaşık Yapı.....	16
Telafisel Modeller.....	17
Telafisel Olmayan Modeller.....	21
ÇBMTK' da Madde Parametreleri ve Vektörleri.....	22
Değişen Madde Fonksiyonu.....	23
Değişen Madde Fonksiyonu Belirleme Yöntemleri.....	25
Test Geçerliği.....	32
Bölüm 2 İlgili Araştırmalar.....	35

Bölüm 3 Yöntem.....	42
Araştırmanın Modeli	42
Çalışma Grubu	42
Simülasyon Koşulları.....	42
Verilerin Analizi	49
1. Tip Hata ve İstatistiksel Gücün Belirlenmesi	49
Bölüm 4 Bulgular ve Yorumlar.....	50
Alt Problem 1 e Ait Bulgular	50
Alt Problem 2'ye Ait Bulgular.....	54
İki boyutlu Veri Seti İçin ANOVA Sonuçları	59
Alt Problem 3'e Ait Bulgular	62
Alt Problem 4'e Ait Bulgular	67
Üç boyutlu Veri Seti İçin ANOVA Sonuçları	72
Alt Problem 5'e Ait Bulgular	75
Bölüm 5 Sonuç, Tartışma ve Öneriler	78
Sonuçlar.....	78
Tartışma.....	81
Uygulayıcılara Yönelik Öneriler.....	84
İleri Araştırmalara Yönelik Öneriler	84
Kaynaklar	87
EKLER DİZİNİ	104
EK-A1: Madde Parametreleri İki Boyutlu Basit Yapılı Test	105
EK-A2: Madde Parametreleri İki Boyutlu Karmaşık Yapılı Test.....	106
EK-A3: Madde Parametreleri Üç Boyutlu Basit Yapılı Test	107
EK-A4: Madde Parametreleri Üç Boyutlu Karmaşık Yapılı Test	108
EK-B: Güç Oranlarına ait Grafikler	109
EK-C: Gerçek Veri Seti İçin DMF Analizleri Çıktıları.....	113

EK-Ç: SAS Programı Kodları.....	119
EK-D: Multi-MIMIC Yöntemi için MPLUS Kodu	123
EK-E: Etik Komisyonu Onay Bildirimi	124
EK-F: Etik Beyanı	125
EK-G: Doktora Tez Çalışması Orijinallik Raporu	126
EK-Ğ: Dissertation Originality Report	127
EK-H: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı.....	127

Tablolar Dizini

Tablo 1 Veri Yapısına göre MTK Modelleri.....	10
Tablo 2 Basit yapıda parametreler	15
Tablo 3 MH Yönteminde Veri Düzeni	26
Tablo 4 MH Yöntemi ile DMF Düzeyi Belirlenmesi	27
Tablo 5 LR Yöntemi ile DMF Düzeyi Belirlenmesi	28
Tablo 6 SIBTEST Yöntemi ile DMF Düzeyi Belirlenmesi.....	30
Tablo 7 MTK-OO Yöntemi ile DMF Düzeyi Belirlenmesi	30
Tablo 8 Odak-Referans Grup Büyüklükleri.....	44
Tablo 9 Simülasyon Koşulları.....	45
Tablo 10 İki Boyutlu Basit Yapılı ve Orta Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları.....	50
Tablo 11 İki Boyutlu Basit Yapılı ve Yüksek Düzeyde DMF Etki Büyüklüğüne (C düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları	52
Tablo 12 İki Boyutlu Karmaşık Yapılı ve Orta Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları	55
Tablo 13 İki Boyutlu Karmaşık Yapılı ve Yüksek Düzeyde DMF Etki Büyüklüğüne (C düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları	56
Tablo 14 1. Tip Hata ve Güç oranları için ANOVA Sonuçları.....	60
Tablo 15 Üç Boyutlu Basit Yapılı ve Orta Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları.....	63
Tablo 16 Üç Boyutlu Basit Yapılı ve Yüksek Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları	64
Tablo 17 Üç Boyutlu Karmaşık Yapılı ve Orta Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları	68
Tablo 18 İki Boyutlu Karmaşık Yapılı ve Yüksek Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları	69
Tablo 19 1. Tip Hata ve Güç oranları için ANOVA Sonuçları.....	73
Tablo 20 PISA Verisi DMF Analizi Sonuçları.....	76

Şekiller Dizini

Şekil 1. Madde karakteristik eğrisi	8
Şekil 2. 1 PLM için madde karakteristik eğrisi	11
Şekil 3. 2 PLM için madde karakteristik eğrisi	12
Şekil 4. 3 PLM için madde karakteristik eğrisi	13
Şekil 5. Veri setinin yapısına göre ÇBMTK modelleri	14
Şekil 6. Basit yapı.....	15
Şekil 7. Karmaşık yapı.....	16
Şekil 8. İki boyutlu Rasch model MKY.....	18
Şekil 9. Çok boyutlu 2 PLM için MKY ve eş olasılık konturu.....	19
Şekil 10. Çok boyutlu 3 PLM için MKY ve eş olasılık konturu.....	20
Şekil 11. Çok boyutlu normal ogive model için MKY ve eş olasılık konturu.....	21
Şekil 12. Çok boyutlu telafisel olmayan modeller için MKY	22
Şekil 13. Çok boyutlu telafisel olmayan modeller için eş olasılık konturları	22
Şekil 14. Maddelerin üç boyutlu düzlemde gösterilmesi	23
Şekil 15. Tek biçimli DMF	24
Şekil 16. Tek biçimli olmayan DMF	25
Şekil 17. İki boyutlu basit yapı DMF (1:1) için 1. tip hata oranları	53
Şekil 18. İki boyutlu basit yapı DMF (1:2) için 1. tip hata oranları	54
Şekil 19. İki boyutlu karmaşık yapı DMF (1:1) için 1. tip hata oranları	58
Şekil 20. İki boyutlu karmaşık yapı DMF (1:2) için 1. tip hata oranları	59
Şekil 21. Üç boyutlu basit yapı DMF (1:1) için 1. tip hata oranları	66
Şekil 22. Üç boyutlu basit yapı DMF (1:2) için 1. tip hata oranları	67
Şekil 23. Üç boyutlu karmaşık yapı DMF (1:1) için 1. tip hata oranları	71
Şekil 24. Üç boyutlu karmaşık yapı DMF (1:2) için 1. tip hata oranları	72
Şekil 25. DMF gösteren madde sayıları	77
Şekil 26. İki boyutlu basit yapı DMF (1:1) için güç oranları	109
Şekil 27. İki boyutlu basit yapı DMF (1:2) için güç oranları	109
Şekil 28. İki boyutlu karmaşık yapı DMF (1:1) için güç oranları	110
Şekil 29. İki boyutlu karmaşık yapı DMF (1:2) için güç oranları	110
Şekil 30. Üç boyutlu basit yapı DMF (1:1) için güç oranları	111
Şekil 31. Üç boyutlu basit yapı DMF (1:2) için güç oranları	111
Şekil 32. Üç boyutlu karmaşık yapı DMF (1:1) için güç oranları	112

Şekil 33. Üç boyutlu karmaşık yapılı DMF (1:2) için güç oranları 112

Simgeler ve Kısaltmalar Dizini

AGFI: Adjusted Goodness of Fit Index (Uyum İyiliği İndeksi)

ANOVA: Analysis of Variance (Varyans Analizi)

CFI: Comparative Fit Index (Karşılaştırmalı Uyum İndeksi)

ÇBDMF: Çok Boyutlu Değişen Madde Fonksiyonu

ÇBMA: Çok Boyutlu Madde Ayırıcılığı

ÇBMTK: Çok Boyutlu Madde Tepki Kuramı

DMF: Değişen Madde Fonksiyonu

DMG: Dönüştürülmüş Madde Güçlüğü

GFI: Goodness of Fit Index (Uyum İyiliği İndeksi)

HGLM: Hierarchical Generalized Linear Model (Aşamalı Genelleştirilmiş Doğrusal Model)

KMO: Keiser-Meyer-Olkin (Örnekleme Uygunluğu Testi)

KTK: Klasik Test Kuramı

LR: Lojistik Regresyon Yöntemi

MH: Mantel-Haenszel Yöntemi

MKY: Madde Karakteristik Yüzeyi

ÇBRTÇDLM: Çok Boyutlu Rasch Tabanlı Çok Düzeyli Lojistik Modeli

MTK: Madde Tepki Kuramı

MTK-OO: Madde Tepki Kuramı Olabilirlik Oran Testi Yöntemi

Multi-MTK-OO: Çok Boyutlu Madde Tepki Kuramı Olabilirlik Oran Testi Yöntemi

MULTI-MIMIC: Multi Multiple Indicators Multiple Causes Model (Çok Boyutlu Çoklu Nedenler Çoklu Göstergeler)

MULTISIB: Multidimensional Simultaneously Item Bias Test (Çok Boyutlu Eş Zamanlı Madde Yanlılığı Testi)

NNFI: Non-Normed Fit Index (Norm Dışı Veriler İçin Uyum İndeksi)

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

RAM: Raju Area Measurement (Raju'nun Alan İndeksi)

RMSE: Root mean Square Error (Hataların Kareleri Ortalamalarının Karekökü)

RMSEA: Root mean Square Error of Approximation (Tahmini Hataların Kareleri Ortalamalarının Karekökü)

RMSR: Root Mean Square Residual (Artık Değerlerin Kareleri Ortalamalarının Karekökü)

SBM: Sınırlandırılmış Bilgi Modeli

SIBTEST: Simultaneously Item Bias Test (Eş Zamanlı Madde Yanlılığı Testi)

WRMSR: Weighted Root Mean Square Residual (Ağırlıklandırılmış Artık Değerlerin Kareleri Ortalamalarının Karekökü)

1 PLM: 1 Parametrelili Lojistik Model

2 PLM: 2 Parametrelili Lojistik Model

3 PLM: 3 Parametrelili Lojistik Model

Bölüm 1

Giriş

Bu bölümde problem durumu, araştırmanın amacı ve önemi, problem cümlesi, sayıtlar ve sınırlamalar üzerinde durulmuştur.

Problem Durumu

Eğitim-öğretimde bireylerde hedeflenen gelişimin adil bir şekilde gözlenmesi, eksik öğrenmelerin tespit edilmesi, bireylerin başarı düzeylerinin saptanması ancak tarafsız ve hatalardan arınık bir ölçme değerlendirme süreci ile mümkün olabilir. Ölçme ve değerlendirme sürecinin niteliği ne kadar yüksek olursa birey için verilecek geri bildirimler de o kadar doğru ve faydalı olacaktır.

Testlerden elde edilen sonuçlar bireyin ölçülen özelliğe sahip olup olmadığını gösterir. Bireyler testlere sadece öğrenme eksikliklerinin belirlenmesi için alınmazlar. Eğitim ve öğretim sürecinde izleme testleri olabileceği gibi seçme ve yerleştirme testleri de bulunmaktadır. Öğrenciler okullarını bitirip bir sonraki eğitim aşamasına geçerken seçme ve yerleştirme sınavlarına tabi tutulabilirler. Zaman zaman bu sınavlar personel seçimleri, görev değişiklikleri ve atamalar için işveren tarafından kurumlara yaptırılabilir veya direkt olarak devlet eliyle yapılabilir. Bu sebeplerden dolayı sınavlar hayatımızda çok önemli bir konuma sahiptir.

Eğitimde kullanılan testlerin amacına uygun olarak yapılabilmesi o testin geçerliğini gösterir. Bir test bireyin ölçmek istediği davranışlarını ne kadar yüksek düzeyde ve amacına uygun olarak ölçüyorsa, o testin o kadar geçerli olduğu söylenebilir (Mellor, 1995). Bir testin geçerli olmasının yanında, tesadüfi hatalardan arınık olması da testin güvenilirliği olarak tanımlanmaktadır. Adil bir ölçme ve değerlendirme süreci ancak amacına uygun ve hazırlanma, uygulama ve puanlama sürecinde hataların minimize edildiği testler ile gerçekleştirilebilir. Bir testin nitelikli olması testin yeterli seviyede geçerli ve güvenilir olmasıyla sağlanabilir (Horst, 1966). Testlerin, bireyler özelinde olduğu kadar toplumsal ve siyasi olarak da etkili sonuçları olduğu (Clauser ve Mazor, 1988) düşünülürse, testlerin niteliğinin yüksek olması gerektiği anlaşılmaktadır.

Sınava katılan grupların özellikleri de testin sonuçlarını etkileyebilmektedir. Bu özelliklere cinsiyet, dil, ırk, sosyo kültürel farklılıklar örnek olarak verilebilir

(Zenisky, Hambleton ve Robin, 2003; Engelhard,1990; Harris ve Carlton 1993). Birey ve grupların, sahip olduğu bu özelliklerden dolayı testlerden aldıkları puanlar farklılaşabilmektedir. Diğer bir deyişle test ile ölçülmek istenilen özellikler bilinçli veya bilinçsiz olarak sınavı alanların özelliklerinden etkilenerek hataya sebep olmaktadır (Romberg, Zarinnia ve Collins, 1990; Shepard, 1989).

Testi alan kişiler aynı yetenek düzeyine sahipse test sonuçlarının da aynı olması beklenir. Ancak kişinin ölçülmek istenen özelliğine etki eden farklılıkları kendisine avantaj veya dezavantaj sağlaması yanlılığa sebep olmaktadır. Yanlılık, sistematik hata olarak görülmekte ve geçerliğe zarar vermektedir (Zumbo,1999; Camilli ve Shepard, 1994; Cole ve Moss, 1989; Hambleton ve Rogers,1989; Adam ve Rowe 1988; Tittle, 1988). Madde yanlılığından söz edebilmek için maddelerin Değişen Madde Fonksiyonuna (DMF) sahip olup olmadığı tespit edilmelidir.

Değişen Madde Fonksiyonu (DMF) belirleme teknikleri istatistiksel bir süreçtir. DMF'nin var olup olmadığı belirlenirken, yetenek düzeylerinin eşit olduğu varsayılan grupların, cinsiyet, dil, kültürel özellikler gibi değişkenler ile oluşturulmuş odak ve referans gruplarında, maddelere verdikleri cevapların dağılımı dikkate alınmaktadır (Steinberg ve Thissen, 2006; Cohen ve Bolt, 2005; Hambleton, Swaminathan ve Rogers, 1991; Mellenberg, 1989; Adams ve Rowe, 1988). Eğer bu iki grup arasında fark varsa o maddenin değişen madde fonksiyonuna sahip olduğu anlamına gelmektedir. Değişen madde fonksiyonunun varlığını tespit etmek iki temel amaca dayanır. Bunlardan ilki, test puanlarının alt grupların özelliklerinden dolayı oluşan değişkenlik kaynaklarından etkilenip etkilenmediği, ikincisi ise bu özelliklerin avantaja dönüşüp dönüşmediğinin saptanmasıdır (Crocker ve Algina, 1986).

DMF içeren maddeler yanlılığa sebep olabileceğinden, olası yanlı maddeler olarak da tanımlanmaktadır (Mellenberg, 1983) . Ancak son olarak yanlılığın tespiti için uzmanların görüşüne başvurulması gerekmektedir. Yanlı madde içeren testleri yorumlamak, anlamlandırmak ve sonucuna göre kararlar almak hatalara sebep olmakta ve adillik kavramına zarar vermektedir. Bu nedenle DMF ve dolayısı ile yanlılık içermesi muhtemel maddelerin tespiti doğru bir şekilde yapılmalıdır. Yanlılık çalışmaları 20. Yüzyıl başlarında Alfred Binet'in düşük sosyo- ekonomik düzeylerden çocukları zekâ testine tabi tutmasıyla başlamıştır (Camilli ve Shepard, 1994; Harold, 1989). Binet anadili İngilizce olmayan, düşük ekonomik ve sosyal

imkânlarla sahip çocuklarda, bazı maddelerin zihinsel potansiyelden daha çok bu çocukların evde veya okuldaki kültürel eğitim-öğretimlerinin sonucu ile alakalı olduğunu belirlemiştir. Bu sebeple bu farklılıklara sebep olan maddeleri testten çıkarmayı uygun görmüştür. Ülkemizde gerçekleştirilen farklı sınavlarda da değişen madde fonksiyonunu ya da madde yanlılığını inceleyen birçok çalışmanın alan yazında bulunduğu görülmektedir (Öğretmen ve Doğan, 2004; Acar, 2008; Anıl ve Güzeller, 2010; Anıl, Güzeller, Çokluk ve Şekercioğlu, 2010; Kelecioğlu, Atalay ve Öztürk, 2010; Doğan ve Sevindik, 2011; Çepni, 2011; Karakaya ve Kutlu, 2012; Kan, Sünbül ve Ömür, 2013; Arslan 2014).

Birçok çalışmada DMF belirlenirken verilerin tek boyutlu olduğu varsayımında bulunmaktadır (Snow ve Oshima, 2007). Test ile tek bir örtük özelliğin ölçülüyor olması tek boyutluluk olarak tanımlanır ancak bu her testte mümkün değildir (Hambleton ve Swaminathan, 1985; Embretson ve Reise, 2000). Testlerin yapısı incelendiğinde farklı örtük özellikleri ölçen farklı alt testler ve madde gruplarından oluşabildiği belirlenmiştir (Snow ve Oshima, 2007). Birçok eğitimsel ve psikolojik test çok boyutlu olarak hazırlanmaktadır (Ackerman, Gierl ve Walker, 2003). Çok boyutlu testler birden fazla özelliği ölçmeyi hedefleyen testlerdir. Bu testlere, başarı testleri, psikolojik testler, yetenek testleri örnek olarak gösterilebilir. Önceki çalışmalar tek boyutluluk varsayımı göz ardı edilmesinin MTK parametre tahminlerinde ciddi etkiye neden olduğunu göstermektedir (Ackerman, 1992; Oshima ve Miller, 1992; Bolt ve Stout, 1996; Mazor, Hambleton ve Clauser, 1998; Walker ve Beretvas, 2003; Kim ve Jang, 2009; Paek ve Cai, 2014). Çok boyutlu testlerde tek boyutluluk varsayımı hatalı sonuçlara ve değerlendirmelere sebep olabilir. Bu nedenle çok boyutlu olduğu belirlenen testlerde yapılan analizlerin de testin boyut sayısını dikkate alması daha doğru sonuçlar elde etmek açısından önem arz etmektedir.

Araştırmanın Amacı ve Önemi

Sınıf içi testlerden uluslararası büyük ölçekli sınavlara kadar, uygulamalardan elde edilen ölçme sonuçları eğitimin bütün paydaşlarını yakından ilgilendirmektedir. Test sonuçlarına bakılarak alınan kararların önemi düşünüldüğünde, cevaplayıcıların yeteneklerini mümkün olduğunca yansız bir şekilde ölçme gerekliliği ortaya çıkmaktadır. Sonuçlarına dayanarak seçme, yerleştirme,

mezuniyet, işe alma gibi önemli kararların verildiği sınavların adil olabilmesi için sınavı oluşturan maddelerin DMF gösterip göstermediği belirlenmelidir.

DMF belirleme uygulamaları genellikle testin boyut sayısını dikkate almadan kullanılmaktadır. Bu durum hataların artmasına sebep olmaktadır. Ayrıca DMF belirleme yöntemleri de testin boyut sayısına göre değişmektedir. Bu nedenle DMF belirlemek için kullanılan yöntemlerin sadece tek boyutluluk varsayımı altında değil, testin boyut sayısına da duyarlı olacak şekilde yapılmasının gerekliliği ortaya çıkmaktadır.

Testin boyut sayısının yanında basit veya karmaşık yapıda olması, farklı örneklem büyüklükleri, odak ve referans grup oranları ve DMF etki büyüklüklerinin DMF belirleme süreçlerine etkisi de belirlenmek istenmektedir. Ayrıca DMF analizlerini SIBTEST ve çok boyutlu çoklu nedenler çoklu göstergeler yöntemi (Multi-MIMIC) yöntemi ile gerçekleştirmenin bulgulara nasıl yansıtacağı, 1. tip hataları ve güç oranlarını nasıl etkileyeceği de merak konusudur. Bu sorulara cevap bulmak için 100 tekrarlı simülasyon veri seti ile gerçek veri seti birlikte kullanılacaktır. Farklı yetenek düzeylerini bir arada bulunduran çok boyutlu testlerde, yetenek düzeylerinin DMF'ye sebep olabileceğinin tartışıldığı durumlara cevaben, çok boyutlu model ile yapılan DMF analizi neticesinde çok boyutluluğun bu farklılaşmaya etkisinin nasıl olduğu ve neticesinde 1. tip hataların ve istatistiksel güç oranlarının değişiminin belirlenmesi de çalışmanın amaçlarından biridir. Gerçek veri setinde de örneklem büyüklüğü, odak ve referans grup oranı ve farklı DMF belirleme tekniklerinden elde edilen DMF'li madde sayılarının karşılaştırılması amaçlanmaktadır.

Alan yazındaki benzer çalışmalar incelendiğinde çok boyutluluğu dikkate alan ve testin yapısına uygun model kullanan çalışmalar olsa da çok boyutlu yapılara hem klasik hem de çok boyutlu yöntemi uygulayan çalışmalar oldukça azdır. Ayrıca aynı çalışma içerisinde gerçek veri setini muhteva eden çalışmalar da yine sınırlı kalmıştır. Bu çalışmada testin yapısının dikkate alınması ve doğru koşullara karar verebilmek için hem yapay hem gerçek veri setlerinin kullanılacak olmasının alan yazına katkıda bulunması hedeflenmektedir.

Araştırma bu yönüyle DMF belirlerken boyutluluğu dikkate almanın sonuçlarını gösterecek bir çalışma olması açısından önem taşımaktadır.

Araştırma Problemi

İki ve üç boyutlu testlerdeki maddeler, tek boyutlu ve çok boyutlu DMF belirleme teknikleri ile analiz edildiğinde 1. tip hata ve istatistiksel güç oranları DMF etki büyüklüğü düzeyine, örneklem büyüklüğüne, odak-referans grup oranlarına göre nasıl değişmektedir? İki boyutlu gerçek veri setindeki maddeler tek ve çok boyutlu DMF belirleme teknikleri ile analiz edildiğinde DMF'li madde sayıları nasıl değişmektedir?

Alt Problemler

1) İki boyutlu basit yapıları verilere iki boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, 1. tip hata ve istatistiksel güç değerleri,

a) DMF etki büyüklüğü orta düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

b) DMF etki büyüklüğü yüksek düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

2) İki boyutlu karmaşık yapıları verilere iki boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, 1. tip hata ve istatistiksel güç değerleri,

a) DMF etki büyüklüğü orta düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

b) DMF etki büyüklüğü yüksek düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

3) Üç boyutlu basit yapıları verilere üç boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, 1. tip hata ve istatistiksel güç değerleri,

a) DMF etki büyüklüğü orta düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

b) DMF etki büyüklüğü yüksek düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

4) Üç boyutlu karmaşık yapıları verilere üç boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, 1. tip hata ve istatistiksel güç değerleri,

a) DMF etki büyüklüğü orta düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak -referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

b) DMF etki büyüklüğü yüksek düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak -referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

5) Gerçek veri setinden elde edilen, iki boyutlu veri, tek boyutlu (SIBTEST) ve iki boyutlu olarak analiz edildiğinde (MULTI-MIMIC) öğrencilerin cinsiyeti açısından, farklı örneklem büyüklüklerinde (600, 1200, 3600), odak grup ve referans grupları oranlarında (1:1, 1:2) DMF gösteren madde sayısı nasıl değişmektedir?

Sınırlılıklar

Araştırma, kullanılan verilerin yapısı, koşulları, analiz için kullanılan yöntem ve uygulamalar ile sınırlıdır.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Klasik Test Kuramı

Sınavların hayatımızda bu kadar önemli rol oynamasından kaynaklı olarak ölçme ve değerlendirme sürecinin en az hata ile gerçekleştirilmesi hedeflenmektedir. Bu sebeple ortaya çıkan ilk ve üzerinde yıllarca çalışılan teori klasik test teorisidir (KTK). KTK gözlenen puanlardan hareketle gerçek puanlara ulaşmayı hedefler. Bu sebeple bu kurama gerçek puan teorisi de denmektedir (Baykul, 2010). Gözlenen puan, ölçme işleminin gerçekleştirilmesinden sonra elde edilen ve içinde gerçek ve hata puanı bulunan puandır. Gerçek puan, ölçülen özelliğe ait ve hata puanı karışmamış puandır. Hata puanı ise her ölçme işleminde karşılaşılan ve kaynağı bilinmeyen puan türüdür. Gözlenen, gerçek ve hata puanları arasındaki ilişki aşağıdaki denklem ile ifade edilmektedir (Sperman 1904; Croker ve Algina, 1986; Baykul, 2010; Cohen ve Swerdik, 2013).

$$X_{Gözlenen} = X_{Gerçek} + X_{Hata}$$

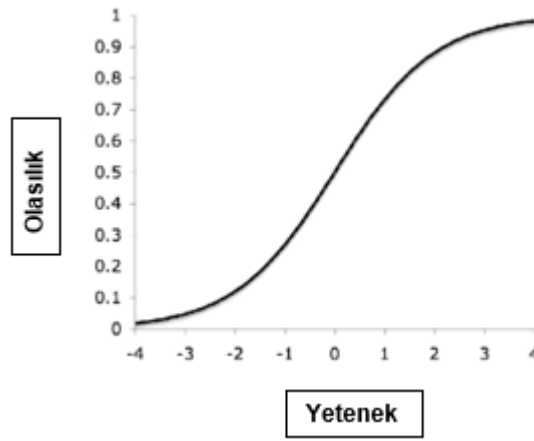
KTK' da temel varsayımlar aşağıda ifade edilmiştir (Gulliksen, 1950; Magnusson, 1968; Baykul, 2000).

- Beklenen hata puanları değeri sifıra yakınsar.
- Gerçek puanlar ile hata puanları arasında korelasyondan bahsedilemez.
- Farklı ölçme işlemleri neticesinde ortaya çıkan hata puanları arasında ilişki gözlenmez.

Bu varsayımların çok olması ve her testte karşılanmasındaki güçlük, madde ve test istatistiklerinin bireyden ziyade guruba bağımlılığı KTK' ya olan eleştirileri arttırmış ve neticesinde Madde Tepki Kuramı (MTK) geliştirilmeye başlanmıştır (Lord ve Novick, 1968; Lord, 1980; Crocker ve Algina, 1986; Embretson ve Reise, 2000).

Madde Tepki Kuramı

Madde Tepki Kuramı (MTK) 20. yüzyılın ikinci yarısından itibaren gelişmeye başlamıştır (Embretson ve Reise; 2000). MTK bireylerin yetenekleri ile ilgilenir. Yetenek parametresi θ (theta) simgesi ile sembolize edilir. Kuram bireyin maddelere verdiği yanıtlar ile yetenek parametresini kestirmeyi amaçlar. MTK ile bireyin gözlenemeyen yeteneği ve nitelikleri arasında matematiksel modellemeler yoluyla fonksiyonlar elde edilir. Madde puanının kestirimi ile elde edilen fonksiyon madde karakteristik eğrisi (MKE) olarak tanımlanır. MKE, bireyin başarılı olma ihtimalini test ile ölçülen özellik/yetenek ile ilişkilendiren matematiksel bir fonksiyondur (Crocker ve Algina, 1986). MKE' nin genel gösterimi Şekil 1'de gösterilmiştir.



Şekil 1. Madde karakteristik eğrisi

Şekil 1'de yatay eksen ($y=0$) bireylerin yetenek düzeylerini ifade ederken, dikey eksen ($x=0$) bu yetenek düzeylerindeki bireylerin ölçülen özelliğe sahip olma (soruyu doğru yanıtlama) düzeylerini göstermektedir. Bu eğri fonksiyonda yer alan parametrelere ve modellere göre değişkenlik göstermektedir.

Ancak bu modellerin kurulması ve olası doğru ilişkilerin ortaya konması için temel varsayımlar bulunmaktadır (Hambleton ve Swaminathan, 1985; Embretson ve Reise, 2000). Bu varsayımlar tek boyutluluk ve yerel bağımsızlıktır.

Tek Boyutluluk. Bireyin sadece bir özelliğinin ölçülmesi durumudur. Yani bu özelliğe sahip ölçme araçları bireyin bir yeteneğini ölçmeye odaklanmaktadır (Lord ve Novick, 1968; Hambleton ve Swaminathan, 1985; Crocker ve Algina, 1986; Embretson ve Reise, 2000). Örneğin; matematik başarı testinde ölçülen cebirsel işlem becerisine, okuduğunu anlama becerisi karışlıyorsa bu testi tek boyutlu

olmaktan çıkarmaktadır. Tek boyutluluk durumunun incelenmesi için verilere faktör analizi yapılmaktadır.

Yerel Bağımsızlık. Benzer düzeyde yeteneğe sahip bireylerin, farklı test maddelerine verdikleri cevapların birbirini etkilememesi durumudur (Hambleton ve Swaminathan, 1985; Reckase, 2000). Yerel bağımsızlık varsayımı için tek boyutluluk varsayımı ön koşuldur. Yerel bağımsızlıkta maddeler arası ilişki yetenek parametresi koşulu altında incelendiğinden, bireylerin maddelere verdiği yanıtların doğru olma olasılığı olasılıktaki bağımsızlık ilkesi gereği çarpımlar ile aşağıdaki gibi incelenmektedir.

$$P(U_1 U_2 U_3 \dots U_n | \theta) = P(U_1 | \theta) P(U_2 | \theta) P(U_3 | \theta) \dots P(U_n | \theta) = \prod_{i=1}^n P(U_i | \theta)$$

U_i : Rastgele seçilmiş bireyin i . maddeye vereceği yanıt

$P(U_i | \theta)$: θ yetenek düzeyindeki bireyin yanıt olasılığı

$P(U_i = 1 | \theta)$: θ yetenek düzeyindeki bireyin doğru yanıt olasılığı

$P(U_i = 0 | \theta)$: θ yetenek düzeyindeki bireyin doğru yanıt olasılığı

MTK için en önemli varsayımlardan biri de hız testi olmaması gerekliliğidir. Bir öğrenci sınavda yeteri kadar süresi olmaması (kalmaması) sebebiyle soruya doğru yanıt veremez veya boş bırakmak zorunda kalırsa madde ve birey parametreleri kestirimi bundan olumsuz etkilenir. Çünkü ölçülmek istenen özelliğe hız değişkeni de karışmış ve tek boyutluluk önkoşulu zarar görmüştür. MTK tüm bu varsayımlar altında, bireylerin yeteneklerinin gözlenebilir olma durumu ve madde ve test yapılarından kaynaklanan çeşitlilikten dolayı farklı modeller altında incelenmektedir.

MTK Modelleri

MTK modelleri McDonald (1982) tarafından boyutluluğa göre tek boyutlu ve çok boyutlu olarak sınıflandırılmıştır.

1) Tek boyutlu modeller

- a) Bir parametrelili lojistik model (1PLM)-Rasch modeli
- b) İki parametrelili lojistik model (2 PLM)- Birnbaum modeli
- c) Üç parametrelili lojistik model (3 PLM)

2) Çok boyutlu modeller

a) Telafisel (compensatory) model

b) Telafisel olmayan (noncompensatory) model

Bu sınıflandırmanın yanı sıra Hambleton ve Swaminathan da (1985) iki kategorili olarak puanlanan verileri yapılarına göre Tablo 1'deki gibi sınıflandırmışlardır.

Tablo 1

Veri Yapısına göre MTK Modelleri

Verinin yapısı	Model	Referans
İki kategorili	Örtük doğrusal (latent linear)	Lzarsfeld ve Henry (1968)
	Mükemmel ölçek (perfect scale)	Guttman (1944)
	Örtük uzaklık (latent distance)	Lzarsfeld ve Henry (1968)
	Bir-iki-üç parametrelili normal eğrileri	Lord (1952)
Çok kategorili	Bir-iki-üç-dört parametrelili lojistik	Birnbaum (1968), Barton ve Lord (1981), McDonald (1982),
	Kategorik yanıt	Bock (1972)
	Dereceli yanıt	Samejima (1969)
Sürekli	Kısmi bilgi modeli (partial credit model)	Master (1982)
	Sürekli Yanıt (continous response)	Samejima (1972)

İki kategorili puanlanan veriler için en sık kullanılan MTK modelleri Lojistik modellerdir. Lojistik modeller 1 PLM, 2 PLM, 3 PLM ve 4 PLM olmak üzere, sıklıkla ilk üçü kullanılmaktadır. Aşağıda tek boyutlu model olarak da bilinen 1 PLM, 2 PLM ve 3 PLM anlatılmıştır.

1 Parametrelili Lojistik Model (1 PLM) ve Rasch Model. Şans başarısının olmadığı bu modelde her maddenin eşit ayırt ediciliğe sahip olduğu varsayılır. Bu modelde madde güçlük parametresi dikkate alınır ve yetenek kestirimleri gerçekleşir (Embretson ve Reise, 2000). 1 PLM ait olasılık dağılımı fonksiyonu aşağıdaki gibidir.

$$P(U_{is} = 1 | \theta_s, \beta_i) = \frac{e^{\alpha_i(\theta_s - \beta_i)}}{1 + e^{\alpha_i(\theta_s - \beta_i)}}$$

U_{is} =s bireyinin i. maddeye verdiği yanıt (1-0)

θ_s = s bireyinin yetenek seviyesi

β_i = maddenin güçlük parametresi

α_i = i maddesinin ayırt ediciliği

Rasch model 1 PLM den farklı olarak madde ayırt edicilik parametrelerinin 1' e eşit olduğu kabul edilmektedir. Rasch modeli aşağıdaki fonksiyon ile ifade edilmektedir.

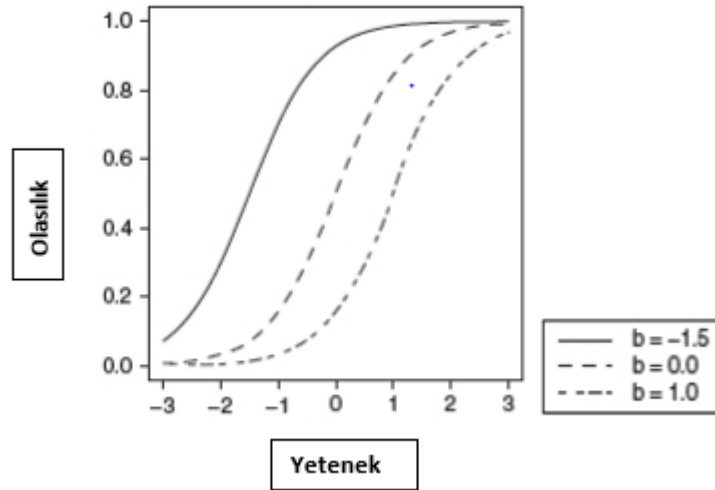
$$P(U_{is} = 1 | \theta_s, \beta_i) = \frac{e^{(\theta_s - \beta_i)}}{1 + e^{(\theta_s - \beta_i)}}$$

U_{is} =s bireyinin i. maddeye verdiği yanıt (1-0)

θ_s = s bireyinin yetenek seviyesi

β_i = maddenin güçlük parametresi

Rasch modelin MKE'si Şekil 2'de verilmiştir. Madde güçlük parametresinin değişimine göre eğrilerin konumu da değişmektedir.



Şekil 2. 1 PLM için madde karakteristik eğrisi

2 Parametrelili Lojistik Model (2 PLM). Bimbaum modeli olarak da bilinen bu modelde 1 PLM' den farklı olarak ayırt edicilik parametresi de eklenmiştir. Bu değer teoride $(-\infty, +\infty)$ aralığında değerler olsa da kullanımda 0 ile +2 arasında değerler aldığı gözlenmektedir. Ayırıcılıkları yüksek maddelerin MKE' lerinin dik olduğu bilinmektedir (Linden ve Hambleton, 1997; Harvey ve Hammer, 1999). 2 PLM, 3 PLM' nin özel bir halidir. 3 PLM' deki şans parametresi burada olmadığı için 2 PLM

de şans parametresinin 0 olduğu varsayılmaktadır (Lee, 2007; Seungho Yang, 2007). 2 PLM aşağıdaki fonksiyon ile ifade edilmektedir.

$$P(U_{is} = 1 | \theta_s, \beta_i, \alpha_i) = \frac{e^{\alpha_i(\theta_s - \beta_i)}}{1 + e^{\alpha_i(\theta_s - \beta_i)}}$$

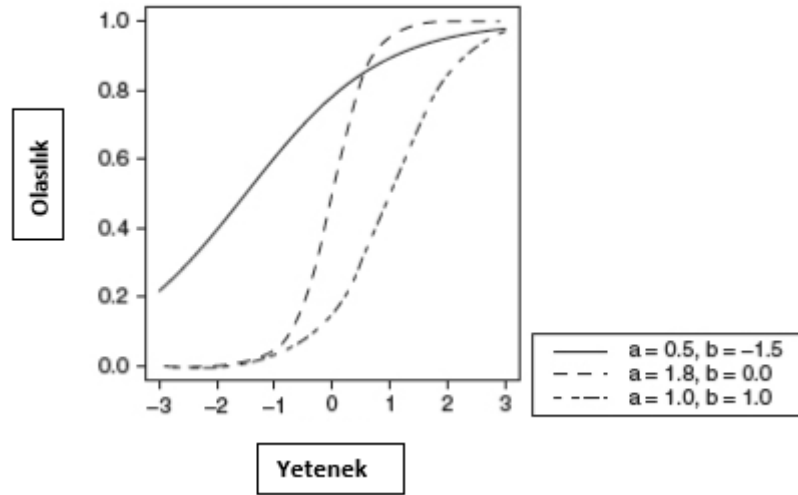
U_{is} =s bireyinin i. maddeye verdiği yanıt (1-0)

θ_s = s bireyinin yetenek seviyesi

β_i = maddenin güçlük parametresi

α_i = i maddesinin ayırt ediciliği

2 PLM' nin MKE' si Şekil 3'te verilmiştir. Madde ayırt edicilik ve güçlük parametrelerinin değişimine göre eğrilerin konumu da değişmektedir.



Şekil 3. 2 PLM için madde karakteristik eğrisi

3 Parametrelili Lojistik Model (3 PLM). 2 PLM 'den farkı şans parametresine sahip olmasıdır. Şans parametresi; düşük yetenek düzeyine sahip bireylerin maddeyi doğru cevaplama olasılığıdır. Bireyin ölçülen özelliğe sahip olmasa bile maddeyi doğru yanıtlandırma olasılığı bulunduğundan dolayı, tüm yetenek düzeylerinde sıfırdan yüksek olduğu kabul edilmektedir (Lord, 1980). 3 PLM aşağıdaki fonksiyon ile ifade edilmektedir.

$$P(U_{is} = 1 | \theta_s, \beta_i, \alpha_i, c_i) = c_i + (1 - c_i) \frac{e^{\alpha_i(\theta_s - \beta_i)}}{1 + e^{\alpha_i(\theta_s - \beta_i)}}$$

U_{is} =s bireyinin i. maddeye verdiği yanıt (1-0)

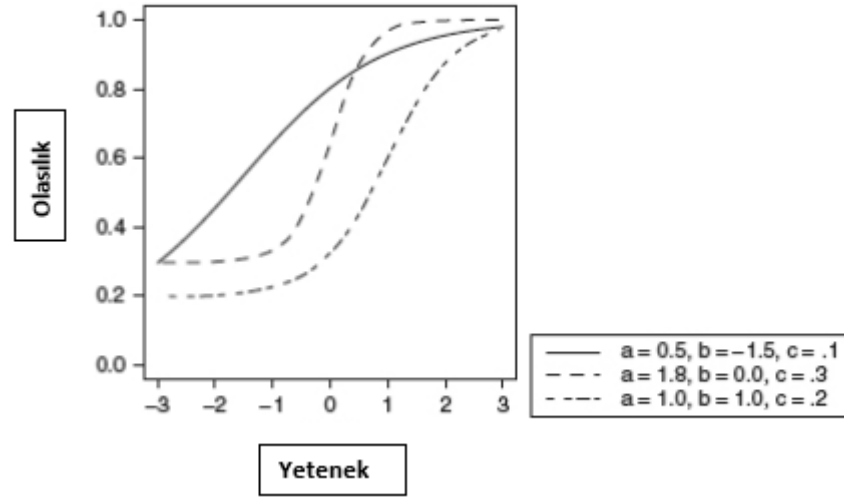
θ_s = s bireyinin yetenek seviyesi

β_i = maddenin güçlük parametresi

α_i = i maddesinin ayırt ediciliği

c_i = şans parametresi

3 PLM' nin MKE' si Şekil 4'te verilmiştir. Madde ayırt edicilik, güçlük ve şans parametrelerinin değişimine göre eğrilerin konumu da değişmektedir.



Şekil 4. 3 PLM için madde karakteristik eğrisi

Çok Boyutlu MTK Modelleri (ÇBMTK)

Tek Boyutluluk varsayımlarının karşılanması birçok kez mümkün olmamaktadır. Bu nedenle farklı örtük özellikleri ölçen çok boyutlu testler için daha uygun modeller ile çalışılmalıdır (Ackerman, Gierl ve Walker, 2003; Reckease, 2007). Bu modellere çok boyutlu MTK modelleri de (ÇBMTK) denmektedir. ÇBMTK MTK modellerindeki tek boyutluluk varsayımının sağlanmadığına yönelik eleştirilerin artmasıyla ortaya çıkmış ve geliştirilmiştir. Tek boyutlu MTK, örtük bir özellik ile bireyin soruyu doğru yanıtlama olasılığı arasındaki matematiksel ilişki modelini kurmak için kullanılırken, ÇBMTK bu modelleme sürecine birden fazla örtük özelliği veya yapıyı dâhil edebilir (Zhang, Xie ve Li, 2010). Tek boyutlu MTK tek bir yeteneğin fonksiyonu iken ÇBMTK' da çok sayıda yetenek aynı fonksiyonda ifade edilebilir. Bu nedenle ÇBMTK ile sade bir puan yerine karma yetenek görüntüsü elde edilir (Hartig ve Höhler, 2009).

ÇBMTK modelleri kullanmak için verinin çok boyutluluk özelliği taşımasının yanında yerel bağımsızlık ve monotonluk özelliklerine de ihtiyaç duyulmaktadır.

Yerel bağımsızlık tek boyutluluktaki ile aynıdır. Monotonluk ise bireyin yeteneğinin artması ile doğru cevap verme olasılığının artması veya bireyin yeteneğinin azalmasıyla doğru cevap verme olasılığının azalmasını ifade etmektedir. Buna göre olasılık dağılım fonksiyonu ile yetenek parametrelerini temsil eden θ vektörleri arasında pozitif değişimli bir ilişki vardır (Reckase, 2009). ÇBMTK iki kategorili puanlanan veriler (1-0) için aşağıdaki gibi tanımlanmaktadır.

$$U = (u_1, u_2, u_3, \dots, u_n), \quad R_j = (u_{1j}, u_{2j}, u_{3j}, \dots, u_{lj})$$

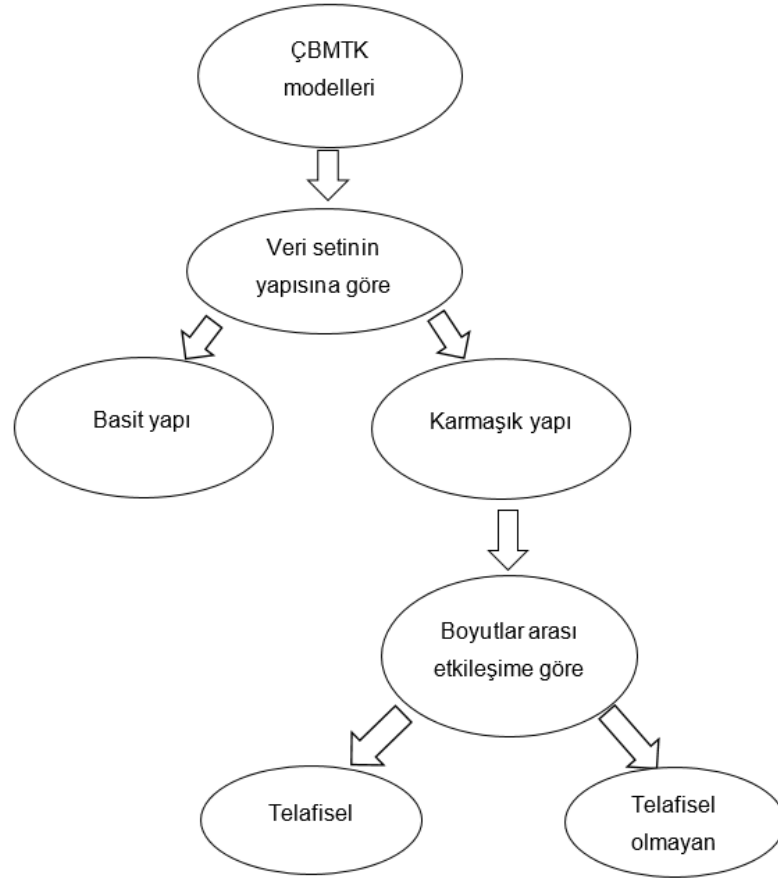
$$P_{i=}(\theta_{1j}, \theta_{2j}, \theta_{3j}, \dots, \theta_{nj}) = P(r_{1j} = 1 | \theta_{1j}, \theta_{2j}, \theta_{3j}, \dots, \theta_{nj}) = P(\theta_j)$$

l =madde sayısını

U_j = yanıt veren j . bireyin yanıt vektörünü

n = yanıt veren birey sayısını göstermektedir.

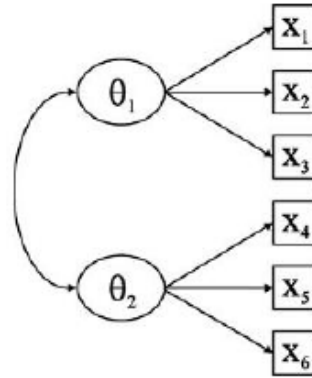
ÇBMTK modellerinin veri setinin yapısına göre sınıflandırılması Şekil 5'te verilmiştir (Hartig ve Höhler, 2009).



Şekil 5. Veri setinin yapısına göre ÇBMTK modelleri

Basit Yapı

Testi oluşturan alt testlerin her birinin tek bir yeteneği ölçtüğü ve kendi içinde homojen olduğu yapı basit yapı olarak adlandırılır. Genellikle geniş ölçekli sınavlarda kullanılır. Örneğin matematik yapısını ölçmeyi hedefleyen bir testte cebir, sayılar, geometri gibi her biri kendi içinde farklı bilişsel süreçleri ölçen alt testler bulunur (Sheng ve Wikle, 2007; Bulut, 2013). Basit yapıyı gösteren şema Şekil 6'da (Hartig ve Höhler, 2009) verilmiştir.



Şekil 6. Basit yapı

Basit yapıda yer alan alt testler ile belirlenmiş özellikler maddeler aracılığı ile ölçülebilir. Bu alt testler tek boyutlu iken, testin bütününe çok boyutlu olduğu varsayılır. Bu sebeple bu modele çoklu-tek boyutlu model (multi-unidimensional) denmektedir. Basit yapıda örnek parametreler Tablo 2'de verilmiştir.

Tablo 2

Basit yapıda parametreler

Alt test (i)	Madde (j)	α_{1j}	α_{2j}	β_{ij}	c_{ij}
1	1	α_{11}	0	β_{11}	γ_{11}
1	2	α_{12}	0	β_{12}	γ_{12}
1	3	α_{13}	0	β_{13}	γ_{13}
2	1	0	α_{24}	β_{24}	γ_{24}
2	2	0	α_{25}	β_{25}	γ_{25}
2	3	0	α_{26}	β_{26}	γ_{26}

Tablo 2'de verilen parametre değerleri ile i bireyinin s alt testinde bulunan j maddesini doğru yanıtlama olasılığı aşağıdaki eşitlik ile hesaplanır.

$$P(U_{sij} = 1 | \theta_{si}, \alpha_{sj}, \beta_{sj}, c_{sj}) = c_{sj} + \frac{1 - c_{sj}}{1 + e^{[\sum_{i=1}^s \theta_{ji} \beta_{ji}] + \alpha_{sj}}}$$

U_{sij} = i bireyinin s alt testinin j maddesine verdiği yanıt (1-0)

θ_{si} = i bireyinin s alt testindeki yetenek düzeyi

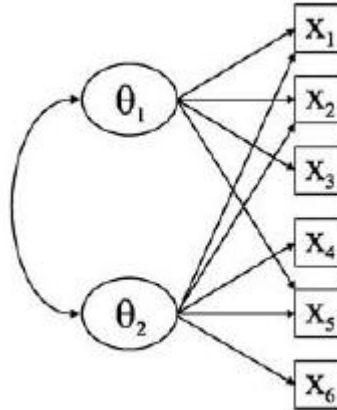
α_{sj} = s alt testindeki j maddesinin ayırt ediciliği

β_{sj} = s alt testindeki j maddesinin güçlük düzeyi

c_{sj} = s alt testindeki j maddesinin şans parametresi

Karmaşık Yapı

Maddeler içi model veya basit olmayan yapı olarak adlandırılan bu yapıda maddeler birden fazla yetenek ile matematiksel ilişkiye sahip olabilir (Bulut, 2013). Karmaşık yapıyı gösteren şema Şekil 7'de (Hartig ve Höhler, 2009) verilmiştir.



Şekil 7. Karmaşık yapı

Maddeler içi modelde boyutları oluşturan maddelerin doğru cevaplanma olasılıklarının birbirinden etkilenip etkilenmemesi durumuna, yani boyutlar arası etkileşime göre Telafisel (compensatory) ve Telafisel olmayan (noncompensatory) modeller bulunur (Ackerman, 1996). Telafisel modelde bir yetenek düzeyindeki eksiklik diğer yetenek düzeylerindeki yetkinlik ile telafi olabilirken, telafisel olmayan modelde, ölçülen yeteneklerdeki eksiklik diğer yetenekler ile telafi edilemez. Telafisel modelde, bireylerin düşük yetenek düzeyleri telafi edildiği için maddeyi doğru yanıtlama olasılığı artarken, telafisel olmayan modelde bireyin maddeyi doğru cevaplama olasılığı, farklı yetenek olasılıklarının çarpımına eşit olmaktadır

(Ackerman, Gierl ve Walker, 2003). Bu modeller de tek boyutlu MTK' da olduğu gibi parametre sayılarına göre modellere ayrılır (Reckase, 2009).

Telafisel Modeller

- a) ÇBMTK için genişletilmiş Rasch Modeli
- b) ÇBMTK için genişletilmiş 2 PLM
- c) ÇBMTK için genişletilmiş 3 PLM
- d) ÇBMTK için genişletilmiş normal ogive model

ÇBMTK için genişletilmiş Rasch modeli. Tek boyutlu Rasch modelinden farklı olarak yetenek parametresi tek bir yüzey yerine koordinatlar toplamının bileşkesi olarak aşağıdaki formül ifade edilir (McKinley ve Reckase, 1982).

$$P(U_{is} = 1 | \theta_s, d_i) = \frac{e^{\sum_{l=1}^m \theta_{sl} + d_i}}{1 + e^{\sum_{l=1}^m \theta_{sl} + d_i}}$$

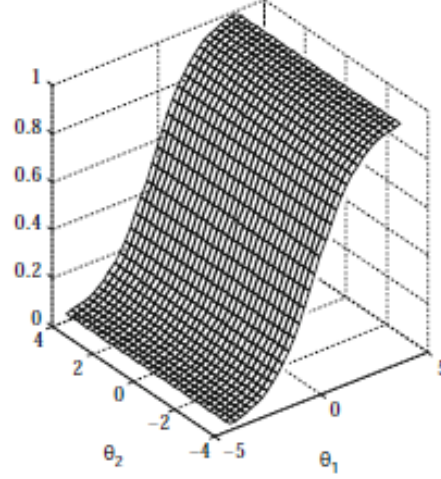
U_{is} =s bireyinin i. maddeye verdiği yanıt (1-0)

θ_s = s bireyinin yetenek seviyesi

θ_{sl} =s bireyinin l. boyuttaki yetenek seviyesi

d_i = i maddesinin kesim noktası

Modelde maddelerin eşit güçlükte olduğu kabul edilmektedir. Yukarıdaki eşitlikte yer alan d parametresi tek boyutlu modellerdeki $(\alpha(\theta - \beta))$ ifadesinin açılımıyla elde edilen $\alpha\theta - \alpha\beta$ ifadesinde, $-\alpha\beta$ yerine d yazılmasıyla bulunur. $\alpha=1$ olduğu varsayılır. Bu modele ait madde karakteristik yüzeyi (MKY) $a_i = [1 \ 0]$ Şekil 8'de verilmiştir (Reckase, 2009).



Şekil 8. İki boyutlu Rasch model MKY

ÇBMTK için genişletilmiş 2 PLM. Bu modelde 2 PLM' de s bireyinin i maddesine doğru yanıt verme olasılığı aşağıdaki fonksiyonla ifade edilmiştir.

$$P(U_{is} = 1 | \theta_s, d_i, \alpha_i) = \frac{e^{\sum_{l=1}^m \alpha_{il} \theta_{sl} + d_i}}{1 + e^{\sum_{l=1}^m \alpha_{il} \theta_{sl} + d_i}}$$

U_{is} =s bireyinin i. maddeye verdiği yanıt (1-0)

θ_s = s bireyinin yetenek seviyesi

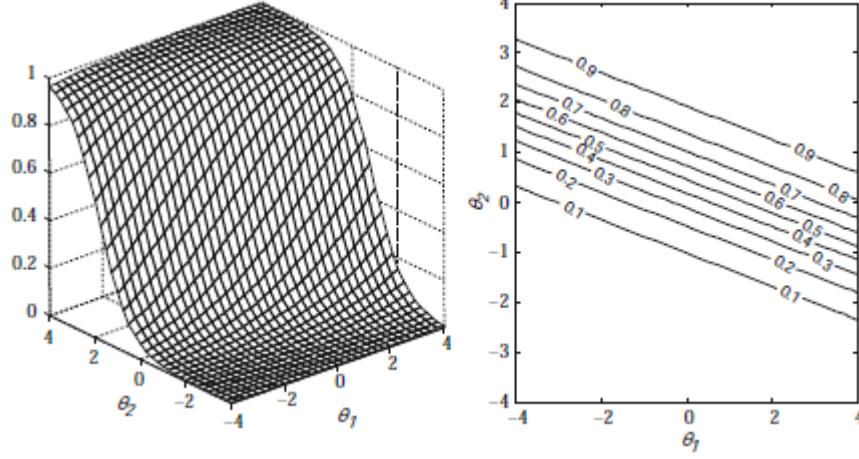
d_i = maddenin güçlük parametresi

α_i = i maddesinin ayırt ediciliği

α_{il} =l. boyuttaki i maddesinin ayırt ediciliği

θ_{sl} =s bireyinin l. boyuttaki yetenek seviyesi

Bu modele $\alpha_i = [0,5 \ 1,5]$ $d=-0,7$ ait MKY ve eş olasılık konturları Şekil 9'da verilmiştir (Reckase, 2009).



Şekil 9. Çok boyutlu 2 PLM için MKY ve eş olasılık konturu

ÇBMTK için genişletilmiş 3 PLM. Bu modelde 3 PLM' de s bireyinin i maddesine doğru yanıt verme olasılığı şans parametresi eklenmesiyle aşağıdaki fonksiyonla hesaplanır.

$$P(U_{is} = 1 | \theta_s, d_i, \alpha_i) = c_i + (1 - c_i) \frac{e^{\sum_{l=1}^m \alpha_{il} \theta_{sl} + d_i}}{1 + e^{\sum_{l=1}^m \alpha_{sl} \theta_{sl} + d_i}}$$

U_{is} =s bireyinin i. maddeye verdiği yanıt (1-0)

θ_s = s bireyinin yetenek seviyesi

d_i = maddenin güçlük parametresi

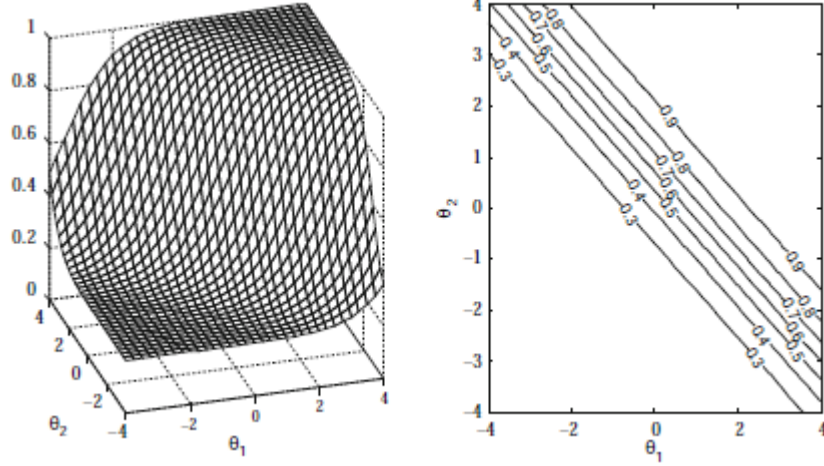
α_i = i maddesinin ayırt ediciliği

c_i = i maddesinin şans parametresi

α_{il} =l. boyuttaki i maddesinin ayırt ediciliği

θ_{sl} =s bireyinin l. boyuttaki yetenek seviyesi

Bu modele $\alpha_i = [1,3,1,4]$ $d=-1$ $c = 0,2$ ait MKY ve eş olasılık konturları Şekil 10'da verilmiştir (Reckase, 2009).



Şekil 10. Çok boyutlu 3 PLM için MKY ve eş olasılık konturu

ÇBMTK için genişletilmiş normal ogive model. Bu model s bireyinin i maddesinin doğru yanıt olma olasılığını farklı ağırlıklandırılmış yetenek düzeylerinin bileşkesini kullanarak hesaplar (Embretson and Reise, 2000). Şans parametresi içeren ve içermeyen formüller aşağıda verilmiştir (Bock ve Schilling, 2003).

$$P(U_{is} = 1 | \theta_s, d_i, \alpha_i) = \int_{-z_{is}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$$P(U_{is} = 1 | \theta_s, d_i, \alpha_i, c_i) = c_i + (1 - c_i) \int_{-z_{is}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

U_{is} =s bireyinin i. maddeye verdiği yanıt (1-0)

θ_s = s bireyinin yetenek seviyesi

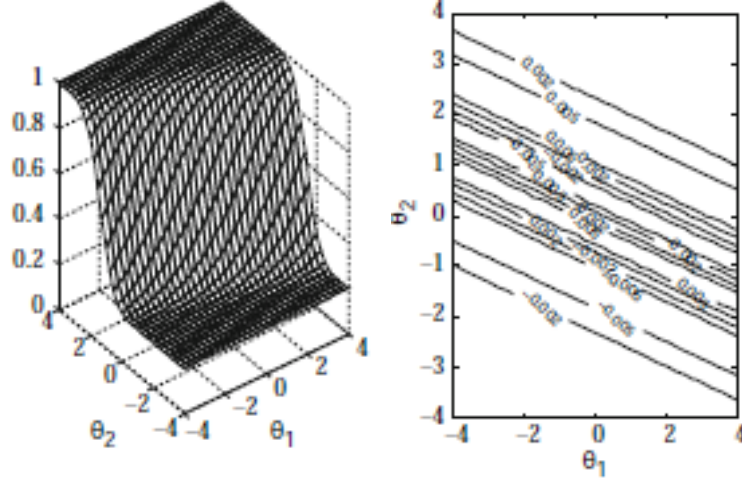
d_i = maddenin güçlük parametresi

α_i = i maddesinin ayırt ediciliği

c_i = i maddesinin şans parametresi

$z_{is} = \alpha_i \theta_s + d_i$

Bu modele $\alpha_i = [0,5 \ 1,5]$ $d=0$ $c = 0,2$ ait MKY ve eş olasılık konturları Şekil 11'de verilmiştir (Reckase, 2009).



Şekil 11. Çok boyutlu normal ogive model için MKY ve eş olasılık konturu

Telifisel Olmayan Modeller

Bu modelde maddeye ilişkin yetenekler arasında telifisel ilişki yoktur. Dolayısı ile maddeye doğru yanıt verme olasılığı tüm yetenekler için elde edilen olasılık değerlerinin çarpımına eşittir. Şans parametresi içeren ve içermeyen formüller aşağıda verilmiştir Sympson (1978).

$$P(U_{is} = 1 | \theta_s, d_i, \alpha_i) = \prod_{l=1}^m \frac{e^{1,7\alpha_{il}(\theta_{sl} + d_{il})}}{1 + e^{1,7\alpha_{il}(\theta_{sl} + d_{il})}}$$

$$P(U_{is} = 1 | \theta_s, d_i, \alpha_i, c_i) = c_i + (1 - c_i) \prod_{l=1}^m \frac{e^{1,7\alpha_{il}(\theta_{sl} + d_{il})}}{1 + e^{1,7\alpha_{il}(\theta_{sl} + d_{il})}}$$

U_{is} =s bireyinin i. maddeye verdiği yanıt (1-0)

θ_s = s bireyinin yetenek seviyesi

d_i = maddenin güçlük parametresi

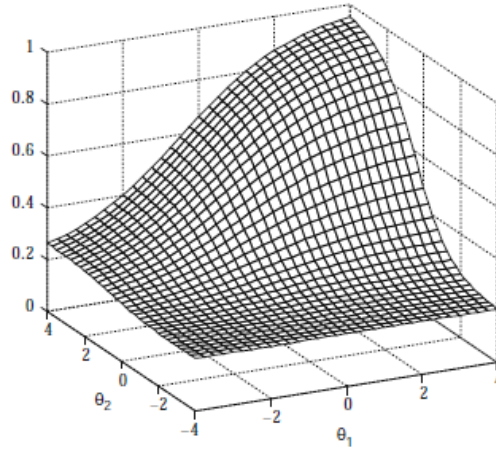
α_i = i maddesinin ayırt ediciliği

c_i = i maddesinin şans parametresi

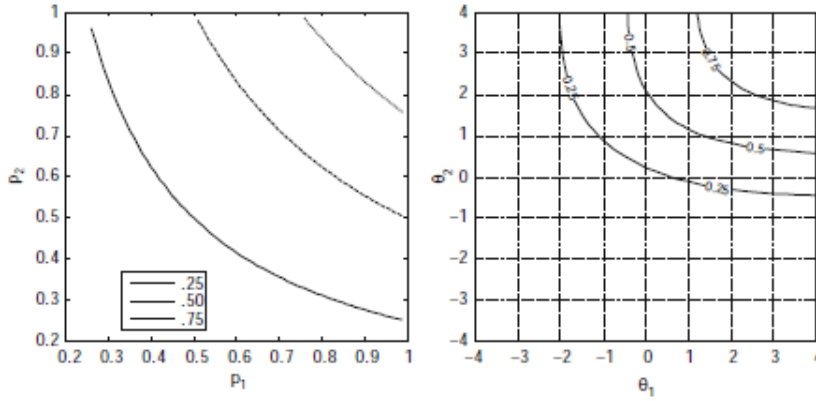
m = Toplam boyut sayısı

Telifisel olmayan modelde her bir madde için toplam boyut sayısı kadar α ve d parametresi kestirilir. Ancak şans parametresi telifisel modelde olduğu gibi bir tanedir (Reckase, 2009). Telifisel olmayan modele ait $a_i = [0,7 \ 1,1]$, $b_i = [-0,5 \ 0,5]$

ve $\gamma = 0,2$ ait MKY ve eş olasılık konturları Şekil 12'de ve Şekil 13'te verilmiştir (Reckase, 2009).



Şekil 12. Çok boyutlu telafisel olmayan modeller için MKY



Şekil 13. Çok boyutlu telafisel olmayan modeller için eş olasılık konturları

ÇBMTK' da Madde Parametreleri ve Vektörleri

Her madde için bu madde ile ilişkili yetenekleri ve eş olasılık konturlarını daha kolay yorumlamak amacıyla madde vektörleri belirlenir. Bir vektörün tanımlanması için, uzunluğunun, yönünün ve başlangıç noktasının bilinmesi gereklidir. ÇBMTK' da bu bileşenler ayırt edicilik, güçlük ve yön parametreleridir.

Madde vektörünün normu (uzunluğu) en büyük ayırt edicilik parametresi olarak tanımlanır. Çok boyutlu madde ayıricılığı (ÇBMA) aşağıdaki formül ile hesaplanır.

$$\text{ÇBMA} = \sqrt{\sum_{k=1}^m \alpha_k^2}$$

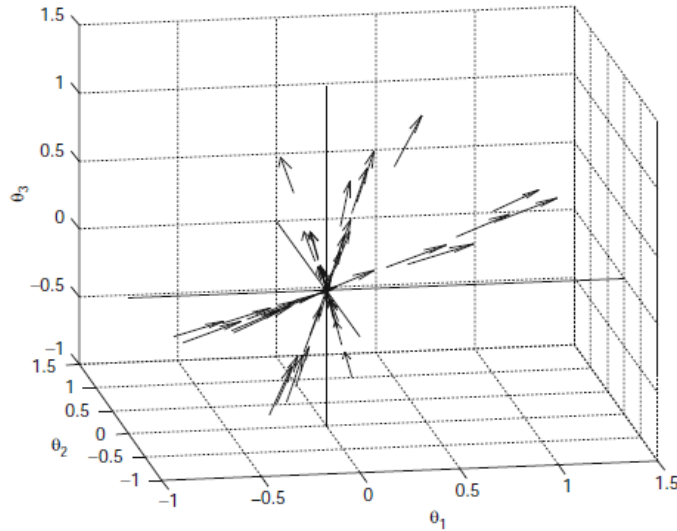
Madde güçlük parametresi tek boyutlu MTK' da olduğu gibidir. Madde vektörünün başlangıç noktası ile orijin arasındaki uzaklığı temsil eder. Çok boyutlu madde güçlük parametresi (ÇBMG) aşağıdaki formül ile hesaplanır. Buradaki d_i yer parametresidir.

$$\text{ÇBMG} = \frac{-d_i}{\text{ÇBMA}}$$

Son olarak madde vektörünün yönü için orijinden geçen ve θ_1 eksenine göre pozitif yönde açı yapan yer vektörü kullanılır. θ_1 Eksenine daha yakın olan maddeler θ_1 yeteneğini, θ_2 eksenine yakın olan maddelerin de θ_2 yeteneğini ölçer (Ackerman, 1994; Ackerman, Gierl ve Walker, 2003). Bir i maddesinin yönü aşağıdaki formül ile hesaplanır.

$$\alpha_i = \text{Arccos}\left(\frac{\alpha_{i1}}{\text{ÇBMA}_i}\right)$$

Kartezyen koordinat sisteminde birden fazla madde vektörü aynı anda çizilebilir. Aşağıdaki örnekte 45 maddenin üç boyutlu koordinat sisteminde nasıl gösterildiği ve kümelendiği Şekil 14'te verilmiştir.



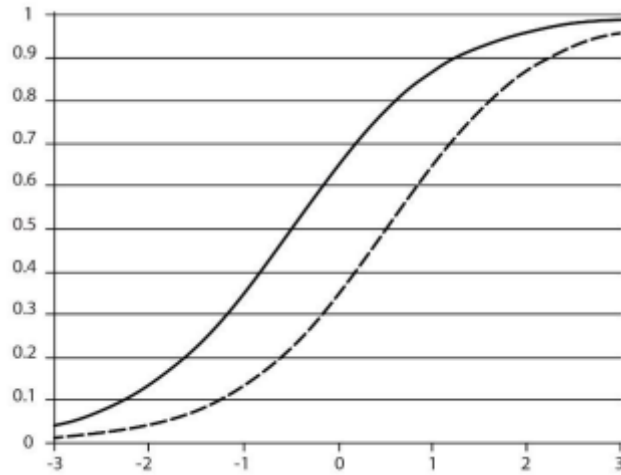
Şekil 14. Maddelerin üç boyutlu düzlemde gösterilmesi

Değişen Madde Fonksiyonu

Değişen Madde Fonksiyonu (DMF) yeteneğin kontrol edildiği durumlarda grupların maddeyi doğru cevaplandırma olasılıklarının değişmesi (farklılaşması)

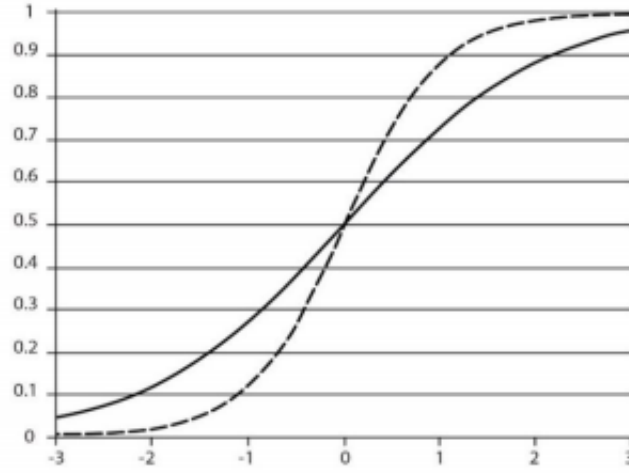
anlamına gelmektedir (Mellenberg, 1982; Clauser ve Mazor, 1988). DMF tek biçimli ve tek biçimli olmayan DMF olmak üzere iki şekilde ortaya çıkmaktadır.

Tek biçimli DMF, her bir yetenek düzeyinin gruplardan sadece birisi için avantaj sağladığı DMF türüdür. Tek biçimli DMF’de yetenek ölçeği boyunca DMF sadece bir grubun lehine gözlenmektedir ve madde karakteristik eğrileri birbirini kesmez. Madde ayırt ediciliklerin gruplar arasında değişim göstermediği tek biçimli DMF’de, madde güçlüğü, referans ve odak gruplar arasında farklılaşmaktadır (Li ve Stout, 1996; Zumbo, 1999; De Ayala 2002). Tek biçimli DMF’ye ilişkin MKE Şekil 15’te verilmiştir. Yatay eksen yetenek düzeyini gösterirken, dikey eksen maddeyi doğru yanıtama olasılıklarını göstermektedir. Şekil 15 incelendiğinde, tüm yetenek düzeylerinde 1. grubun (kesiksiz çizgi) maddeyi doğru yanıtama olasılıkları daha yüksektir.



Şekil 15. Tek biçimli DMF

Tek biçimli olmayan DMF ise maddenin bazı yetenek düzeylerinde bir gruba avantaj sağlarken, bazı yetenek düzeylerinde ise diğer gruba avantaj sağladığı durumdur. DMF tek biçimli olmaması durumunda grupların MKE’leri kesişmektedir. Maddeye ilişkin ayırt edicilik ve güçlük değerleri referans ve odak grup için farklılık göstermektedir (Zumbo, 1999; De Ayala, Kim, Stapleton ve Dayton, 2002). Tek biçimli DMF’ye ilişkin MKE Şekil 16’da verilmiştir. Yatay eksen yetenek düzeyini gösterirken, dikey eksen maddeyi doğru yanıtama olasılıklarını göstermektedir. Şekil 16 incelendiğinde, maddenin, yetenek düzeyi $\theta = 0$ oluncaya kadar 1. gruba (kesiksiz çizgi), daha yüksek yetenek düzeylerinde ise 2. gruba (kesikli çizgi) avantaj sağladığı görülmektedir.



Şekil 16. Tek biçimli olmayan DMF

Değişen Madde Fonksiyonu Belirleme Yöntemleri

DMF çalışmaları KTK çerçevesinde farklı yöntem ve sınıflandırmalar ile yapılmaya başlansa da (1970) Madde Tepki Kuramı'nın (MTK) ilerlemesi ile KTK ve MTK kaynaklı olmak üzere iki ana başlıkta incelenmiştir (Millsap ve Everson, 1993; Camilli ve Shepard, 1994). KTK 'ya göre Dönüştürülmüş Madde Güçlüğü (DMG), varyans analizi, Mantel-Haenszel (MH) (Mantel ve Haenszel, 1959; Holland, 1985), Lojistik Regresyon (LR) (Swaminathan ve Rogers, 1990; Zumbo, 1999) yöntemleri gözlenen puanlar üzerinden hesaplamalar yapılarak sıklıkla kullanılmaktadır.

MTK' ya dayalı olarak ise Simultaneous Item Bias Test (SIBTEST) (Shealy ve Stout, 1993, Stout ve Roses, 1995) Multidimensional Simultaneous Item Bias Test (MULTISIB), Raju'nun Alan İndeksi (RAM) (Raju,1988), Hierarchical Generalized Linear Model (HGLM), Lord' un Ki-kare testi (Lord's Chi-Square- LCS), (Lord, 1980) yöntemleri vardır. Ayrıca Delta Yöntemi, MTK Olabilirlik Oranı (MTK-OO) (Thissen, Steinberg ve Wainer, 1993; Thissen, 2001), Sınırlandırılmış Bilgi Modeli (SBM), yetenek farkları testi, çoklu nedenler çoklu göstergeler modeli (MIMIC), Multi-MIMIC ve madde karakteristik eğrisi gibi yöntemler kullanılmaktadır.

Bu yöntemlerden sık kullanılanları aşağıdaki alt başlıklarda açıklanmıştır.

Mantel-Haenszel Yöntemi. Mantel-Haenszel (MH), parametrik olmayan bir DMF belirleme tekniğidir. Mantel ve Haenszel (1959) tarafından geliştirilen yöntemde aynı yetenek seviyesine sahip olduğu varsayılan iki grubun, yanıtladıkları

soruların doğru ve yanlış yanıtlanma olasılıkları oranları arasındaki farklar dikkate alınır. Test puanlarından yola çıkılarak, benzer puanları alan bireylerden yetenek grupları oluşturulur. Bu yetenek düzeyleri için 2x2'lik çapraz tablolar oluşturulur (Gierl, 2000; Holland ve Thayer, 1988). Tablo 3'te referans ve odak gruplarında maddeyi doğru ve yanlış cevaplama sayıları ile toplam değerleri verilmiştir.

Tablo 3

MH Yönteminde Veri Düzeni

Grup	Doğru	Yanlış	Toplam
Referans	A_j	B_j	n_{rj}
Odak	C_j	D_j	n_{fj}
Toplam	m_{1j}	m_{0j}	T_j

Tablo 3'teki değerleri kullanarak α_{MH} (MH istatistiği) aşağıdaki eşitlik ile elde edilir.

$$\alpha_{MH} = \frac{\sum_j \frac{A_j D_j}{T_j}}{\sum_j \frac{B_j C_j}{T_j}}$$

Bu değer için kurulan hipotez, bu değer 1 olduğunu, karşı hipotez de 1 den farklı olduğunu varsayar. MH istatistiği; $\alpha_{MH} > 1$ ise referans grup lehine, $\alpha_{MH} < 1$ ise odak grup lehine maddeyi doğru cevaplama olasılığının daha yüksek olduğu anlamına gelir. MH istatistiğini yorumlamayı kolaylaştırmak için Camili ve Shepard (1994) tarafından $\Delta_{MH} = -2,35 \ln(\alpha_{MH})$ logaritmik dönüşüm uygulanması önerilmiştir. Bu değer 0 olması DMF'nin yokluğunu ifade eder. Dorans ve Holland (1993) tarafından Δ_{MH} için Tablo 4 kullanılarak DMF büyüklüğünün tespit edilebileceği öne sürülmüştür.

Tablo 4

MH Yöntemi ile DMF Düzeyi Belirlenmesi

Düzyey	Aralık	DMF Büyüklüğü
A	$ \Delta_{MH} < 1$	İhmal edilebilir
B	$1 \leq \Delta_{MH} < 1,5$	Orta düzey
C	$ \Delta_{MH} \geq 1,5$	Yüksek düzey

MH yöntemi, kolay yorumlanması, küçük gruplar için de kullanışlı olması ve MTK' daki yöntemler ile yakın sonuçlar vermesi sebebiyle sıklıkla tercih edilir. MH yöntemi için en büyük sınırlılık sadece tek biçimli DMF için kullanılmasıdır (Bertrand ve Boiteau, 2003).

Lord' un χ^2 (Ki-Kare) Yöntemi. Bu yöntem referans ve odak gruba ait madde parametrelerinin karşılaştırılması için, madde parametreleri farklarından oluşan varyans-kovaryans matrisi kullanılarak, χ^2 istatistiği ile tek biçimli ve tek biçimli olmayan DMF'yi belirleme yöntemlerinden biridir (Lord, 1980). Aynı ölçek düzeyine getirilmiş kestirilmiş parametreler kullanılarak aşağıdaki eşitlik ile hesaplanmaktadır.

$$\chi^2 = \frac{(d_{j1} - d_{j2})}{V_j}$$

d_{j1} =referans grubu için madde güçlük değeri

d_{j2} = odak grubu için madde güçlük değeri

V_j = güçlük değerleri varyansı

Lojistik Regresyon (LR). Kullanım kolaylığı açısından (hesaplama ve yorumlama) MH yöntemi ile benzer özelliklere sahip olsa da tek biçimli olmayan DMF'yi tespit edebilmesi ve çok kategorili verilerde de kullanılabilmesi yönüyle güçlü ve çok sık tercih edilen bir yöntemdir (Swaminathan ve Rogers, 1990). Adından da anlaşılacağı üzere bu yöntem bağımlı değişkenin kategorik olduğu durumlarda kullanılan bir tür regresyon analizidir. Bu regresyon yönteminde kategorik değişkeni yordamak için kategorik veya sürekli bağımsız değişkenlerin bulunduğu model

oluşturulur (Pedrajita ve Talisayon, 2009). Bir maddenin doğru yanıtlanma olasılığı lojistik regresyon modeli ile bulunur (Swaminathan ve Rogers, 1990).

$$P(U = 1) = \frac{e^z}{1 + e^z}$$

$$z = \tau_0 + \tau_1\theta + \tau_2g + \tau_3(\theta g)$$

U = Bireyin maddeye verdiği yanıt

θ = Bireyin yetenek düzeyi

g = Bireyin ait olduğu grup (referans-odak)

τ = Grup farklılığı

LR yöntemi ile model karşılaştırması yapılır. Bu sebeple yetenek, grup ve aralarındaki etkileşimden kaynaklanan değişkenlikler modele art arda eklenerek modelde farklılığa neden olup olmadığı incelenir. Bu aşamalı ekleme aşağıdaki eşitliklerdeki gibi olmaktadır (Gierl, 2000).

$$z = \tau_0 + \tau_1\theta$$

$$z = \tau_0 + \tau_1\theta + \tau_2g$$

$$z = \tau_0 + \tau_1\theta + \tau_2g + \tau_3(\theta g)$$

Başlangıç modelinde toplam puan bulunurken, sonraki modelde grup değişkeni, en son modelde ise etkileşim değişkeni yer almaktadır. Bu modeller ile regresyon analizi yapılırken ki-kare (χ^2) ve standartlaştırılmış regresyon katsayıları-etki büyüklüğü değeri (R^2) değerleri birlikte hesaplanır. R^2 değerinin büyüklüğüne göre DMF düzeyi ve büyüklüğü hakkında yorum yapılabilir. Bu sınıflama için aralıklar Tablo 5'te verilmiştir (Zumbo ve Thomas, 1996).

Tablo 5

LR Yöntemi ile DMF Düzeyi Belirlenmesi

Düzyey	Aralık	DMF Büyüklüğü
A	$\Delta R^2 < 0,13$	İhmal edilebilir
B	$0,13 \leq \Delta R^2 < 0,26$	Orta düzey
C	$\Delta R^2 \geq 0,26$	Yüksek düzey

Bir maddede DMF'nin varlığından söz edilebilmesi için grup ve etkileşim etkileri arasında yapılan ki-kare testinin anlamlı olması ve ΔR^2 değerinin en az 0,13 olması gereklidir. DMF tespitinin ardından tek biçimli olup olmadığını belirlemek için modeller arası R^2 farkları incelenir. Tek biçimli DMF için ilk ve ikinci model arasındaki, tek biçimli olmayan DMF için ikinci ve üçüncü model arasındaki R^2 farklar dikkate alınır (Jodoin ve Gierl 2001).

SIBTEST Yöntemi. Shealy ve Stout (1993) tarafından geliştirilen bu yöntemde iki kategorili verilerde DMF belirlemek amaçlanmıştır. Standartlaştırma yöntemi ile benzese de 1. tip hatayı kontrol edebilmesi en büyük avantajıdır. Bu yöntemde 1. tip hatanın kontrolü için regresyon kaynaklı düzeltme kullanılır (Clauser ve Mazor, 1998; Fang, 1999). Poly-SIBTEST, Crossing-SIBTEST ve multi-SIBTEST gibi çeşitleri vardır. Poly-SIBTEST (Chang, Mazzeo ve Roussos, 1996), iki ve çok kategorili değişkenler için kullanılan örtük özellik yöntemiyken, Crossing-SIBTEST (Li ve Stout, 1993) ise iki kategorili verilerde tek biçimli olmayan DMF'yi belirlemeyi amaçlamaktadır. MULTISIB ise çok boyutlu verilerde DMF inceler.

SIBTEST yöntemi ile farklı gruplarda yetenek düzeyleri eşitlenmiş bireylerin doğru yanıt verme olasılıkları karşılaştırılarak DMF belirlenir. Bu farklara ve odak grubun θ yeteneğindeki yoğunluk fonksiyonuna bağlı aşağıda verilen eşitlik ile β değeri bulunur ve DMF etki büyüklüğü hesaplanır (Bolt, 2000).

$$\beta = \int B(\theta) f_o(\theta) d\theta$$

$$B(\theta) = P_r(U = 1 | \theta) - P_r(U = 1 | \theta)$$

$P_r(U = 1 | \theta)$ = Maddenin referans grubundaki bireyler tarafından doğru yanıtlanma olasılığı

$P_r(U = 1 | \theta)$ = Maddenin odak grubundaki bireyler tarafından doğru yanıtlanma olasılığı

$f_o(\theta)$ = odak grubun θ yeteneğindeki yoğunluk fonksiyonu

β Değerlerini yorumlamak için aralıklar Tablo 6'da (Roussos ve Stout, 1996) verilmiştir. Bu değer pozitif ise referans grup lehine negatif ise odak grup lehine farklılaşma vardır.

Tablo 6

SIBTEST Yöntemi ile DMF Düzeyi Belirlenmesi

Düzyey	Aralık	DMF Büyüklüğü
A	$ \beta < 0,059$	İhmal edilebilir
B	$0,059 \leq \beta < 0,088$	Orta düzey
C	$ \beta \geq 0,088$	Yüksek düzey

Osterlind ve Everson' a (2009) göre SIBTEST yöntemi ile referans ve odak grupları daha doğru bir şekilde eşitlenebilmektedir. Clauser ve Mazor (1998) ise SIBTEST' in tek biçimli DMF tespiti için MH ile benzer sonuçlar verse de 1. tip hatanın kontrolü, yüksek istatistiksel güç ve küçük örneklerde de kullanılabilme özelliklerinden dolayı tercih edildiğini belirtmişlerdir.

MTK Olabilirlik Oran (MTK-OO) Testi Yöntemi. MTK yöntemine bağlı olarak, iki modelin (serbest-sınırlı) olabilirlik oran farklarının test edildiği yöntemdir (Thissen, Seinberg ve Wainer, 1993). Serbest model, parametre değerlerinin kestirimi için bazı parametrelerin farklılaşmasına izin verilen modeldir. Bu farklılık testinde elde edilen katsayı ki-kare dağılımı gösterir ve uygun serbestlik derecesinde tablodan elde edilen değerden büyük bir fark varsa gruplar arasında DMF olduğu belirlenir (Teresi, Kleinman, Welikson, 2000). MTK-OO kullanmak için karşılaştırılacak parametrelerin aynı ölçek düzeyine indirgenmiş olması gerekir (Kim ve Cohen, 1995). Analiz neticesinde elde edilen olabilirlik oranı G^2 olarak isimlendirilir. Bu katsayıya ait DMF düzeyi aralıkları Tablo 7'de verilmiştir (Greecer, 2004).

Tablo 7

MTK-OO Yöntemi ile DMF Düzeyi Belirlenmesi

Düzyey	Aralık	DMF Büyüklüğü
A	$3,84 \leq G^2 < 9,4$	İhmal edilebilir
B	$9,4 \leq G^2 < 41,9$	Orta düzey
C	$G^2 \geq 41,9$	Yüksek düzey

Multi-MTK-OO testi yöntemi Suh ve Cho (2014) tarafından MTK-OO dan geliştirilmiştir. Tek boyutlu MTK-OO yönteminde olduğu gibi model veri uyumunun değerlendirildiği süreçtir. Yetenekler arası bağımsızlığın sağlanması için ortalamaların 0, varyansların 1 olması gereklidir. Ayrıca bu yöntemi uygulayabilmek için boyutlar arası korelasyon sıfır olmalıdır.

MIMIC (Multiple Indicators Multiple Causes) Yöntemi. Camili ve Shepard (1994) odak ve referans grup arasındaki muhtemel farklılıkları ikincil bir faktör üzerinden karşılaştırılmasına imkân verdiği için DFA'nın DMF belirlerken kullanılabileceğini savunmuşlardır. DFA kullanılarak DMF belirlenmesini sağlayan yöntemlerden biri de Çoklu Göstergeler ve Çoklu Nedenler yöntemi olarak bilinen MIMIC yöntemidir. MIMIC yönteminde tek biçimli DMF nin bulunup bulunmadığı aşağıda kurulan modelin test edilmesiyle anlaşılmaktadır (Finch, 2005; Wang, Shih ve Yang, 2009).

$$y_i^* = \lambda_i \eta + \beta_i z_k + \varepsilon_i$$

y_i^* : İ değişkeni gizil yanıtı

η : Gizil özellik

λ_i : İ değişkenine ait faktör yükü

z_k : Gruba üyelik belirten yapay değişken

β_i : Eğim parametresi

ε_i : Tesadüfi hata

Bu yöntem kullanılarak gruplama değişkenine ait dolaylı ve doğrudan etkilerin belirlenmesi sağlanır. Dolaylı etki gizil değişkene ait ortalamanın odak ve referans gruba göre değişip değişmediğini belirler. Doğrudan etki ise maddeyi doğru

yanıtlama olasılıklarının farklılaşması ile ilgilendir. Yukarıda verilen model ile testin farklı örtük özellikleri ölçmesinden kaynaklanabilecek grup farklılaşmaları kontrol edilerek DMF belirlenir (Finch ve French, 2007). DFA yönteminde gözlenen değişkenlerin kategorik olması normallik varsayımına zarar vereceğinden maximum likelihood estimation (MLE) açıklanamayan hatanın artmasına sebep olabilir. Bu nedenle WLS veya RWSL kestirimleri kullanılması tavsiye edilir (Klein ve Moosbrugger, 2000).

Multi MIMIC Yöntemi. MIMIC yönteminin çok boyutlu yapılarda daha güvenilir DMF analizi yapabilmesi için Lee, Suh ve Bulut (2017) tarafından çok boyutlu geliştirilmiş MIMIC önerilmiştir. İki ve daha fazla boyut sayısına sahip iki kategorili testlerde kullanılabilen bu yöntem için aşağıdaki model geliştirilmiştir.

$$y_i^* = \lambda_{1i}\eta_1 + \lambda_{2i}\eta_2 + \beta_1 z + \omega_{1i}\eta_1 z + \omega_{2i}\eta_2 z + \varepsilon_i$$

y_i^* : İ değişkeni gizil yanıtı

η_1 ve η_2 : Gizil özellikler

λ_{1i} ve λ_{2i} : Gizil özelliklere ait i değişkenine ait faktör yükü

z_k : Gruplama değişkeni

β_i : Eğim parametresi

ω_{1i} : İ maddesine ait tek biçimli olmayan DMF parametresi

ε_i : Tesadüfi hata

Bu modelde MTK-OO dan farklı olarak boyutlar arası korelasyona izin verilmektedir. Bu sebeple çalışmada çok boyutlu verilerin DMF analizi için Multi-MIMIC yöntemi kullanılmıştır.

Kullanılan tüm bu yöntemler DMF'yi belirlemek, yansız testler geliştirmek ve neticesinde testlerin geçerliğini artırmak içindir. Test geçerliğinin tarihsel gelişimi ve yaklaşımları da test geçerliği başlığı altında incelenmiştir.

Test Geçerliği

Geçerlik kavramı bir testin ölçmek istediği özelliği, amacına uygun olarak ölçmesiyle ilgilidir. Yani bir test ölçmek istediği değişkeni ne kadar çok başka değişkenler ile karıştırmadan ölçebilirse, testin o derece geçerli olduğu söylenebilir

(Turgut ve Baykul, 2013). Ya da en bilinen tanımı ile geçerlik, bir testin, sadece o testle ölçülmek istenen değişkeni ölçmesi diğer değişkenlerle karıştırmamasıdır. Tüm bunların yanında, geçerliğin, hedeflenen ölçme sürecinin gerçekleştirildiğini kanıtlayan bilgileri toplama süreci/işi olduğu sonucuna varılabilir.

Geçerlik kavramı ilk olarak 1915 yılında ölçüte dayalı geçerlik çatısı altında tanımlanmaya ve tartışılmaya başlamıştır (Lissitz ve Samuelsen, 2007). Sonraki yıllarda (1920-1950), ölçüt dayanaklı geçerlik, geçerlik kavramının ilk karşılığı olmuştur (Angoff, 1998; Cronbach, 1971; Moss, 1992; Shepard, 1993). Bu karşılık, testlerden elde edilen puanlar ile puanlara yordayıcı olarak önceden belirlenen ölçütlerin ilişkisini konu almaktadır. Örneğin, bir işe personel alınırken veya bir okula öğrenci seçimi ve yerleştirmesi yapılırken ölçütlerden faydalanılır. Bu ölçütler ile seçme ve yerleştirmeye tabi tutulanların puanları arasındaki uyum ve testlerden alınan puanların yordama gücü o teste ait geçerliğin derecesini gösterir. Ölçüt dayanaklı geçerliğin genellikle iki alt başlığı olduğu kabul edilir. Bunlardan ilki yordama geçerliği, diğeri ise uyum geçerliğidir. Yordama geçerliği, o ölçekten elde edilen yordayıcı puan ile ölçülmek istenen özellikleri ölçtüğü bilinen ölçüt arasındaki korelasyonun hesaplanması ile elde edilir (Gulliksen, 1959; Lord ve Novick, 1968; Cronbach, 1971). Uyum geçerliği ise eşzamanlı olarak, geliştirilen ölçekten elde edilen puanlarla, belirlenen ölçüt arasındaki korelasyondur (Thorndike ve Hagen, 1959; Carmines ve Zeller, 1982).

Devam eden süreçte geçerliğin bir diğer çeşidi olan “kapsam geçerliği tartışılmaya başlanmıştır. Rulon (1946) ölçüt dayanaklı geçerlik türünün aksine, testin geçerliği için dışarıdan bir ölçütün varlığına ihtiyaç bulunmadığını, esas ölçütün test puanlarının ve testin kendisi olduğunu düşünüyordu. Bu yaklaşıma göre, testlerin hedeflenen konu alanı ve bilişsel süreçleri temsil edip etmediğine ilişkin uzman fikirlerinin alınması gerektiği gerçeği ortaya çıktı. Bu şekilde alan uzmanları tarafından oluşturulan standartlar ile testlerin ölçütlerinin ve içeriklerinin belirlendiği geçerlik türü kapsam geçerliği olarak tanımlanmıştır (Cureton, 1951; Cronbach ve Meehl, 1955; Akt. Lissitz ve Samuelsen, 2007; Kelecioğlu ve Göçer Şahin, 2014).

Test geliştiricileri ve kullanıcılarının yeni problemler ve sınırlılıklar ile karşı karşıya kalmaları sonucu geliştirilen çalışmalar ile birlikte geçerlik yöntemi hakkında ortaya çıkan yenilikler artış göstermiştir. 1950’li yıllardan itibaren Cronbach ve

Meehl'in (1955) yapı geçerliğinin, kapsam ve ölçüt dayanaklı geçerliğe alternatif olabileceği düşüncesi bu yeniliklerin başında gelmektedir. Yapı geçerliği, sadece gözaltında tutulan niteliklerin ölçümü değil ayrıca birbirlerine bağlı niteliklerle ilgili olan ağın doğasını içerir. Messick (1989) yapı geçerliğini, test harici olan şeyleri içeren farklı yapıların birbirleri arasındaki formal ilişkiye sahip olan teori gelişimi ve ölçülen yapıların bütünü olarak da tanımlamaktadır. Buradan hareketle yapı geçerliğinin, tek başına yapı çalışması değil, yapı çalışmasının ve diğer yapıların ilişkilerine dair olan çalışmanın bir birleşimi olduğu söylenebilir.

1999 yılında yayınlanan Standartlar kitabında (APA, AERA ve NCME, 1999) yapı geçerliği diğer geçerlik yaklaşımlarından ayrı tutulsa da tüm geçerlik türlerinin yapı geçerliği çatısı altında incelenmesi gerektiğini savunanların sayısı az değildir. Bu yaklaşıma "Birleştirilmiş Model" de denilmektedir. Loevinger'in (1957) yordama, uygunluk ve kapsam geçerliklerinin yapı geçerliği çatısı altında toplanması gerektiği görüşü 1980'li yılların başında kabul görmeye başlamıştır. Böylece yapı geçerliğinin tüm geçerlik türlerini içerdiği konusunda birleşilmiştir (Anastasi, 1986; Embretson, 1983; Guion 1977; Messick 1980,1988,1989, Akt. Kane, 2006).

Bölüm 2

İlgili Araştırmalar

Alan yazın incelendiğinde DMF belirlemek için kullanılan yöntemlerin karşılaştırıldığı birçok araştırmaya rastlanmıştır. Ancak bu araştırmaların büyük bir kısmı tek boyutluluk varsayımı ile sınırlandırılmıştır. Sayıları az da olsa testin çok boyutluluğunu dikkate alan ve çok boyutluluk ile değişen madde fonksiyonu arasında ilişki gözeten araştırmalar da bulunmaktadır.

Chen ve Jin (2018) çalışmalarında birçok geleneksel DMF belirleme yaklaşımlarının tek boyutluluk varsayımı altında bireylerin toplam puanlarına göre yapıldığına vurgu yapmışlardır. Fakat eğitimde ve psikolojide kullanılan değerlendirme testlerinin çok boyutlu yapıda olmasından dolayı toplam puanların çok boyutlu yapılarda DMF belirlemede etkili olmadığını iddia etmişlerdir. Bu nedenle bir ölçeğin tüm alt puanlarının lojistik regresyonda kullanılmasını ve bireylerin toplam puanları ile karşılaştırılmasını önermişlerdir. Araştırmalarında 21 maddeden oluşan iki boyutlu testi, 250-500 ve 1000 kişilik örneklemlere test yapısını, ortak maddede sayısını, örtük özelliklerin gruplara dağılımlarını ve DMF etki büyüklüğünü değiştirerek uygulamışlardır. Çalışma sonunda hata değerlerinin azaldığı belirtilmiştir.

Bulut ve Suh (2017) çalışmalarında iki ya da daha fazla örtük özelliğe sahip, basit olmayan yapıda ve çok boyutlu testlerde DMF belirleme süreçleri için kullanılan MIMIC, MTK-OO ve LR yöntemlerinden ve elde edilen sonuçlar hakkında bilgi vermektedir. Analizler için iki kategorili, iki boyutlu 2 PLM e göre simülasyon veri seti üretilmiştir. Koşullar maddenin DMF içerme durumu ve içeriyorsa biçimi, test uzunluğu, DMF etki büyüklüğü, örneklem büyüklüğü ve referans ve odak grup oranları, örtük özellikler arasında korelasyonlar olarak belirlenmiştir. Çalışma sonucunda LR nin tek biçimli olmayan DMF için kullanılabildiği, tek biçimli DMF için de kısa testlerde MIMIC ve LR nin MTK-OO dan daha iyi sonuçlar verdiği bulgusuna ulaşılmıştır. Uzun testlerde sonuçlar birbirine yakındır. 1. tip hata için LR yönteminde, istatistiksel güç için ise MTK-OO yönteminde daha yüksek ve dengeli sonuçlar gözlenmiştir.

Özdoğan (2017) çok boyutlu testlerde madde öbeği fonksiyonunu incelemiştir. Yapay veri seti ile gerçekleştirilen çalışmada boyutlar arası korelasyon

değerleri 0,10-0,45 ve 0,80 olarak, örneklem büyüklüğü 2000-5000 ve referans-odak grup oranı 1:3, 1:1, 3:1 olarak belirlenmiştir. Test uzunluğu 30 madde olup ilk 10 madde öbek olarak kabul edilmiştir. ITEMGEN (Ackerman, 1994b) ile elde edilen madde parametreleri kullanılarak genişletilmiş 2 PLM, iki kategorili ve çok boyutlu olarak SAS programı tarafından üretilmiştir. DMF belirlemek için kullanılan SIBTEST yine SAS programı ile gerçekleştirilmiştir. Araştırma neticesinde sonuçlar 1. tip hata ve istatistiksel güç değerleri dikkate alınarak değerlendirilmiştir. Sonuçlara göre artan boyutlar arası korelasyon güç değerlerini düşürmekte, örneklem büyüklüğü ise arttırmaktadır. Ayrıca istatistiksel gücün en yüksek olduğu durum odak-referans grup büyüklüklerinin eşit olduğu durumdur. 1. tip hatalar ise boyutlar arası korelasyonun artmasıyla düştüğü ve örneklem büyüklüğüne paralel olarak arttığı belirlenmiştir. Odak ve referans grup oranlarının da eşit olduğu durumda ise 1. tip hata en yüksek değere ulaşmıştır. Çalışma neticesinde elde edilen bir diğer sonuç ise hata oranlarının ANOVA ile incelenmesiyle bulunmuştur. Buna göre, madde öbeğinde DMF'li madde olmadığı ve diğer 20 maddede DMF'li maddeler olduğunda, öbeğe ait hatalar, boyutlar arası korelasyon, odak-referans grup oranları ve örneklem büyüklüklerine göre manidar bulunmuştur.

Bulut, Lee ve Suh (2016) çok boyutlu testlerde farklı koşullarda MIMIC kullanarak 1. tip hatanın ve gücün nasıl değiştiğini belirlemeye çalışmışlardır. Simülasyon koşulları olarak DMF etki büyüklüğü ve biçimi, test uzunluğu, boyutlar arası korelasyon, örneklem büyüklüğü ve odak referans grupları arasındaki ortalama farklılıkları belirlenmiştir. Çalışma neticesinde tek biçimli DMF yi belirleme gücünün daha yüksek olduğu, örneklem büyüklüğü ve ortak madde sayısının artmasının gücü arttırdığı tespit edilmiştir. Ayrıca odak ve referans grupların ortalamalarının yakın olmasının da gücü arttırdığı belirlenmiştir. Sonuç olarak tek biçimli DMF olan maddelerde güvenilir sonuçlar verdiği düşünülen Multi-MIMIC yönteminin tek biçimli olmayan maddeler için kullanılmasının soru işaretleri içerdiği ifade edilmiştir.

Walker ve Göçer Şahin (2016) çok boyutluluğa sebep olan ikincil yetenek düzeyi dağılımının, değişen madde fonksiyonu belirleme süreci nasıl etkilediğini üç farklı teknik (MH, LR ve SIBTEST) kullanarak belirlemeye çalışmışlardır. İki boyutlu ve iki kategorili veri seti çok boyutlu telafisel modele göre simüle edilmiştir. Test 30 madde içermektedir. Bu maddeler önceden belirlenmiş madde parametreleri ile 25

madde ilk yetenek düzeyini, beş madde de ikincil yetenek düzeyini ölçecek şekilde üretilmiştir. Boyutlar arası korelasyon sistematik olarak arttırılmıştır (0,00-0,25-0,50-0,75). Odak ve referans grupları ikinci yetenek düzeyindeki ortalama farklılığına göre belirlenmiştir. Üç farklı DMF belirleme tekniği, SIBTEST, MH ve LR, kullanılmıştır. Her bir analiz neticesinde 1. tip hatalar DMF içermeyen maddelerde ve istatistiksel güç oranları da DMF içeren maddelerden yola çıkarak hesaplanmıştır. Ardından elde edilen sonuçların etkileşim ve değişimlerini belirlemek için ANOVA testi uygulanmıştır. SIBTEST ile boyutlar arası korelasyon arttığında 1. tip hatanın azaldığı ve 0,05 e yaklaştığı görülmüştür. Ayrıca MH ile güç oranlarının SIBTEST e göre daha iyi sonuçlar verdiği tespit edilmiştir. Sonuç olarak bu araştırma neticesinde, ikinci yetenek düzeyindeki küçük farklılaşmalarının değişen madde fonksiyonuna sebep olabileceği ortaya çıkmıştır.

Ling Liaw (2015) tez çalışmasında, öncül yeteneğin ölçülmesi sürecinde, çok boyutlu ve tek boyutlu modellerin DMF belirleme ve kontrol durumlarının karşılaştırması yapılmıştır. 40 maddelik veri seti ve 2000 kişilik örneklem kullanılmıştır. Maddeler 2 PLM olup iki boyutlu telafisel olmayan model kullanılmıştır. Boyutlar arası korelasyon 0,0 ve 0,3 olarak, referans ve odak grup oranı 1:1 ve 9:1, öncül yetenek etki düzeyi 0,5 ve 0,8, ikincil yetenek etki düzeyi 0,2 ve 0,5 ve DMF içeren madde oranı %0, %10, %20 ve %30 olarak koşulları oluşturmaktadır. 500 tekrarın yapıldığı çalışmada koşul ve model etkileşimleri için varyans analizi (ANOVA) yapılmıştır. Sonuç olarak tek boyutlu model yanlılık konusunda en kötü sonuçları verirken, çok boyutlu modeller ile sınırlı madde havuzlarında bile DMF'li maddeleri kaldırmadan da uygun neticeler elde edilmiştir.

Huang (2010) tez çalışmasında, uluslararası büyük ölçekli sınavlarda, (PISA, PIRLS ve TIMSS), maddelerde gözlenen DMF nin dilden mi, eğitim müfredatından mı yoksa kültürden mi kaynaklandığını tek boyutlu ve çok boyutlu DMF belirleme yöntemleri kullanarak belirlemeye çalışmış ve sonuçlarını karşılaştırmıştır. Analizler neticesinde klasik yöntemlerin çok boyutlu testlerde çok fazla DMF'li madde tespit ettiği belirlenirken, çok boyutlu Rasch tabanlı çok düzeyli lojistik modeli (ÇBRTÇDLM) kullanarak daha az sayıda ve daha az hata ile DMF'li madde tespiti yapmıştır.

Furlow, Ross ve Gagne (2009) çalışmalarında, referans grubu odak gruptan çok daha üstün yetenek düzeyi parametresine sahip olarak tasarlamışlardır.

DMF'ye sebep olan bu durum 1. tip hata ve istatistiksel güç olarak incelenmiş, örneklem büyüklükleri, referans ve odak grup oranları, boyutlar arası korelasyonlar, DMF büyüklükleri, test uzunlukları ve test içinde madde öbekleri oranları değiştirilerek koşullar oluşturulmuştur. Tek boyutlu veriler IRTGEN (Whitaker, Fitzpatrick, Williams, & Dodd, 2003), çok boyutlu veriler SAS/IML program GENMIRT (Kromrey, Parshall, Chason, & Yi, 1999) ile üretilmiştir. Tüm analizlerde SIBTEST kullanılmış ve analizler neticesinde istatistiksel gücün, DMF etki büyüklüğü, madde öbek oranı yüzdesi, örneklem büyüklüğü arttıkça arttığı gözlenmiştir. Ayrıca 1. tip hatanın tüm koşullarda beklenen değere yakın olduğu belirlenmiştir.

Snow ve Oshima (2009) ise iki boyutlu bir veri setine tek boyutlu ve üç boyutlu DMF analizi yapmışlardır. Araştırma için 40 maddelik iki boyutlu veri seti 2000 kişilik örneklem büyüklüğü (1000 referans grup-1000 odak grup) için SAS 9.13 (SAS Institute, Cary, NC) programı kullanılarak üretilmiştir. Madde setleri için ölçümlene NOHARM (Fraiser, 1988) ile yapılmıştır. DMF belirlemek için 1D DFIT program (DFIT8; Raju, Oshima, & Wolach, 2009) ve Multidimensional DFIT programı kullanılmıştır. Araştırmada 100 tekrar yapılmıştır. Çalışma neticesinde iki boyutlu veri seti için üç boyutlu olduğu varsayılarak DMF analizi yapıldığında istatistiksel gücün oldukça düşük çıktığı belirlenmiştir. Ayrıca iki boyutlu veri seti için tek boyutlu varsayımı ile DMF analizi yapıldığında 1. tip hatada değişkenlik ve tek biçimli olmayan DMF nin yok olması durumu ortaya çıkmaktadır.

Kim ve Jang (2009) Kanada'da farklı dil ve kültür özelliklerine sahip öğrencilerin Ontario Secondary School Literacy Testinde (OSSLT-Mart 2006) performansına dayalı bir DMF çalışması yürütmüşlerdir. Araştırma örneklemini 5644 10. sınıf öğrencisi oluşturmaktadır. Bu öğrencilerden 1969' u evinde ilk dil olarak İngilizceyi kullananlar (L1), 3675' i ise ilk dil olarak İngilizceyi kullanmayanlardır (ELL). Çalışma için İngilizcenin alt öğrenme alanları ayrı ayrı boyutlar olarak ele alınmış ve L1 ve ELL öğrencileri arasında farklılaşmaya sebep olup olmadığı belirlenmeye çalışılmıştır. Araştırma neticesinde kelime bilgisine sahip L1 öğrencilerinin avantaja sahip olduklarını fakat gramer, okuma ve yazma becerilerinde ELL öğrencilerinin daha iyi olabileceğini ortaya koymuştur.

Walker, Zhang ve Surber (2008) okuma yeteneğinin matematik performansına etkisini çok boyutlu değişen madde fonksiyonu (ÇBDMF) ile

belirlemeye çalışmışlardır. Bu çalışmada ilköğretim 5. ve 6. sınıf öğrencilerine yapılan uygulamadan elde edilen gerçek veri seti kullanılmıştır. 5868 5. Sınıf, 6301 6. Sınıf öğrencisinin katıldığı örneklemin %48,2 si kız öğrencidir. Tera Nova çoklu değerlendirme testi kullanılmış ve öğrenciler okuma becerileri düzeylerine göre gruplandırılmışlardır. Testin çok boyutluluğunu belirlemek için NOHARM (Fraiser, 1988) kullanılmış ve testin yarı telafisel iki boyutlu modele uyumlu olduğu tespit edilmiştir. DMF belirlemek için SIBTEST yöntemi kullanılmıştır. Çok boyutluluk testinde okuduğunu sınıflandırma becerisi düşük olan öğrencilerin matematik sorularını doğru yanıtlamada güçlük çektiği belirlenmiş ve bu durumun maddenin farklı fonksiyonlaşmasına sebep olduğu sonucuna varılmıştır.

Barnett ve Ercikan (2006) çok boyutlu matematik testlerinde cinsiyetten kaynaklanan DMF yi belirlemek için 2000 yılında uygulanan British Columbia Provincial Principles of Mathematics Examination for Grade 12 (PME-12) sınavını kullanmışlardır. Sınava 5069 erkek 4335 kız öğrenci katılmıştır. Problem çözme becerisini ölçen matematik soruları metin içeren ve içermeyen olmak üzere ikiye ayrılmaktadır. DMF analizleri için SIBTEST yöntemi kullanılmıştır. Araştırma neticesinde, bazı metinlerin kızlar lehine bazılarının da erkekler lehine farklılaşmaya sebep olduğu belirtilmiştir. Ayrıca önceki araştırmaların aksine geometri içeren soruların kızlar lehine avantaj sağlamadığı, üst düzey bilişsel süreç gerektiren sorularda erkeklerin, ardışık işlem kümesi gerektiren sorularda kızlar lehine farklılaşma olduğu tespit edilmiştir.

Stout, Li, Nandakumar ve Bolt (1997) araştırmalarında, isteyerek iki boyutlu olarak hazırlanan bir testte DMF belirlemek için kullanılan MULTISIB (Shealy ve Stout, 1993) yöntemini incelemişlerdir. Örneğin; matematik testinde geometri ve cebir becerilerini ölçen maddeler karışık şekilde hazırlanmıştır. Çalışma sonucunda iki boyutlu testlerde ikincil boyutun DMF ye sebep olduğunun fakat bu durumun MULTISIB tekniği ile en aza indirildiği ve yapay veri çalışmaları ile 1. tip hata ve istatistiksel güç değerlerinin iyi sonuçlar verdiği belirtilmiştir. Ayrıca çok boyutlu testlerde toplam test puanı yerine boyutlardan elde edilen puanların alt boyut puanı (subdimension-score) olarak değerlendirilmesinin daha uygun sonuçlar vereceği belirtilmiştir.

Oshima, Raju ve Flowers (1997) araştırmalarında bilinen madde parametreleri ve DMF ile iki boyutlu veri üretmiş ve çok boyutlu DMF belirleme

süreçleri kullanarak sonuçları değerlendirmişlerdir. İki boyutlu ve 2 PLM olarak telafisel modele göre üretilen 40 maddelik veri seti üretilmiştir. Çalışmanın koşullarını, maddelerin tek biçimli ve tek biçimli olmama durumları, referans ve odak grup oranları, DMF'li madde sayısı ve DMF etki büyüklüğü olmuştur. Madde parametreleri ölçümlemesi için NOHARM, DMF analizleri için DFIT programı kullanılmıştır. Araştırma sonucu olarak çok boyutlu testlerde DMF analizlerinin çok boyutluluk esasına göre yapılması ile daha uygun sonuçlar elde edildiği belirtilmiştir.

Mazor ve Ripkey (1996) testlerin toplam puanlarına göre gruplandırılan bireylerin yeteneklerini karşılaştırırken testlerin tek boyutlu olduğu varsayımı yapıldığı fakat tek boyutluluğun sağlanmadığı durumlarda 1. tip hatanın arttığını belirtmişlerdir. Bu çalışmada simülasyon ve gerçek veri seti kullanılmıştır. Simülasyon veri setinde madde parametreleri BILOG ile belirlendikten sonra DATAGEN programı ile üretilmiştir. Örneklem sayısı 2000 ile sınırlandırılmıştır (1000 kız-1000 erkek). Toplam test puanlarına, alt test puanlarına ve çoklu alt test puanlarına göre yapılan DMF analizlerinde MH ve LR yöntemleri kullanılmış, test çok boyutlu olduğunda testin tek boyutlu olan küçük parçalara ayrılması ve eşleştirme kriterinin çoklu alt test olarak belirlenmesinin en iyi sonuçları verdiği belirtilmiştir.

Bolt ve Stout (1996) DMF belirlemek için SIBTEST yönteminin çok boyutlu testlerde nasıl kullanıldığını anlatmışlardır. Bu yöntem ile DMF belirlemek dışında, DMF 'ye neden olan durumların saptanabileceğini öne sürmüşlerdir. SIBTEST yönteminin parametrik olmayan bir yöntem olduğu ve bir veri setinin içerdiği DMF miktarını artan ve kontrol edilen 1. tip hata ile belirleyebileceği sonucuna varmışlardır. Araştırmada çok kategorili gerçek ve üretilmiş veri setleri kullanılmıştır.

Camili (1992) çalışmasında çok boyutlu modellerde DMF belirleme sürecine, gerekliliğine ve önemine vurgu yapmıştır. Araştırmacı DMF' nin sadece grup farklılıklarından kaynaklanmadığını ve ölçmeyi etkileyen ikincil yetenek düzeyi dağılımlarının da dikkate alınması gerektiğini belirtmiştir.

Yapılan çalışmalar incelendiğinde DMF belirleme süreçlerinde testin boyut yapısının önem taşıdığı görülmektedir. Testlerin tek boyutlu olarak ele alınmasının elde edilen bulgularda hatayı arttırdığı ve yanıltıcı sonuçların ortaya çıkmasına sebep olduğu görülmektedir. Çalışmalardan bazıları testin çok boyutlu olmasının

DMF ye sebep olabileceğini belirtirken, bazıları ise DMF belirlerken ölçülen özellikler arasındaki ilişkinin, öncelik durumunu ve etkileşiminin bilinmesi ile DMF süreçlerinin çok daha etkin ve hatasız sürdürülebileceğini iddia etmektedir. Yöntem olarak değerlendirildiğinde SIBTEST, MULTISIB, DFIT, MULTI-DFIT, MH, LR; MTK-OO, MIMIC ve Multi-MIMIC modellerinin sıklıkla tercih edildiğinin ve yapay veri setleri için 1. tip hata ve istatistiksel güçlerin kıyaslandığı görülmüştür. Çalışmalarda veri setleri genellikle iki boyutlu olarak tasarlanmaktadır. Ancak iki boyutlu ve üç boyutlu veri setleri ile gerçek veri setinde aynı anda içeren ve aynı koşullara sahip çalışmaya rastlanmamıştır. Bu tez çalışmasında farklı boyutluluk modellerinin, yapay veri setinde DMF belirleme gücü ve 1. tip hata üzerindeki etkisi ile gerçek veri setinde DMF' li madde sayıları belirlenirken testin boyut sayısı dikkate alınmıştır.

Bölüm 3

Yöntem

Araştırmanın Modeli

Araştırmada DMF yöntemleri, farklı koşullarda simülasyon verileri ile farklı boyutluluk modellerinde karşılaştırılmıştır. Bu yönüyle çalışma simülatif betimsel desen araştırmasıdır. Buna ek olarak PISA 2012 verisi kullanılarak var olan durum belirlenmeye çalışıldığından bu çalışma aynı zamanda betimsel çalışma özelliği göstermektedir.

Çalışma Grubu

Araştırmada farklı koşullarda üretilmiş iki ve üç boyutlu veri setlerine DMF analizleri uygulanmıştır. Ayrıca PISA 2012 sınavına katılan ve Türkiye örneklemini oluşturan 4848 (2370 kız, 2478 erkek) öğrenci bu çalışmada kullanılan gerçek veri setini oluşturmaktadır. Seçim yapılırken öğrencilerin yetenek düzeyleri arasında belirgin farklılık olmaması için okul türleri ve başarı durumları dikkate alınmıştır.

Simülasyon Koşulları

Araştırmada iki boyutlu ve üç boyutlu olarak üretilen verilere DMF analizleri uygulanarak, istatistiksel güç ve 1. tip hata belirlenmiştir. Veriler üretilirken farklı koşullar ele alınmıştır.

Test Uzunluğu. Benzer çalışmalar incelendiğinde Bulut ve Suh (2017) değişen test uzunluğunu (20-22), Özdoğan (2017), Walker ve Şahin (2016) 30 madde, Ling-Liaw (2015), Huang (2010), Snow ve Oshima (2009), Oshima, Raju ve Flowers (1997) 40 maddeye sahip çok boyutlu testleri seçtikleri görülmüştür. Bu çalışma üç boyutlu veri de içerdiğinden çalışmada maddeler arası etkileşimden dolayı yetersiz madde miktarından kaynaklanan hatayı arttırmaması ve ilgili çalışmalar da göz önüne alınarak bu çalışmada test uzunluğu 40 madde olarak belirlenmiştir. Maddelerin yarısı DMF içerirken yarısı içermemektedir.

Testin Yapısı. Test iki boyutlu ve üç boyutlu olarak basit ve karmaşık yapıda üretilmiştir. Basit yapıda yer alan alt testler ile belirlenmiş özellikler maddeler aracılığı ile ölçülebilir. Alt testler tek boyutlu iken, testin bütününün çok boyutlu olduğu varsayılır. Bu sebeple bu modele çoklu-tek boyutlu model (multi-

unidimensional) de denmektedir. Karmaşık yapı ise maddeler içi model veya basit olmayan yapı olarak adlandırılır. Bu yapıda maddeler birden fazla yetenek ile matematiksel ilişkiye sahip olabilir (Bulut, 2013). Boyutlar arası korelasyon tek boyutluluğun engellenmesi için 0.3 ile sınırlı tutulmuştur.

DMF Etki Büyüklüğü. Literatür incelendiğinde DMF etki büyüklüğünün 0 dan 1.50 ye kadar değerler aldığı görülmektedir (Johnson Frotman, 2007; Hidalgo ve Lopez-Pina, 2004; Samuelsen, 2005). Bu çalışmada ise DMF etki büyüklüğü 0,50 ve 1,00 olarak belirlenmiştir. 0,50 orta düzeyde etki büyüklüğünü, 1,00 ise yüksek düzeyde etki büyüklüğü olarak kabul edilmiştir.

Örneklem Büyüklüğü. DMF çalışmalarında etkisi incelenmek istenen değişkenlerden biri de örneklem büyüklüğüdür. DMF çalışmaları incelendiğinde doğru sonuçlar alabilmek için referans ve odak grup için 200 den yüksek olması beklenmektedir (Carter, 2011; Johnson Frotman, 2007; Jodoin ve Gierl, 2001; Li, Brooks ve Johanson, 2012; Narayanan ve Swaminathan, 1996; Wyse ve Mapuranga, 2009). Ancak çok boyutlu testlerin en az 1000 kişilik örneklem büyüklüğüne sahip olması gerektiğini savunan araştırmalar da bulunmaktadır (Bolt ve Lall, 2003; Lee, 2012). Bu sebeple araştırmada örneklem büyüklükleri 600, 1200, 2400, 3600, 4800, 6000 olarak belirlenmiştir.

Odak ve Referans Grup Oranı. Referans ve odak grup büyüklükleri ve bu büyüklüklerin oranı istatistiksel analizlerin sağlamlığını etkileyebilecek potansiyele sahiptir (Swaminathan ve Rogers, 1990). Bu oranlar da 1:1 ve 1:2 olarak belirlenmiştir ve örneklem büyüklüklerine göre dağılımı Tablo 8'de gösterilmiştir.

Tablo 8

Odak-Referans Grup Büyüklükleri

Örneklem Büyüklüğü	1:1 (Odak-Referans)	1:2 (Odak-Referans)
600	300-300	200-400
1200	600-600	400-800
2400	1200-1200	800-1600
3600	1800-1800	1200-2400
4800	2400-2400	1600-3200
6000	3000-3000	2000-4000

Çalışmada yer alan tüm koşullar Tablo 9'da gösterilmiştir.

Tablo 9

Simülasyon Koşulları

Testin Yapısı	Boyut Sayısı	DMF Etki Büyüklüğü	Odak-Referans Grup Oranı	Örneklem Büyüklükleri
Basit	2	Orta	1:1	600, 1200, 2400, 3600, 4800, 6000
			1:2	
		Yüksek	1:1	
			1:2	
Karmaşık	2	Orta	1:1	
			1:2	
		Yüksek	1:1	
			1:2	
Basit	3	Orta	1:1	
			1:2	
		Yüksek	1:1	
			1:2	
Karmaşık	3	Orta	1:1	
			1:2	
		Yüksek	1:1	
			1:2	

Testin iki boyutlu ve üç boyutlu olması, basit veya karmaşık yapıli olması, DMF etki büyüklüğünün orta ve yüksek düzeyde olması, odak ve referans grup oranlarının 1:1 ve 1:2 olması ile altı farklı örneklem büyüklüğü (600-1200-2400-3600-4800-6000) olması araştırmanın koşullarındadır. Elde edilen veri setinin hem boyut sayısı dikkate alınarak hem de tek boyutlu olarak analiz edilmesinden dolayı toplam koşul sayısı 192'dir $((2*2*2*2*6)*2)=192$. Analiz neticelerinin güvenilirliğini sağlamak için, 100 tekrar yapılmıştır. Her bir analiz için aynı madde parametrelerine sahip farklı veri setleri ile analizler tekrarlanmıştır. DMF belirleme ve 1. tip hata ve istatistiksel güç çalışmalarında tekrar sayıları 25-50-100 olarak karşımıza çıkmaktadır. Bu sebeple çalışmada tekrar sayısı 100 olarak belirlenmiştir.

Verilerin Türetilmesi

Veriler koşullara uygun olarak simülasyon yolu ile üretilmiştir. Simülasyon veri kullanırken iki önemli husus vardır. İlki araştırmacı tarafından uygun yöntem belirlenmesi, ikincisi ise korelasyon ile elde edilmiş veri setinin gerçek durumu yansıttığından emin olunmasıdır (Way, Ansley ve Forstyh, 1988).

Çalışma iki ve üç boyutlu yapay veri setleri içermektedir. Çok boyutlu veriler üretilirken maddelerin ayırt edicilik değerleri, hangi yetenek düzeyine daha yakın oldukları, birbirilerini telafi etme durumları ve boyutlar arası korelasyonları oldukça önemlidir (Ackerman, Gierl ve Walker, 2003).

Maddenin ayırt edicilik gücü maddenin ölçmek istediği yetenek ile ilgili bilgi verir (Reckase, 2009). Çok boyutlu testlerde bir madde birden fazla örtük özelliği ölçüyor olabilir. Ancak maddenin hangi yeteneği daha çok ölçtüğü veya ölçülen yeteneklerden hangisinin öncül hangisinin ardıl olduğu, ölçülen yetenek eksenleri ile arasındaki açı belirlenerek bulunur. Bu açı öncül olarak yeteneğin bulunduğu yatay eksenle ardıl yeteneğin ölçüldüğü dikey eksen arasındaki pozitif yönlü dar açıdır. Eksenler arasındaki açı 0° ve 90° arasında değerler alır. Açının hesaplanması için aşağıdaki eşitlik kullanılır (Reckase, 2009).

$$\alpha = \arctan\left(\frac{a_2}{a_1}\right)$$

Verilerin üretilmesi için gerekli olan madde parametreleri ITEMGEN (Ackerman, 1994) programı ile elde edilmiştir. 40 madde ile oluşturulan testler için maddelerin ayırt edicilik güçleri ve ölçtükleri yeteneklere göre sahip oldukları açı, çok boyutlu ayırıcılık değerleri EK-A'da verilmiştir. Oluşturulan madde parametreleri ile SAS programı kullanılarak iki kategorili (1-0) ve çok boyutlu genişletilmiş 2PLM'ye göre simülasyon koşullarına uygun olarak veri setleri oluşturulmuştur.

Üretilen veri setlerinin doğruluğunun kontrol edilmesi için koşullar arasından rastgele seçilen yapay veriler FACTOR 10.8 (Lorenzo-Seva, Ferrando, 2016) ile analiz edilmiştir. Yapılan bu testler neticesinde iki boyutlu veri setlerinde Kaiser-Meyer-Olkin (KMO) test sonuçları 0,838 ile 0,973 arasında değerler almıştır. Yapılan analizlerde örneklem büyüklükleri arttıkça bu testin sonucunun yükseldiği belirlenmiştir. İki boyutlu veriler için yapılan analizler neticesinde; Root Mean Square of Error of Approximation (RMSEA), 0,000 ile 0,030 arasında, Root Mean Square of

Residuals (RMSR) 0,028 ile 0,065 arasında ve Weighted Root Mean Square of Residuals (WRMSR) 0,034 ile 0,064 arasında deęişen deęerlere sahip olduęu tespit edilmiştir. Bu deęerler de hatalar için öngörülen üst sınırdan düşük olduęu için uygunluk göstermektedir (Kelley,1935; Harman, 1962). Uyum iyilięi testlerinden Non-Normed Fit Index (NNFI) 0.988 ile 1 arasında deęerler alırken, Comperative Fit Index'in (CFI) 0,990 ile 1, Goodness of Fit Index'in (GFI) 0,962 ile 0,995, Adjusted Goodness of Fit Index'in (AGFI) 0,953 ile 0,995 arasında deęerler aldıęı belirlenmiştir. Yu & Muthen'e göre (2002) bu aralıkta deęerler yüksek uyuma işaret etmektedir.

Yapılan bu testler neticesinde üç boyutlu veri setlerinde KMO test sonuçları 0,846 ile 0,975 arasında deęerler almıştır. Yapılan analizlerde örneklem büyüklükleri arttıkça bu testin sonucunun yükseldięi belirlenmiştir. üç boyutlu veriler için yapılan analizler neticesinde; RMSEA deęerleri 0,000 ile 0,027 arasında, RMSR 0,028 ile 0,065 arasında ve WRMSR 0,034 ile 0,064 arasında deęişen deęerlere sahip olduęu tespit edilmiştir. Bu deęerler de hatalar için öngörülen üst sınırdan düşük olduęu için uygunluk göstermektedir (Kelley,1935; Harman, 1962). Uyum iyilięi testlerinden NNFI 0,990 ile 1 arasında deęerler alırken, CFI'nın 0,992 ile 1, GFI 0,966 ile 0,997, AGFI 0,960 ile 0,996 arasında deęerler aldıęı belirlenmiştir. Yu & Muthen'e göre (2002) bu aralıkta deęerler yüksek uyuma işaret etmektedir. Sonuç olarak iki boyutlu ve üç boyutlu olarak farklı test yapılarında ve farklı koşullar altında üretilen yapay verilerin araştırmaya uygun olduęu belirlenmiştir.

Gerçek veri seti için Uluslararası Öğrenci Deęerlendirme Programı- Programme For International Student Assessment (PISA) 2012 Türkiye örneklemini kullanılmıştır. PISA Ekonomik İşbirlięi ve Kalkınma Örgütü-OECD'nin 15 yaş grubunda ve örgün eğitime devam eden öğrenciler için üçer yıllık periyodlar halinde düzenledięi sınavdır. Öğrencilerin bu sınavda matematik, fen bilimleri ve okuma alanlarında sahip oldukları bilgileri yeni durumlara, günlük hayata uyarlama ve bu bilgileri kullanabilme becerileri ölçülmektedir. Bu sınav sonuçlarına göre ülkeler kendi eğitim programlarını gözden geçirmekte ve çağın gereklerine ayak uydurabilmek için gerekli güncellemeleri yapma şansı bulmaktadırlar.

PISA uygulaması öğrencinin becerilerini okuryazarlık başlıęı altında ele alır. Okuryazarlık öğrencilerin temel alanlarda karşılaşılan problemleri çözerken, bilgilerini, becerilerini kullanmaları, analiz etmeleri ve neticelerini mantık muhakeme

süreçlerinden geçirerek sonuca ulaşmalarını ifade etmektedir (MEB, 2010). PISA uygulamasında her sınavda farklı bir alana ağırlık verilmektedir. Ağırlıklı alanlar 2000 yılında okuma becerileri, 2003 yılında matematik okuryazarlığı, 2006 yılında fen bilimleri okuryazarlığı, 2009 yılında okuma becerileri olarak belirlenmiştir.

2012 yılında gerçekleştirilen PISA uygulamasında ağırlıklı alan ise matematiktir. Matematik okuryazarlığı bir problemi matematiksel olarak kurgulama, işlem, mantık, muhakeme ve çözme süreçlerini ele alır. Ayrıca öğrencinin elde ettiği sonucun gerçek hayata uygun olup olmadığına da karar verme becerisini de kapsar (MEB, 2010). PISA 2012 uygulamasına, 65 ülkeden yaklaşık 510 bin öğrenci katılmıştır (MEB, 2013). Bu öğrencilerden 4848'i Türkiye'de örgün eğitime devam eden 955.349 öğrenci arasından seçilmiştir. Türkiye için sınav sonucu rapor edilirken (MEB 2013), bölgeler arası farklılıklar, okul türleri arası farklılıklar, öğrencilerin sahip olduğu olanaklar ve okul iklimi dikkate alınmaktadır.

PISA 2012 matematik okuryazarlığını ölçen 30 soru ile okuma becerilerini ölçen 30 soru alınarak ve gerekli boyutluluk analizleri ile iki boyutlu olduğu tespit edilen 60 soru kullanılmıştır. Boyutluluk analizleri için farklı örneklem büyüklükleri gerektiğinden önce 600 kişilik (300-300, 200-400) ardından 1200 kişilik (600-600, 400-800) ve son olarak da 3600 kişilik (1800-1800, 1200-2400) örneklem seçilmiştir. Odak grup kızlardan, referans grup erkeklerden oluşmaktadır. Odak ve referans grup oranı 1:1 ve 1:2 olarak alınmıştır. Testlerin iki boyutlu olduğunun belirlenmesi için yine FACTOR 10.8 (Lorenzo-Seva, Ferrando, 2016) programı kullanılmıştır. KMO test sonuçları 0,837 ile 0,959 arasında değerler almıştır. Yapılan analizlerde örneklem büyüklükleri arttıkça bu testin sonucunun yükseldiği belirlenmiştir. İki boyutlu veriler için yapılan analizler neticesinde; RMSEA değerleri 0,007 ile 0,043 arasında, RMSR 0,032 ile 0,077 arasında ve WRMSR 0,041 ile 0,066 arasında değişen değerler almıştır. Bu değerler de hatalar için öngörülen üst sınırdan düşük olduğu için uygunluk göstermektedir (Kelley, 1935; Harman, 1962). Uyum iyiliği testlerinden NNFI 0,984 ile 0,998 arasında değerler alırken, CFI'nin 0,988 ile 1, GFI 0,957 ile 0,975, AGFI 0,958 ile 0,978 arasında değerler aldığı belirlenmiştir. Yu & Muthen'e göre (2002) bu aralıkta değerler yüksek uyuma işaret etmektedir. Faktör analizi sonuçlarına göre, PISA testinden alınan bu 60 maddenin 2 boyutlu 1 test oluşturduğu kabul edilerek DMF analizlerine geçilmiştir.

Verilerin Analizi

Koşullara uygun olarak üretilen veri setleri ile gerçek veri setindeki boyutlar arası ilişki ve boyut sayısının kontrolü için FACTOR 10.8 programı kullanılmıştır. Ardından SAS programı (SAS 9.4, SAS Institute, Cary, NC) ile tek boyutlu veri setleri için SIBTEST analizleri gerçekleştirilmiştir. Çok boyutlu verilerde Multi-MIMIC yöntemi ile DMF analizleri için de Mplus 7 (Muthen ve Muthen, 1998) programı kullanılmıştır. Yapay veri setleri ile 100 tekrar yapılmıştır. Ayrıca analizler neticesinde elde edilen bulgular için varyans analizleri, karşılaştırma ve ilişki testleri için ise SPSS 23 (IBM, 2019) programından faydalanılmıştır. Verilerin analizinde kullanılan kodlar EK-C ve EK-Ç'de verilmiştir.

1. Tip Hata ve İstatistiksel Gücün Belirlenmesi

1. tip hata; DMF göstermeyen bir maddenin DMF'li bir madde olarak belirlenmesi durumuyken, istatistiksel güç ise DMF'li olduğu bilinen maddenin DMF'li olarak sınıflanmasıdır (Vaughn ve Wang, 2010). Çalışmanın yapay veri içeren kısmında testin boyut sayısının, basit veya karmaşık yapıda olmasının, orta veya yüksek düzeyde DMF içermesinin, örneklem büyüklüğü ile odak ve referans grup oranlarının 1. tip hata ve güç üzerindeki etkileri araştırılmıştır. Çalışmada anlamlılık seviyesi (α düzeyi) 0,05 olarak belirlenmiştir.

Bölüm 4

Bulgular ve Yorumlar

Bu bölümde araştırma neticesinde elde edilen bulgulara ve bu bulguların diğer çalışmalar çerçevesinden yorumlanmasına yer verilmiştir.

Alt Problem 1 e Ait Bulgular

a) İki boyutlu basit yapıli verilere iki boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, 1. tip hata ve istatistiksel güç değerleri, DMF etki büyüklüğü orta düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

Bu alt probleme cevap bulabilmek için belirlenen koşullara uygun olarak gerçekleştirilen DMF analizleri neticesinde elde edilen 1. tip hata ve güç oranları Tablo 10'da verilmiştir.

Tablo 10

İki Boyutlu Basit Yapılı ve Orta Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları

Testin Yapısı	DMF Etki Düzeyi	O/R	Örneklem Büyüklüğü	SIBTEST		Multi-MIMIC	
				1.Tip Hata	Güç	1.Tip Hata	Güç
Basit	Orta	1:1	600	0,78	0,95	0,25	0,96
			1200	0,60	0,95	0,21	0,97
			2400	0,33	0,95	0,17	0,97
			3600	0,18	0,95	0,13	0,97
			4800	0,10	0,95	0,07	0,96
			6000	0,06	0,95	0,05	0,96
Basit	Orta	1:2	600	0,81	0,95	0,26	0,96
			1200	0,66	0,95	0,22	0,97
			2400	0,38	0,95	0,17	0,96
			3600	0,21	0,95	0,14	0,97
			4800	0,13	0,95	0,07	0,96
			6000	0,07	0,95	0,06	0,96

Testin İki boyutlu, basit yapıda olduğu ve maddelerin orta düzeyde DMF içerdiği bu durumda her iki yöntemde de 1. tip hata örneklem büyüklüğü arttıkça azalmaktadır. En yüksek 1. tip hatanın (0,81) olduğu örneklem büyüklüğü 600 olurken, en düşük sonuç 6000 (0,06) kişilik örneklemden elde edilmiştir. Odak ve

referans gruplarının oranlarına göre (1:1, 1:2) karşılaştırıldığında, 1. tip hatalarda, 1:2 olduğu koşulda hafif düzeyde artış görülmüştür. 1. tip hatalar DMF yöntemine göre kıyaslandığında özellikle düşük örneklem büyüklüklerinde ciddi gözlenmiştir. Örneğin odak ve referans grup büyüklüklerin eşit olduğu 600 kişilik örneklem büyüklüğünde SIBTEST ile elde edilen 1. tip hata 0,78 olarak ölçülürken, aynı koşullar altında Multi-MIMIC ile gerçekleştirilen İki boyutlu DMF analizinde bu değer 0,25 olarak bulunmuştur. Örnekleme ve yöntemine göre fark olmasına rağmen 1. tip hatanın 1 koşul hariç 0,05 ve altına düşmediği ve yüksek olduğu belirlenmiştir.

DMF analizlerinden elde edilen istatistiksel güçler incelendiğinde ise DMF'li maddeyi belirleme gücünün %95 ile %97 arasında değiştiği belirlenmiştir. Testin tek boyutlu (SIBTEST) ya da iki boyutlu (Multi-MIMIC) analiz edilme durumuna göre çok küçük düzeyde farklılıklar oluşmuştur. Bunun yanı sıra örneklem büyüklüğünün değişmesinin testin gücünü etkilememesi de tespit edilen bulgulardan biridir.

b) İki boyutlu basit yapıları verilere iki boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, DMF etki büyüklüğü yüksek düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

Bu alt probleme cevap bulabilmek için belirlenen koşullara uygun olarak gerçekleştirilen DMF analizleri neticesinde elde edilen 1. tip hata ve güç oranları Tablo 11'de verilmiştir.

Tablo 11

İki Boyutlu Basit Yapılı ve Yüksek Düzeyde DMF Etki Büyüklüğüne (C düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları

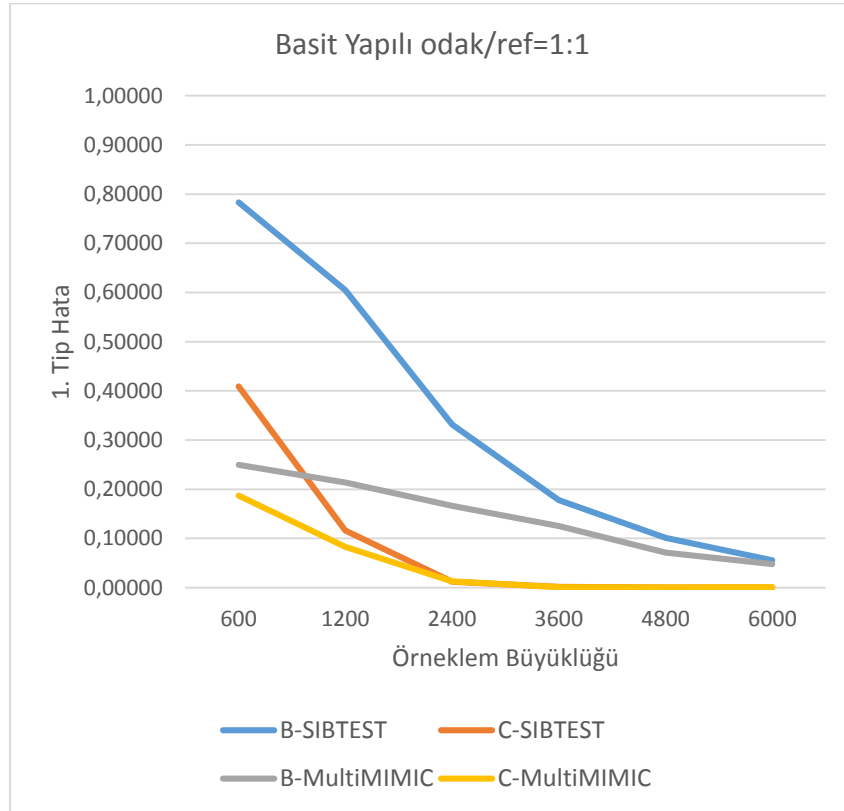
Testin Yapısı	DMF Etki Düzeyi	O/R	Örneklem Büyüklüğü	SIBTEST		Multi-MIMIC	
				1.Tip Hata	Güç	1.Tip Hata	Güç
Basit	Yüksek	1:1	600	0,41	0,91	0,19	0,95
			1200	0,12	0,90	0,08	0,95
			2400	0,01	0,90	0,01	0,95
			3600	0,00	0,90	0,00	0,95
			4800	0,00	0,90	0,00	0,95
			6000	0,00	0,90	0,00	0,95
Basit	Yüksek	1:2	600	0,42	0,90	0,19	0,96
			1200	0,16	0,90	0,09	0,95
			2400	0,02	0,90	0,02	0,95
			3600	0,00	0,90	0,00	0,95
			4800	0,00	0,90	0,00	0,95
			6000	0,00	0,90	0,00	0,95

Testin iki boyutlu, basit yapıda olduğu ve maddelerin yüksek düzeyde DMF içerdiği bu durumda 1. tip hata örneklem büyüklüğü arttıkça azalmaktadır. Her iki O/R oranında ve her iki yöntemde 2400 örneklem büyüklüğünden itibaren 1. tip hatanın 0,01 ve altında olduğu gözlenmektedir. En yüksek 1. Tip hatanın (0,42) olduğu örneklem büyüklüğü 600 olurken, en düşük sonuç 6000 (0,00) kişilik örneklemde elde edilmiştir. 1. Tip hatalar odak ve referans gruplarının oranlarına göre (1:1, 1:2) karşılaştırıldığında 1:2 olduğu koşulda hafif düzeyde artış görülmüştür. DMF yöntemine göre kıyaslandığında özellikle düşük örneklem büyüklüklerinde ciddi düşüşler gözlenmiştir. Örneğin odak ve referans grup büyüklüklerinin eşit olduğu 600 kişilik örneklem büyüklüğünde 1. tip hata 0,41 olarak ölçülürken, aynı koşullar altında Multi-MIMIC ile gerçekleştirilen iki boyutlu DMF analizinde bu değer 0,19 olarak bulunmuştur. Örneklem büyüklüğünün 2400 ve daha fazla olması durumunda 1. tip hatanın 0,05 ve altına düştüğü ancak düşük örneklem büyüklüklerinde daha yüksek olduğu belirlenmiştir.

DMF analizlerinden elde edilen istatistiksel güçler incelendiğinde ise DMF'li maddeyi belirleme gücünün %90 ile %96 arasında değiştiği belirlenmiştir. Testin tek boyutlu ya da iki boyutlu analiz edilme durumuna göre çok küçük düzeyde farklılıklar

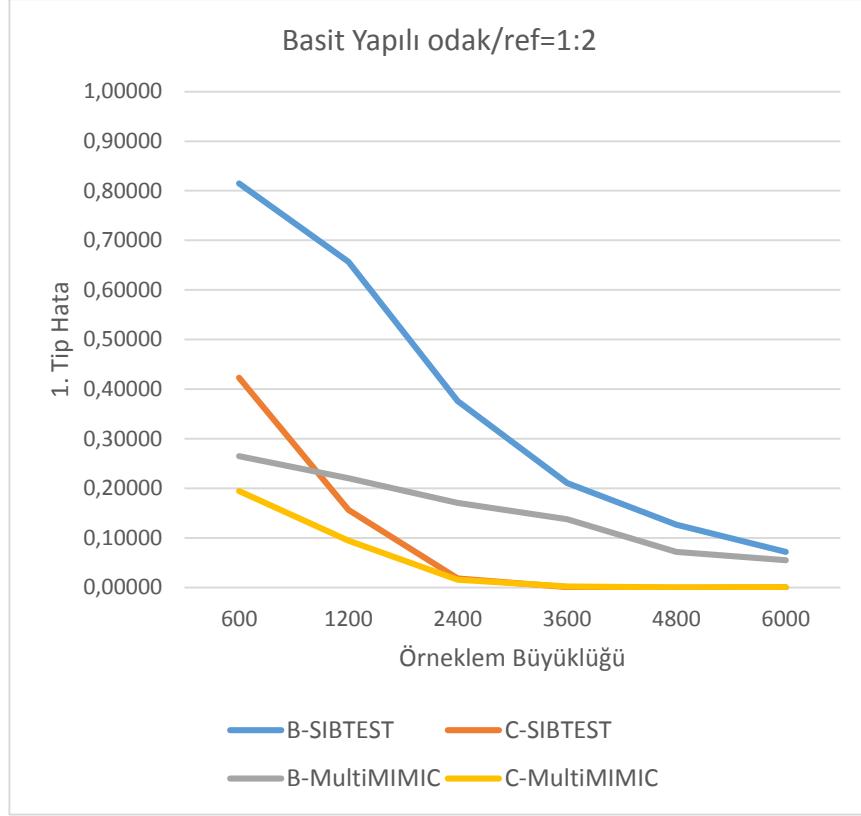
oluşmuştur. Bunun yanı sıra örneklem büyüklüğünün değişmesinin testin gücünü etkilememesi de tespit edilen bulgulardan biridir.

Analizler neticesinde odak ve referans grup oranı eşit olduğunda basit yapıli iki boyutlu testlerden elde edilen 1. tip hata değerlerinin, örneklem büyüklüğüne, DMF etki büyüklüğüne ve DMF analiz yöntemine göre nasıl değiştiğini gösteren grafikler odak ve referans grup oranlarına göre Şekil 17 ve 18’de verilmiştir. Grafikteki B-SIBTEST orta düzeyde etki büyüklüğü koşulunda SIBTEST ile yapılan analizi, B-MultiMIMIC orta düzeyde etki büyüklüğü koşulunda Multi-MIMIC yöntemi ile yapılan analizi ifade etmektedir. C-SIBTEST ve C-MultiMIMIC ise aynı analiz yöntemlerinin yüksek etki büyüklüğü düzeyinde gerçekleştiğini göstermektedir.



Şekil 17. İki boyutlu basit yapıli DMF (1:1) için 1. tip hata oranları

Şekil 17 incelendiğinde en düşük 1. tip hataların yüksek DMF etki düzeyinde 2400 ve daha yüksek örneklem büyüklüklerinde elde edildiği belirlenmiştir. En yüksek 1. tip hatalara ise orta düzey DMF etki büyüklüğüne sahip koşullarda testlerin tek boyutlu olarak analiz edilmesi ile ulaşılmıştır.



Şekil 18. İki boyutlu basit yapılu DMF (1:2) için 1. tip hata oranları

Şekil 18 incelendiğinde en düşük 1. tip hataların yüksek DMF etki düzeyinde 2400 ve daha yüksek örneklem büyüklüklerinde elde edildiği belirlenmiştir. En yüksek 1. tip hatalara ise orta düzey DMF etki büyüklüğüne sahip koşullarda testlerin tek boyutlu olarak analiz edilmesi ile ulaşılmıştır. Yöntemlerin güç değerleri ele alınan koşullara göre önemli düzeyde farklılık göstermediğinden güç değerlerine ilişkin grafikler EK-B'de verilmiştir.

Alt Problem 2'ye Ait Bulgular

a) İki boyutlu karmaşık yapılu verilere iki boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, 1. tip hata ve istatistiksel güç değerleri, DMF etki büyüklüğü orta düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

Bu alt probleme cevap bulabilmek için belirlenen koşullara uygun olarak gerçekleştirilen DMF analizleri neticesinde elde edilen 1. tip hata ve güç oranları Tablo 12'de verilmiştir.

Tablo 12

İki Boyutlu Karmaşık Yapılı ve Orta Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları

Testin Yapısı	DMF Etki Düzeyi	O/R	Örneklem Büyüklüğü	SIBTEST		Multi-MIMIC	
				1.Tip Hata	Güç	1.Tip Hata	Güç
Karmaşık	Orta	1:1	600	0,93	0,98	0,41	0,98
			1200	0,90	0,98	0,41	0,98
			2400	0,82	0,98	0,39	0,97
			3600	0,75	0,98	0,34	0,97
			4800	0,70	0,98	0,28	0,97
			6000	0,65	0,98	0,23	0,97
Karmaşık	Orta	1:2	600	0,94	0,98	0,44	0,97
			1200	0,90	0,98	0,42	0,97
			2400	0,84	0,98	0,39	0,97
			3600	0,77	0,98	0,34	0,97
			4800	0,71	0,98	0,29	0,97
			6000	0,68	0,98	0,26	0,97

Testin iki boyutlu, karmaşık yapıda olduğu ve maddelerin orta düzeyde DMF etki büyüklüğü içerdiği bu durumda 1. tip hata örneklem büyüklüğü arttıkça azalmaktadır. En yüksek 1. tip hatanın (0,94) olduğu örneklem büyüklüğü 600 olurken, en düşük sonuç 6000 (0,65) kişilik örneklemden elde edilmiştir. 1. tip hatalar odak ve referans gruplarının oranlarına göre (1:1, 1:2) karşılaştırıldığında, 1:2 olduğu koşulda hafif düzeyde artış görülmüştür. DMF yöntemine göre kıyaslandığında özellikle düşük örneklem büyüklüklerinde ciddi düşüşler gözlenmiştir. Örneğin; odak ve referans grup büyüklüklerin eşit olduğu 600 kişilik örneklem büyüklüğünde 1. tip hata 0,93 olarak ölçülürken, aynı koşullar altında Multi-MIMIC ile gerçekleştirilen iki boyutlu DMF analizinde bu değer 0,41 olarak bulunmuştur. İki boyutlu karmaşık yapı ve B düzeyinde DMF olduğu durumda hataların hiçbir koşulda 0,05'in altına düşmediği gözlenmektedir.

DMF analizlerinden elde edilen istatistiksel güçler incelendiğinde ise DMF'li maddeyi belirleme gücünün %97 ile %98 arasında değiştiği belirlenmiştir. Testin tek boyutlu ya da iki boyutlu analiz edilme durumuna göre çok küçük düzeyde farklılıklar oluşmuştur. Bunun yanı sıra örneklem büyüklüğünün değişmesinin testin gücünü etkilememesi de tespit edilen bulgulardan biridir.

b) İki boyutlu karmaşık yapıli verilere iki boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, 1. tip hata ve istatistiksel güç değerleri, DMF etki büyüklüğü yüksek düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

Bu alt probleme cevap bulabilmek için belirlenen koşullara uygun olarak gerçekleştirilen DMF analizleri neticesinde elde edilen 1. tip hata ve güç oranları Tablo 13'te verilmiştir.

Tablo 13

İki Boyutlu Karmaşık Yapılı ve Yüksek Düzeyde DMF Etki Büyüklüğüne (C düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları

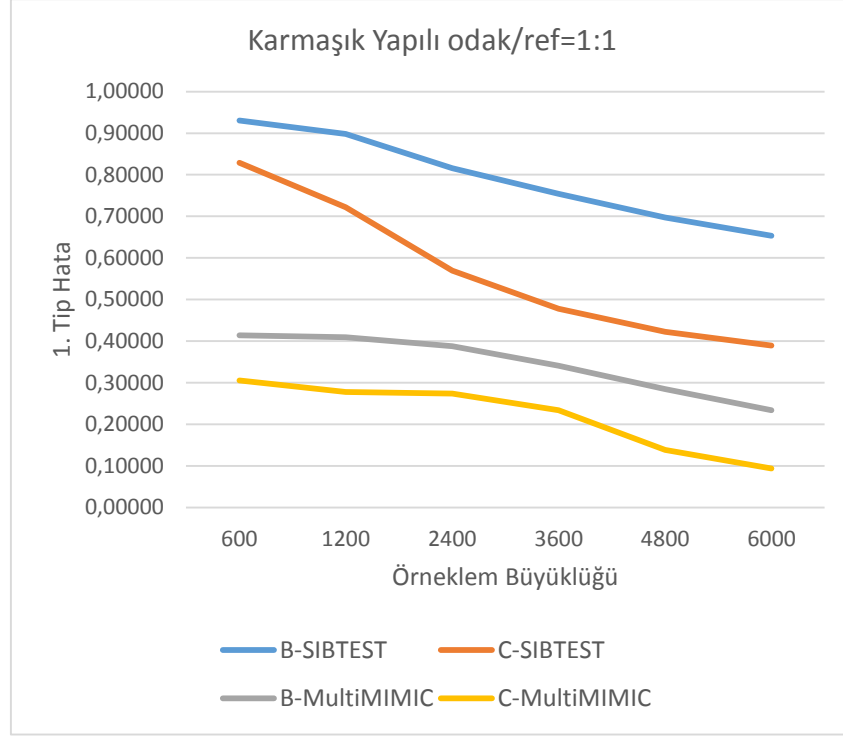
Testin Yapısı	DMF Etki Düzeyi	O/R	Örneklem Büyüklüğü	SIBTEST		Multi-MIMIC	
				1.Tip Hata	Güç	1.Tip Hata	Güç
Karmaşık	Yüksek	1:1	600	0,83	0,96	0,31	0,97
			1200	0,72	0,97	0,28	0,96
			2400	0,57	0,97	0,27	0,96
			3600	0,48	0,97	0,23	0,97
			4800	0,42	0,97	0,14	0,96
			6000	0,39	0,97	0,09	0,96
Karmaşık	Yüksek	1:2	600	0,86	0,96	0,32	0,97
			1200	0,79	0,97	0,30	0,96
			2400	0,59	0,97	0,28	0,97
			3600	0,50	0,96	0,24	0,96
			4800	0,45	0,97	0,15	0,96
			6000	0,40	0,97	0,10	0,96

Testin iki boyutlu, karmaşık yapıda olduğu ve maddelerin yüksek düzeyde DMF etki büyüklüğüne sahip olduğu bu durumda 1. tip hata örneklem büyüklüğü arttıkça azalmaktadır. En yüksek 1. tip hatanın (0,86) olduğu örneklem büyüklüğü 600 olurken, en düşük sonuç 6000 (0,39) kişilik örneklemde elde edilmiştir. 1. tip hatalar odak ve referans gruplarının oranlarına göre (1:1, 1:2) karşılaştırıldığında, 1:2 olduğu koşulda hafif düzeyde artış görülmüştür. 1. tip hatalar DMF yöntemine göre kıyaslandığında özellikle düşük örneklem büyüklüklerinde ciddi düşüşler gözlenmiştir. Örneğin odak ve referans grup büyüklüklerin eşit olduğu 600 kişilik örneklem büyüklüğünde 1. tip hata 0,83 olarak ölçülürken, aynı koşullar altında Multi-MIMIC ile gerçekleştirilen iki boyutlu DMF analizinde bu değer 0,31 olarak bulunmuştur. Bu veriler için de tüm koşullarda hataların 0,05'in altına düşmediği, en

düşük hatanın 6000 örneklem büyüklüğünde O/R oranının eşit olduğu koşulda Multi-MIMIC yönteminde olduğu (0,10) gözlenmiştir. İki boyutlu karmaşık yapı ve C düzeyinde DMF olduğu durumda hataların hiçbir koşulda 0,05'in altına düşmediği gözlenmektedir.

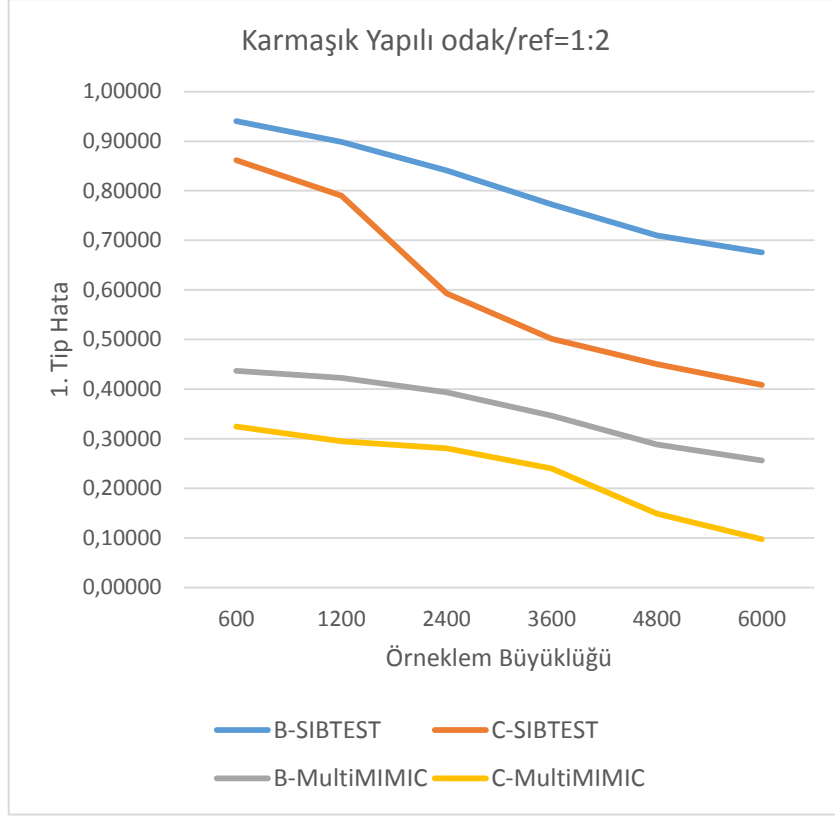
DMF analizlerinden elde edilen istatistiksel güçler incelendiğinde ise DMF'li maddeyi belirleme gücünün %96 ile %97 arasında değiştiği belirlenmiştir. Testin tek boyutlu ya da iki boyutlu analiz edilme durumuna göre çok küçük düzeyde farklılıklar oluşmuştur. Bunun yanı sıra örneklem büyüklüğünün değişmesinin testin gücünü etkilememesi de tespit edilen bulgulardan biridir.

Analizler neticesinde odak ve referans grup oranı eşit olduğunda karmaşık yapıya iki boyutlu testlerden elde edilen 1. tip hata değerlerinin, örneklem büyüklüğüne, DMF etki büyüklüğüne ve DMF analiz yöntemine göre nasıl değiştiğini gösteren grafikler odak ve referans grup oranlarına göre Şekil 19 ve Şekil 20'de verilmiştir. Grafikteki B-SIBTEST orta düzeyde etki büyüklüğü koşulunda SIBTEST ile yapılan analizi, B-MultiMIMIC orta düzeyde etki büyüklüğü koşulunda Multi-MIMIC yöntemi ile yapılan analizi ifade etmektedir. C-SIBTEST ve C-MultiMIMIC ise aynı analiz yöntemlerinin yüksek etki büyüklüğü düzeyinde gerçekleştiğini göstermektedir.



Şekil 19. İki boyutlu karmaşık yapılı DMF (1:1) için 1. tip hata oranları

Şekil 19 incelendiğinde en düşük 1. tip hataların yüksek DMF etki düzeyinde çok boyutlu DMF analiz yöntemi ile elde edildiği belirlenmiştir. En yüksek 1. tip hatalara ise orta düzey DMF etki büyüklüğüne sahip koşullarda testlerin tek boyutlu olarak analiz edilmesi ile ulaşılmıştır.



Şekil 20. İki boyutlu karmaşık yapılu DMF (1:2) için 1. tip hata oranları

Şekil 20 incelendiğinde en düşük 1. tip hataların yüksek DMF etki düzeyinde çok boyutlu DMF analiz yöntemi ile elde edildiği belirlenmiştir. En yüksek 1. tip hatalara ise orta düzey DMF etki büyüklüğüne sahip koşullarda testlerin tek boyutlu olarak analiz edilmesi ile ulaşılmıştır. Yöntemlerin güç değerleri ele alınan koşullara göre önemli düzeyde farklılık göstermediğinden, güç değerlerine ilişkin grafikler EK-B'de verilmiştir.

İki boyutlu Veri Seti İçin ANOVA Sonuçları

Araştırmada İki boyutlu yapay veri setinde 1. tip hatanın ve güç oranlarının örneklem büyüklüğüne, odak ve referans grup oranına, DMF etki düzeyine, testin basit ya da karmaşık yapıda olmasına ve analiz yöntemine göre nasıl değiştiğini belirlemek için varyans analizi yapılmıştır. Varyans analizinden elde edilen bulgular Tablo 14'te verilmiştir.

Tablo 14

1. Tip Hata ve Güç oranları için ANOVA Sonuçları

Koşul Etkileri	Birinci Tip Hata				Güç Oranları		
	<i>Sd</i>	<i>F</i>	<i>p</i>	<i>Eta</i> <i>kare</i> (η^2)	<i>F</i>	<i>p</i>	<i>Eta</i> <i>kare</i> (η^2)
Örneklem Büyüklüğü	5	3,879	0,003*	0,177	0,027	1,000	-
DMF Etki Düzeyi	1	10,461	0,002*	0,100	29,338	0,000*	0,238
Analiz Yöntemi	1	27,706	0,000*	0,228	49,603	0,000*	0,345
Yapı	1	49,603	0,000*	0,345	59,573	0,000*	0,388
Analiz Yöntemi*Yapı	1	15,425	0,000*	0,144	0,034	0,999	-
DMF Etki Düzeyi*Analiz Yöntemi	1	4,517	0,036*	0,047	65,863	0,000*	0,417

* $p < 0,05$

Tablo 14 incelendiğinde 1. tip hata için; örneklem büyüklüğü, DMF etki büyüklüğü düzeyi, testin basit veya karmaşık yapıda olması ile DMF analiz yöntemi arasındaki farklar manidar olduğu belirlenmiştir. Ayrıca analiz yöntemi ve yapı arasındaki etkileşim ile DMF etki düzeyi ve analiz yöntemi etkileşiminin de anlamlı farklılığa sebep olduğu tespit edilmiştir. Araştırma koşullarını oluşturan bağımsız değişkenlerden elde edilen anlamlı farklılığa ilişkin etki büyüklükleri eta kare (η^2) ile hesaplanmıştır. Cohen (1988) tarafından belirtilen kesme değerlerine göre yorumlanmıştır. Bu etki büyüklüğü 0,02 ile 0,05 arasında olduğunda düşük etki, 0,05 ve 0,08 arasında olduğunda orta etki, 0,08 den yüksek olduğunda yüksek düzeyde etki olarak yorumlanır (Tabachnick ve Fidell, 2013). Örneklem büyüklüğü arttıkça 1. tip hata düşmektedir. Örneklem büyüklüğünün 1. tip hatanın farklılaşması üzerindeki etki değeri 0,177 olarak hesaplanmıştır. Bu değere göre örneklem büyüklüğünün 1. tip hata üzerinde yüksek etkisinden söz edilebilir. Yapılan Post-hoc testlerinden Tukey-HSD testi uygulanmasıyla elde edilen sonuçlarda, 600 kişilik örneklem büyüklüğünün 4800 ve 6000 kişilik örneklem büyüklüklerinden manidar düzeyde farklılaştığı diğer örneklem büyüklükleri arasındaki anlamlı farkın olmadığı belirlenmiştir. Post-hoc testlerinden LSD uygulandığında elde edilen sonuçlarda ise 600 kişilik örneklem büyüklüğü 2400, 3600, 4800 ve 6000 kişilik örneklem büyüklüklerinden anlamlı düzeyde farklılaşırken, 1200 kişilik örneklem büyüklüğü 4800 ve 6000 kişilik örneklem büyüklüğüne göre farklılaşmaktadır. 600 kişilik örneklemde elde edilen hatalar diğer örneklem büyüklüklerinde elde edilen 1. tip hatalardan daha büyüktür. Ayrıca 1200 kişilik örneklem büyüklüğünden elde edilen 1. tip hatalar ise 4800 ve

6000 kişilik örneklem büyüklüklerine sahip koşullardan elde edilen 1. tip hatalardan yüksek olmuştur. Bunun yanı sıra yapılan testler neticesinde 2400, 3600, 4800 ve 6000 kişilik örneklemelerin farklılaşmasının manidar düzeyde olmadığı tespit edilmiştir. Testin DMF etki düzeyinin orta düzeyde veya yüksek düzeyde olması da 1. tip hatayı manidar düzeyde etkilemektedir. Buna göre 1. tip hata yüksek DMF etki düzeyine çıkıldıkça azalmaktadır. DMF etki düzeyinde farklılaşmanın etki değeri 0,100 dür. Bu değer de DMF etki düzeyi büyüklüğünün 1. tip hatayı yüksek düzeyde etkilediğini ifade etmektedir. DMF belirlemek için kullanılan SIBTEST ve Multi-MIMIC yöntemlerinden elde edilen 1. tip hatalar kıyaslandığında, 1. tip hatanın çok boyutlu DMF analizinde manidar düzeyde düştüğü belirlenmiştir. Analiz yöntemine göre SIBTEST ile elde edilen 1. tip hatalar ortalaması 0,46 iken Multi-MIMIC yöntemi ile 0,19 olarak belirlenmiştir. DMF analiz yöntemlerinin etki büyüklüğü değeri ise 0,228 olarak bulunmuştur. Bu etki değeri yüksek etki düzeyi olarak yorumlanabilir. İki boyutlu veri setinde 1. tip hatayı etkileyen koşullar incelendiğinde, en büyük etkiye sahip faktörün testin yapısı olduğu söylenebilir. Buna göre testin karmaşık yapıda olması durumunda 1. tip hatanın arttığı 0,345 etki değeri büyüklüğü ile belirlenmiştir. Basit yapıli testlerden elde edilen 1. tip hatalar ortalaması 0,16 iken karmaşık yapıli testlerde 0,49 olarak bulunmuştur. Testin karmaşık yapıda olması durumunda beklenen temel boyut ile yapılan açının büyümesi DMF analizlerini zorlaştırdığı ve bu nedenle 1. tip hatayı arttırdığı düşünülmektedir. Ayrıca koşul etkileşimleri incelendiğinde analiz yöntemi ve yapı arasındaki etkileşim anlamlı farklılaşmaya sebep olduğu belirlenmiştir. Eta kare değerinin 0,144 olduğu belirlenmiş ve bu değer yüksek etki düzeyini işaret etmektedir. Etkileşim değerleri incelendiğinde anlamlı farklılaşmaya sebep olan diğer ikili de DMF etki düzeyi ile analiz yöntemi olmuştur. Ancak etki değeri 0,047 olarak belirlenen bu etkileşimin etkisi düşük düzeyde kalmıştır. Koşullardan odak ve referans grubu oranına ve diğer etkileşim durumlarına göre değerlendirildiğinde anlamlı farklılaşmalar gözlenmemiştir.

Tablo 14 incelendiğinde istatistiksel güç oranları için; DMF etki büyüklüğü düzeyi, testin basit veya karmaşık yapıda olması ile DMF analiz yöntemi arasındaki farkların manidar olduğu belirlenmiştir. Ayrıca, DMF etki düzeyi ve analiz yöntemi etkileşiminin de anlamlı farklılığa sebep olduğu tespit edilmiştir. Bulgulara göre testin DMF etki düzeyinin orta düzeyden yüksek düzeye çıkması güç oranlarını

azaltmaktadır. DMF etki düzeyinde farklılaşmanın etki değeri 0,238'dir. Bu değer de DMF etki düzeyi büyüklüğünün güç oranlarını yüksek düzeyde etkilediğini ifade etmektedir. DMF analiz yönteminde Multi-MIMIC yönteminin tercih edilmesi ise güç oranlarını manidar düzeyde yükseltmektedir. Analiz yöntemine göre SIBTEST ile elde edilen güç oranları ortalaması 0,95 iken Multi-MIMIC yöntemi ile 0,96 olarak belirlenmiştir. DMF analiz yöntemlerinin etki büyüklüğü değeri ise 0,228 olarak bulunmuştur. Bu etki değeri yüksek etki düzeyi olarak yorumlanabilir. Testin yapısının basitten karmaşığa geçmesi DMF'li madde belirleme gücünde artışa neden olmuştur. Testin basit yapılı olması durumunda elde edilen güç oranları ortalaması 0,94 olurken, karmaşık yapılı olması durumunda bu ortalamanın 0,97'ye yükseldiği tespit edilmiştir. Testin yapısının güç oranları üzerindeki etki büyüklüğü ($\eta^2 = 0,388$) yüksek olarak belirlenmiştir. Ayrıca DMF etki düzeyi ve analiz yöntemi etkileşiminin de anlamlı farklılığa sebep olduğu belirlenmiştir. Etki değeri 0,417 olarak belirlenmiş ve yüksek etki değeri olarak yorumlanmıştır. DMF belirleme gücü oranları açısından varyans analizi bulguları incelendiğinde 1. tip hatadan farklı olarak örneklem büyüklüğünün manidar düzeyde etkiye sahip olmadığı tespit edilmiştir ($F=0,027$, $\eta^2=0,002$). Odak ve referans grup oranında da anlamlı farklılaşma yoktur ($F=0,015$, $\eta^2=0,000$). Ayrıca diğer etkileşim durumlarına göre değerlendirildiğinde anlamlı farklılaşmalar gözlenmemiştir.

Analizler neticesinde odak ve referans grup oranının 1:1 ve 1:2 olduğunda basit ve karmaşık yapılı iki boyutlu testlerden elde edilen istatistiksel güç oranlarının, örneklem büyüklüğüne, DMF etki büyüklüğüne ve DMF analiz yöntemine göre nasıl değiştiğini gösteren grafikler güç oranları arasındaki çok büyük değişim gözlenmediğinden EK-B'de verilmiştir.

Alt Problem 3'e Ait Bulgular

a) Üç boyutlu basit yapılı verilere üç boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, 1. tip hata ve istatistiksel güç değerleri, DMF etki büyüklüğü orta düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

Bu alt probleme cevap bulabilmek için belirlenen koşullara uygun olarak gerçekleştirilen DMF analizleri neticesinde elde edilen 1. tip hata ve güç oranları Tablo 15'te verilmiştir.

Tablo 15

Üç Boyutlu Basit Yapılı ve Orta Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları

Testin Yapısı	DMF Etki Düzeyi	O/R	Örneklem Büyüklüğü	SIBTEST		Multi-MIMIC	
				1.Tip Hata	Güç	1.Tip Hata	Güç
Basit	Orta	1:1	600	0,53	0,92	0,23	0,95
			1200	0,36	0,92	0,19	0,94
			2400	0,29	0,92	0,10	0,94
			3600	0,28	0,92	0,07	0,94
			4800	0,26	0,92	0,06	0,94
			6000	0,25	0,92	0,05	0,94
Basit	Orta	1:2	600	0,57	0,92	0,23	0,94
			1200	0,37	0,92	0,20	0,94
			2400	0,30	0,92	0,10	0,95
			3600	0,28	0,92	0,08	0,94
			4800	0,27	0,92	0,06	0,94
			6000	0,26	0,92	0,06	0,94

Testin üç boyutlu, basit yapıda olduğu ve maddelerin orta düzeyde DMF etki büyüklüğü içerdiği bu durumda 1. tip hata örneklem büyüklüğü arttıkça azalmaktadır. En yüksek 1. tip hatanın (0,57) olduğu örneklem büyüklüğü 600 olurken, en düşük sonuç 6000 (0,05) kişilik örneklemden elde edilmiştir. 1. tip hatalar odak ve referans gruplarının oranlarına göre (1:1, 1:2) karşılaştırıldığında, 1:2 olduğu koşulda hafif düzeyde artış görülmüştür. DMF yöntemine göre kıyaslandığında ise özellikle düşük örneklem büyüklüklerinde ciddi düşüşler gözlenmiştir. Örneğin odak ve referans grup büyüklüklerin eşit olduğu 600 kişilik örneklem büyüklüğünde 1. tip hata 0,53 olarak ölçülürken, aynı koşullar altında Multi-MIMIC ile gerçekleştirilen iki boyutlu DMF analizinde bu değer 0,23 olarak bulunmuştur. Üç boyutlu basit yapı ve B düzeyinde DMF olduğu durumda örnekleme ve yöntemine göre fark olmasına rağmen 1. tip hatanın bir koşul hariç 0,05 ve altına düşmediği ve yüksek olduğu belirlenmiştir.

DMF analizlerinden elde edilen istatistiksel güçler incelendiğinde ise DMF'li maddeyi belirleme gücünün %92 ile %95 arasında değiştiği belirlenmiştir. Testin tek boyutlu ya da iki boyutlu analiz edilme durumuna göre çok küçük düzeyde farklılıklar

oluşmuştur. Bunun yanı sıra örneklem büyüklüğünün değişmesinin testin gücünü etkilememesi de tespit edilen bulgulardan biridir.

b) Üç boyutlu basit yapıli verilere üç boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, 1. tip hata ve istatistiksel güç değerleri, DMF etki büyüklüğü yüksek düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

Bu alt probleme cevap bulabilmek için belirlenen koşullara uygun olarak gerçekleştirilen DMF analizleri neticesinde elde edilen 1. tip hata ve güç oranları Tablo 16'da verilmiştir.

Tablo 16

Üç Boyutlu Basit Yapılı ve Yüksek Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları

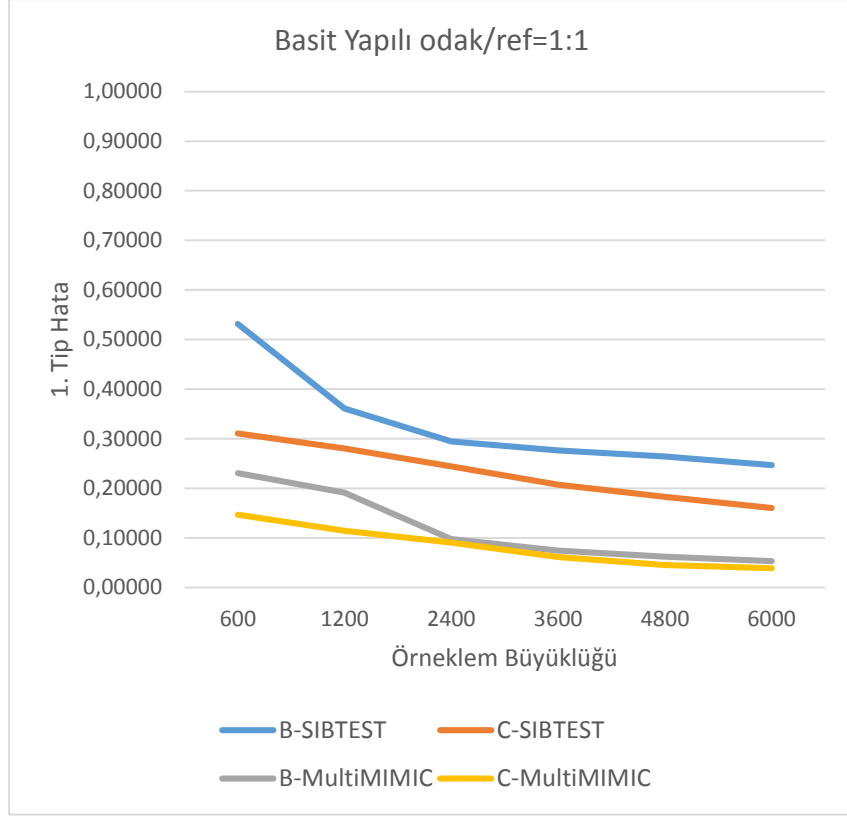
Testin Yapısı	DMF Etki Düzeyi	O/R	Örneklem	SIBTEST		Multi-MIMIC	
				1.Tip Hata	Güç	1.Tip Hata	Güç
Basit	Yüksek	1:1	600	0,31	0,84	0,15	0,93
			1200	0,28	0,85	0,11	0,93
			2400	0,24	0,84	0,09	0,93
			3600	0,21	0,85	0,06	0,93
			4800	0,18	0,85	0,05	0,93
			6000	0,16	0,85	0,04	0,93
Basit	Yüksek	1:2	600	0,31	0,84	0,15	0,93
			1200	0,28	0,85	0,12	0,93
			2400	0,25	0,84	0,09	0,93
			3600	0,22	0,84	0,07	0,93
			4800	0,19	0,85	0,05	0,93
			6000	0,16	0,85	0,04	0,94

Testin üç boyutlu, basit yapıda olduğu ve maddelerin yüksek düzeyde DMF etki büyüklüğü içerdiği bu durumda 1. tip hata örneklem büyüklüğü arttıkça azalmaktadır. En yüksek 1. tip hatanın (0,31) olduğu örneklem büyüklüğü 600 olurken, en düşük sonuç 6000 (0,04) kişilik örneklemde elde edilmiştir. 1. tip hatalar odak ve referans gruplarının oranlarına göre (1:1, 1:2) karşılaştırıldığında, 1:2 olduğu koşulda hafif düzeyde artış görülmüştür. 1. tip hatalar DMF yöntemine göre kıyaslandığında özellikle düşük örneklem büyüklüklerinde ciddi düşüşler gözlenmiştir. Örneğin odak ve referans grup büyüklüklerin eşit olduğu 600 kişilik

örneklem büyüklüğünde 1. tip hata 0,31 olarak ölçülürken, aynı koşullar altında Multi-MIMIC ile gerçekleştirilen iki boyutlu DMF analizinde bu değer 0,15 olarak bulunmuştur. Üç boyutlu basit yapı ve C düzeyinde DMF olduğu durumda 1. tip hatanın 4 koşul hariç 0,05 ve altına düşmediği ve yüksek olduğu belirlenmiştir. Bu düşük 1. tip hatalar da çok boyutlu çoklu nedenler çoklu göstergeler yönteminde 4800 ve 6000 kişilik örneklemelerden elde edilmiştir.

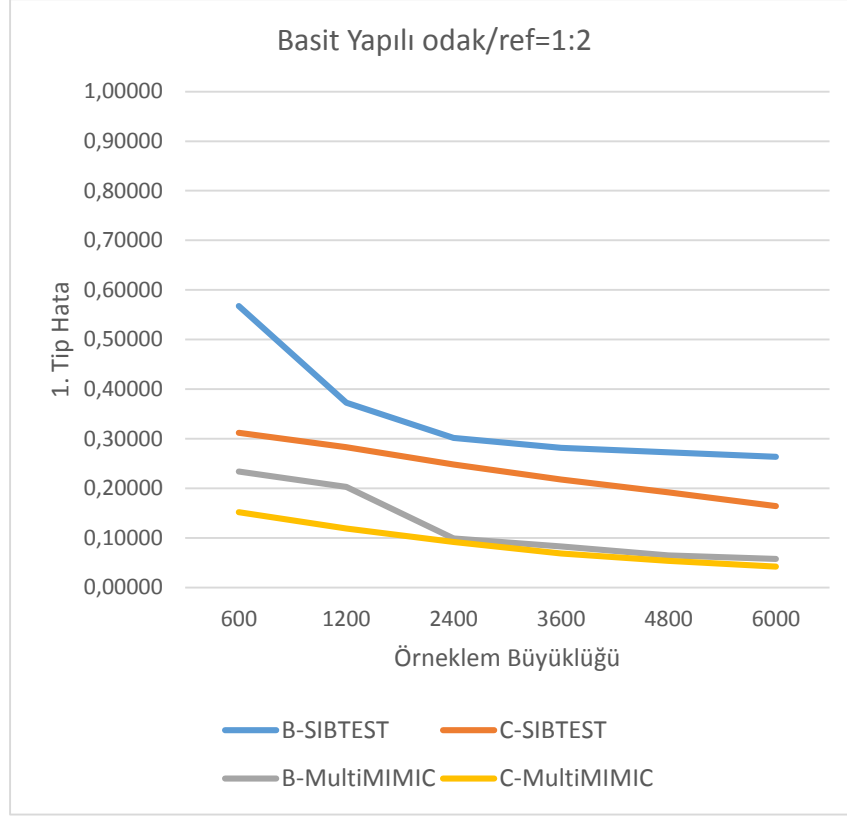
DMF analizlerinden elde edilen istatistiksel güçler incelendiğinde ise DMF'li maddeyi belirleme gücünün %84 ile %94 arasında değiştiği belirlenmiştir. Testin tek boyutlu ya da iki boyutlu analiz edilme durumuna göre çok küçük düzeyde farklılıklar oluşmuştur. Bunun yanı sıra örneklem büyüklüğünün değişmesinin testin gücünü etkilememesi de tespit edilen bulgulardan biridir.

Analizler neticesinde odak ve referans grup oranı eşit olduğunda basit yapıli üç boyutlu testlerden elde edilen 1. tip hata değerlerinin, örneklem büyüklüğüne, DMF etki büyüklüğüne ve DMF analiz yöntemine göre nasıl değiştiğini gösteren grafikler odak ve referans grup oranlarına göre Şekil 21 ve 22'de verilmiştir. Grafikteki B-SIBTEST orta düzeyde etki büyüklüğü koşulunda SIBTEST ile yapılan analizi, B-MultiMIMIC orta düzeyde etki büyüklüğü koşulunda Multi-MIMIC yöntemi ile yapılan analizi ifade etmektedir. C-SIBTEST ve C-MultiMIMIC ise aynı analiz yöntemlerinin yüksek etki büyüklüğü düzeyinde gerçekleştiğini göstermektedir.



Şekil 21. Üç boyutlu basit yapılu DMF (1:1) için 1. tip hata oranları

Şekil 21 incelendiğinde en düşük 1. tip hataların yüksek DMF etki düzeyinde 6000 kişilik örneklem büyüklüklerinde Multi-MIMIC yöntemi ile elde edildiği belirlenmiştir. En yüksek 1. tip hatalara ise orta düzey DMF etki büyüklüğüne sahip koşullarda testlerin tek boyutlu olarak analiz edilmesi ile ulaşılmıştır.



Şekil 22. Üç boyutlu basit yapılu DMF (1:2) için 1. tip hata oranları

Şekil 22 incelendiğinde en düşük 1. tip hataların yüksek DMF etki düzeyinde 6000 kişilik örneklem büyüklüklerinde Multi-MIMIC yöntemi ile elde edildiği belirlenmiştir. En yüksek 1. tip hatalara ise orta düzey DMF etki büyüklüğüne sahip koşullarda testlerin tek boyutlu olarak analiz edilmesi ile ulaşılmıştır. Yöntemlerin güç değerleri ele alınan koşullara göre önemli düzeyde farklılık göstermediğinden, güç değerlerine ilişkin grafikler EK-B'de verilmiştir.

Alt Problem 4'e Ait Bulgular

a) Üç boyutlu karmaşık yapılu verilere üç boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, 1. tip hata ve istatistiksel güç değerleri, DMF etki büyüklüğü orta düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

Bu alt probleme cevap bulabilmek için belirlenen koşullara uygun olarak gerçekleştirilen DMF analizleri neticesinde elde edilen 1. tip hata ve güç oranları Tablo 17'de verilmiştir.

Tablo 17

Üç Boyutlu Karmaşık Yapılı ve Orta Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları

Testin Yapısı	DMF Etki Düzeyi	O/R	Örneklem	SIBTEST		Multi-MIMIC	
				1.Tip Hata	Güç	1.Tip Hata	Güç
Karmaşık	Orta	1:1	600	0,73	0,95	0,38	0,97
			1200	0,51	0,95	0,28	0,96
			2400	0,34	0,95	0,16	0,96
			3600	0,29	0,95	0,12	0,96
			4800	0,26	0,95	0,08	0,96
			6000	0,23	0,95	0,09	0,96
Karmaşık	Orta	1:2	600	0,72	0,95	0,39	0,96
			1200	0,54	0,95	0,29	0,96
			2400	0,37	0,95	0,16	0,96
			3600	0,30	0,95	0,12	0,96
			4800	0,27	0,95	0,09	0,96
			6000	0,24	0,95	0,09	0,95

Testin üç boyutlu, karmaşık yapıda olduğu ve maddelerin orta düzeyde DMF etki büyüklüğü içerdiği bu durumda 1. tip hata örneklem büyüklüğü arttıkça azalmaktadır. En yüksek 1. tip hatanın (0,73) olduğu örneklem büyüklüğü 600 olurken, en düşük sonuç 6000 (0,08) kişilik örneklemden elde edilmiştir. 1. tip hatalar odak ve referans gruplarının oranlarına göre (1:1, 1:2) karşılaştırıldığında, 1:2 olduğu koşulda hafif düzeyde artış görülmüştür. DMF yöntemine göre kıyaslandığında ise özellikle düşük örneklem büyüklüklerinde ciddi düşüşler gözlenmiştir. Örneğin odak ve referans grup büyüklüklerin eşit olduğu 600 kişilik örneklem büyüklüğünde 1. tip hata 0,73 olarak ölçülürken, aynı koşullar altında Multi-MIMIC ile gerçekleştirilen iki boyutlu DMF analizinde bu değer 0,38 olarak bulunmuştur. Üç boyutlu karmaşık yapı ve B düzeyinde DMF olduğu durumda 1. tip hatanın hiçbir koşulda 0,05 ve altına düşmediği ve yüksek olduğu belirlenmiştir.

DMF analizlerinden elde edilen istatistiksel güçler incelendiğinde ise DMF'li maddeyi belirleme gücünün %95 ile %97 arasında değiştiği belirlenmiştir. Testin tek boyutlu ya da iki boyutlu analiz edilme durumuna göre çok küçük düzeyde farklılıklar oluşmuştur. Bunun yanı sıra örneklem büyüklüğünün değişmesinin testin gücünü etkilememesi de tespit edilen bulgulardan biridir.

b) Üç boyutlu karmaşık yapıli verilere üç boyutlu ve tek boyutlu olarak DMF analizleri yapıldığında, 1. tip hata ve istatistiksel güç değerleri, DMF etki büyüklüğü yüksek düzeyde olduğunda örneklem büyüklüklerine (600, 1200, 2400, 3600, 4800, 6000) ve odak-referans gruplarının oranına (1:1, 1:2) göre nasıl değişmektedir?

Bu alt probleme cevap bulabilmek için belirlenen koşullara uygun olarak gerçekleştirilen DMF analizleri neticesinde elde edilen 1. tip hata ve güç oranları Tablo 18'de verilmiştir.

Tablo 18

İki Boyutlu Karmaşık Yapılı ve Yüksek Düzeyde DMF Etki Büyüklüğüne (B düzeyi) Sahip Veri Seti için 1. Tip Hata ve Güç Oranları

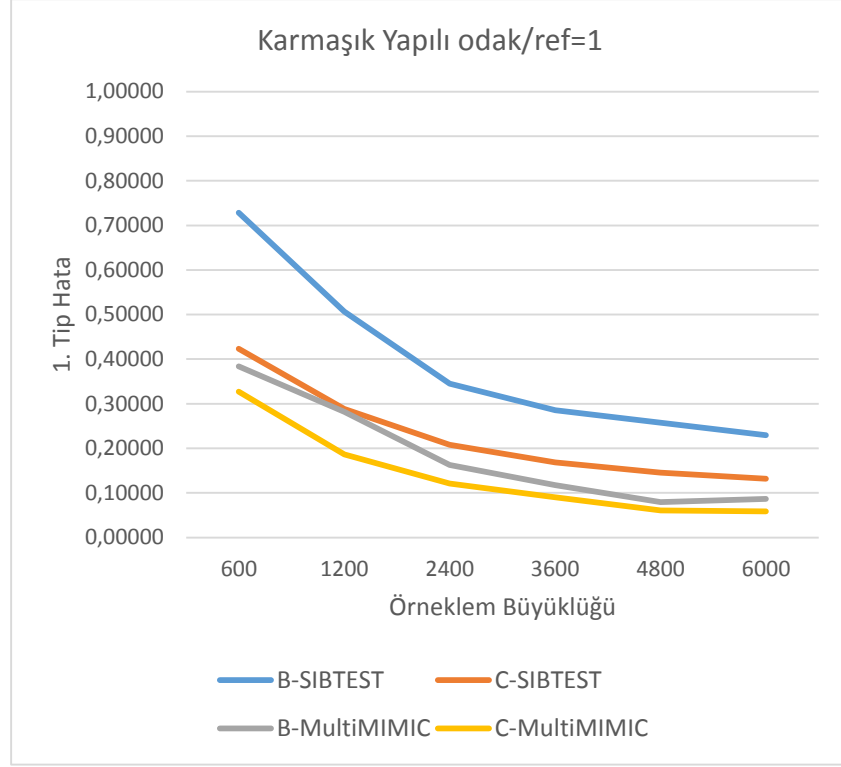
Testin Yapısı	DMF Etki Düzeyi	O/R	Örneklem	SIBTEST		Multi-MIMIC	
				1.Tip Hata	Güç	1.Tip Hata	Güç
Karmaşık	Yüksek	1:1	600	0,42	0,89	0,33	0,96
			1200	0,29	0,90	0,19	0,96
			2400	0,21	0,90	0,12	0,95
			3600	0,17	0,90	0,09	0,95
			4800	0,15	0,90	0,06	0,95
			6000	0,13	0,90	0,06	0,95
Karmaşık	Yüksek	1:2	600	0,44	0,89	0,33	0,95
			1200	0,30	0,89	0,19	0,96
			2400	0,21	0,89	0,12	0,95
			3600	0,18	0,90	0,09	0,95
			4800	0,19	0,90	0,06	0,95
			6000	0,13	0,90	0,06	0,94

Testin üç boyutlu, karmaşık yapıda olduğu ve maddelerin yüksek düzeyde DMF etki büyüklüğüne sahip olduğu bu durumda 1. tip hata örneklem büyüklüğü arttıkça azalmaktadır. En yüksek 1. tip hatanın (0,44) olduğu örneklem büyüklüğü 600 olurken, en düşük sonuç 6000 (0,06) kişilik örneklemde elde edilmiştir. 1. tip hatalar odak ve referans gruplarının oranlarına göre (1:1, 1:2) karşılaştırıldığında, 1:2 olduğu koşulda hafif düzeyde artış görülmüştür. 1. tip hatalar DMF yöntemine göre kıyaslandığında özellikle düşük örneklem büyüklüklerinde ciddi düşüşler gözlenmiştir. Örneğin odak ve referans grup büyüklüklerinin eşit olduğu 600 kişilik örneklem büyüklüğünde 1. tip hata 0,42 olarak ölçülürken, aynı koşullar altında Multi-MIMIC ile gerçekleştirilen iki boyutlu DMF analizinde bu değer 0,33 olarak

bulunmuştur. Üç boyutlu basit yapı ve B düzeyinde DMF olduğu durumda 1. tip hatanın hiçbir koşulda 0,05 ve altına düşmediği ve yüksek olduğu belirlenmiştir. Ancak Multi-MIMIC yönteminin kullanıldığı 4 koşulda 0,06 olduğu görülmektedir.

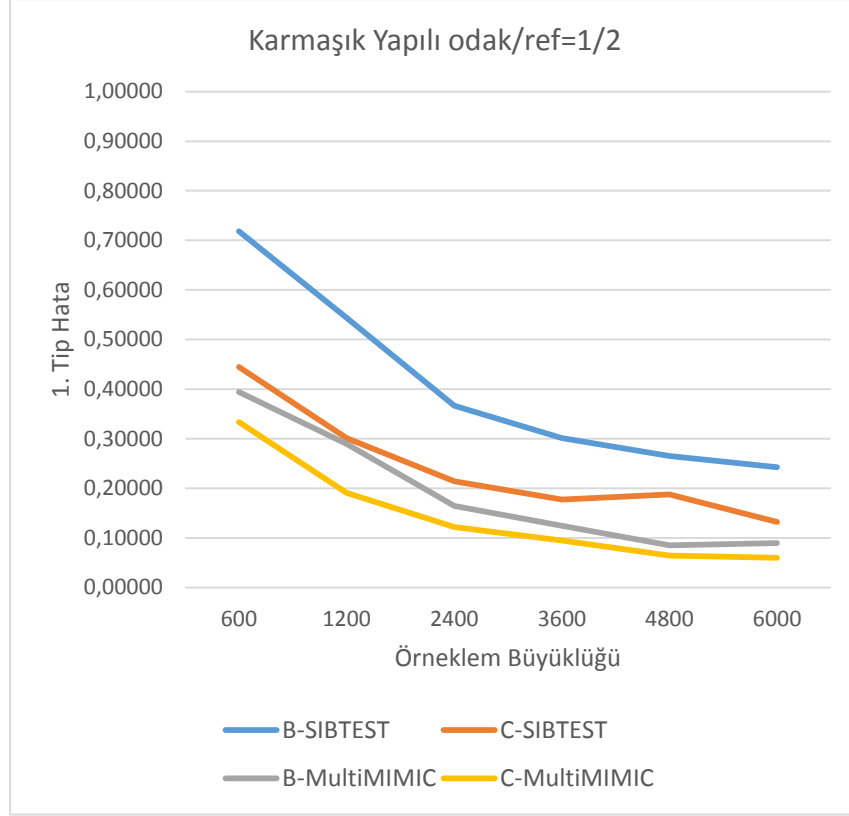
DMF analizlerinden elde edilen istatistiksel güçler incelendiğinde ise DMF'li maddeyi belirleme gücünün %89 ile %96 arasında değiştiği belirlenmiştir. Testin tek boyutlu ya da iki boyutlu analiz edilme durumuna göre çok küçük düzeyde farklılıklar oluşmuştur. Bunun yanı sıra örneklem büyüklüğünün değişmesinin testin gücünü etkilememesi de tespit edilen bulgulardan biridir.

Analizler neticesinde odak ve referans grup oranı eşit olduğunda karmaşık yapıları üç boyutlu testlerden elde edilen 1. tip hata değerlerinin, örneklem büyüklüğüne, DMF etki büyüklüğüne ve DMF analiz yöntemine göre nasıl değiştiğini gösteren grafikler odak ve referans grup oranlarına göre Şekil 23 ve Şekil 24'te verilmiştir. Grafikteki B-SIBTEST orta düzeyde etki büyüklüğü koşulunda SIBTEST ile yapılan analizi, B-MultiMIMIC orta düzeyde etki büyüklüğü koşulunda Multi-MIMIC yöntemi ile yapılan analizi ifade etmektedir. C-SIBTEST ve C-MultiMIMIC ise aynı analiz yöntemlerinin yüksek etki büyüklüğü düzeyinde gerçekleştiğini göstermektedir.



Şekil 23. Üç boyutlu karmaşık yapı DMF (1:1) için 1. tip hata oranları

Şekil 23 incelendiğinde en düşük 1. tip hataların yüksek DMF etki düzeyinde 4800 ve daha yüksek örneklem büyüklüklerinde çok boyutlu DMF analiz yöntemi ile elde edildiği belirlenmiştir. En yüksek 1. tip hatalara ise orta düzey DMF etki büyüklüğüne ve düşük örneklem büyüklüğüne sahip koşullarda testlerin tek boyutlu olarak analiz edilmesi ile ulaşılmıştır.



Şekil 24. Üç boyutlu karmaşık yapılu DMF (1:2) için 1. tip hata oranları

Şekil 24 incelendiğinde en düşük 1. tip hataların yüksek DMF etki düzeyinde 6000 kişilik örneklem büyüklüklerinde çok boyutlu DMF analiz yöntemi ile elde edildiği belirlenmiştir. En yüksek 1. tip hatalara ise orta düzey DMF etki büyüklüğüne ve düşük örneklem büyüklüğüne sahip koşullarda testlerin tek boyutlu olarak analiz edilmesi ile ulaşılmıştır. Yöntemlerin güç değerleri ele alınan koşullara göre önemli düzeyde farklılık göstermediğinden, güç değerlerine ilişkin grafikler EK-B'de verilmiştir.

Üç boyutlu Veri Seti İçin ANOVA Sonuçları

Araştırmada üç boyutlu yapay veri setinde 1. tip hatanın ve güç oranlarının örneklem büyüklüğüne, odak ve referans grup oranına, DMF etki düzeyine, testin basit ya da karmaşık yapıda olmasına ve analiz yöntemine göre nasıl değiştiğini belirlemek için varyans analizi yapılmıştır. Varyans analizinden elde edilen bulgular Tablo 19'da verilmiştir.

Tablo 19

1. Tip Hata ve Güç oranları için ANOVA Sonuçları

Koşul Etkileri	Birinci Tip Hata				Güç Oranları		
	<i>Sd</i>	<i>F</i>	<i>p</i>	<i>Eta</i> <i>kare</i> (η^2)	<i>F</i>	<i>p</i>	<i>Eta</i> <i>kare</i> (η^2)
Örneklem Büyüklüğü	5	12,766	0,000*	0,415	0,005	1,000	-
DMF Etki Düzeyi	1	9,629	0,003*	0,093	34,526	0,000*	0,269
Analiz Yöntemi	1	48,884	0,000*	0,342	58,701	0,000*	0,384
Yapı	1	2,655	0,107	-	18,300	0,000*	0,163
Analiz Yöntemi*Yapı	1	0,402	0,527	-	4,197	0,043*	0,044
DMF Etki Düzeyi*Analiz Yöntemi	1	4,517	0,036*	0,047	65,863	0,000*	0,417

* $p < 0,05$

Tablo 19 incelendiğinde 1. tip hata için; örneklem büyüklüğü, DMF etki büyüklüğü düzeyi ile DMF analiz yöntemi arasındaki farkların manidar olduğu belirlenmiştir. Ayrıca DMF etki düzeyi ve analiz yöntemi etkileşiminin de anlamlı farklılığa sebep olduğu tespit edilmiştir. Araştırma koşullarını oluşturan bağımsız değişkenlerden elde edilen anlamlı farklılığa ilişkin etki büyüklükleri eta kare (η^2) ile hesaplanmıştır. Cohen (1988) tarafından belirtilen kesme değerlerine göre yorumlanmıştır. Bu etki büyüklüğü 0,02 ile 0,05 arasında olduğunda düşük etki, 0,05 ve 0,08 arasında olduğunda orta etki, 0,08 den yüksek olduğunda yüksek düzeyde etki olarak yorumlanır (Tabachnick ve Fidell, 2013). Örneklem büyüklüğü arttıkça 1. tip hata düşmektedir. Örneklem büyüklüğünün 1. tip hatanın farklılaşması üzerindeki etki değeri 0,415 olarak hesaplanmıştır. Bu değere göre örneklem büyüklüğünün 1. tip hata üzerinde yüksek etkisinden söz edilebilir. Yapılan Post-hoc testlerinden Tukey-HSD testi uygulanmasıyla elde edilen sonuçlarda, 600 kişilik örneklem büyüklüğünün 2400, 3600, 4800 ve 6000 kişilik örneklem büyüklüklerinden manidar düzeyde farklılaştığı belirlenmiştir. 600 kişilik örneklemde elde edilen hatalar diğer örneklem büyüklüklerinde elde edilen 1. tip hatalardan daha büyüktür. Ayrıca 1200 kişilik örneklem büyüklüğünden elde edilen 1. tip hatalar ise 3600, 4800 ve 6000 kişilik örneklem büyüklüklerine sahip koşullardan elde edilen 1. tip hatalardan yüksek olmuştur. Post-hoc testlerinden LSD testi uygulandığında elde edilen sonuçlarda ise 600 ve 1200 kişilik örneklem büyüklüklerinin tüm örneklem büyüklükleri ile anlamlı düzeyde farklılaştığı belirlenmiştir. 600 kişilik ve 1200 kişilik

örneklerde elde edilen hatalar diğer örneklem büyüklüklerinde elde edilen 1. tip hatalardan daha büyüktür. Bunun yanı sıra yapılan bu testler neticesinde 2400, 3600, 4800 ve 6000 kişilik örneklemelerin kendi aralarında farklılaşmalarının manidar düzeyde olmadığı tespit edilmiştir. Testin DMF etki düzeyinin orta düzeyde veya yüksek düzeyde olması da 1. tip hatayı manidar düzeyde etkilemektedir. Buna göre 1. tip hata yüksek DMF etki düzeyine çıktıkça azalmaktadır. DMF etki düzeyinde farklılaşmanın etki değeri 0,093 dür. Bu değer de DMF etki düzeyi büyüklüğünün 1. tip hatayı yüksek düzeyde etkilediğini ifade etmektedir. DMF belirlemek için kullanılan SIBTEST ve Multi-MIMIC yöntemlerinden elde edilen 1. tip hatalar kıyaslandığında, 1. tip hatanın çok boyutlu DMF analizinde manidar düzeyde düştüğü belirlenmiştir. Analiz yöntemine göre SIBTEST ile elde edilen 1. tip hatalar ortalaması 0,30 iken Multi-MIMIC yöntemi ile 0,11 olarak belirlenmiştir. DMF analiz yöntemlerinin etki büyüklüğü değeri ise 0,342 olarak bulunmuştur. Bu etki değeri yüksek etki düzeyi olarak yorumlanabilir. Ayrıca koşul etkileşimleri incelendiğinde DMF etki büyüklüğü düzeyi ve analiz yöntemi etkileşiminin anlamlı farklılaşmaya sebep olduğu belirlenmiştir. Ancak eta kare değerinin 0,047 olduğu belirlenmiş ve bu değer düşük etki düzeyini işaret etmektedir. Koşullardan odak ve referans grubu oranına, testin yapısına ve diğer etkileşim durumlarına göre değerlendirildiğinde anlamlı farklılaşmalar gözlenmemiştir.

Tablo 19 incelendiğinde istatistiksel güç oranları için; DMF etki büyüklüğü düzeyi, testin basit veya karmaşık yapıda olması ile DMF analiz yöntemi arasındaki farkların manidar olduğu belirlenmiştir. Ayrıca testin yapısı ve analiz yöntemi etkileşimi ile DMF etki düzeyi ve analiz yöntemi etkileşiminin anlamlı farklılığa sebep olduğu tespit edilmiştir. Bulgulara göre testin DMF etki düzeyinin orta düzeyden yüksek düzeye çıkması güç oranlarını azaltmaktadır. DMF etki düzeyinde farklılaşmanın etki değeri 0,269'dur. Bu değer de DMF etki düzeyi büyüklüğünün güç oranlarını yüksek düzeyde etkilediğini ifade etmektedir. DMF analiz yönteminde Multi-MIMIC yönteminin tercih edilmesi ise güç oranlarını manidar düzeyde yükseltmektedir. Analiz yöntemine göre SIBTEST ile elde edilen güç oranları ortalaması 0,90 iken Multi-MIMIC yöntemi ile 0,95 olarak belirlenmiştir. DMF analiz yöntemlerinin etki büyüklüğü değeri ise 0,384 olarak bulunmuştur. Bu etki değeri yüksek etki düzeyi olarak yorumlanabilir. Testin yapısının basitten karmaşığa geçmesi DMF'li madde belirleme gücünde artışa neden olmuştur. Testin basit yapıları

olması durumunda elde edilen güç oranları ortalaması 0,91 olurken, karmaşık yapılı olması durumunda bu ortalamanın 0,94'e yükseldiği tespit edilmiştir. Testin yapısının güç oranları üzerindeki etki büyüklüğü ($\eta^2 = 0,163$) yüksek olarak belirlenmiştir. DMF etki düzeyi ve analiz yöntemi etkileşiminin etki değeri 0,417 olarak belirlenmiş ve yüksek etki değeri olarak yorumlanmıştır. Analiz yöntemi ve yapı arasındaki etkileşimin de düşük de olsa etki büyüklüğüne sahip olduğu belirlenmiştir. DMF belirleme gücü oranları açısından varyans analizi bulguları incelendiğinde örneklem büyüklüğünün, odak ve referans grup oranının ve diğer etkileşimlerin manidar düzeyde etkiye sahip olmadığı tespit edilmiştir.

1. tip hata ve güç oranları için koşullar arasındaki etkileşim etkileri değerlendirildiğinde manidar bir farklılık gözlenmemiş ve bu sebeple tabloda yer verilmemiştir.

Analizler neticesinde odak ve referans grup oranının 1:1 ve 1:2 olduğunda basit ve karmaşık yapılı üç boyutlu testlerden elde edilen istatistiksel güç oranlarının, örneklem büyüklüğüne, DMF etki büyüklüğüne ve DMF analiz yöntemine göre nasıl değiştiğini gösteren grafikler güç oranları arasındaki çok büyük değişim gözlenmediğinden EK-B'de verilmiştir.

Alt Problem 5'e Ait Bulgular

Gerçek veri setinden elde edilen, iki boyutlu veri, tek boyutlu (SIBTEST) ve iki boyutlu olarak analiz edildiğinde (MULTI-MIMIC) öğrencilerin cinsiyeti açısından, farklı örneklem büyüklüklerinde (600, 1200, 3600), odak grup ve referans grupları oranlarında (1:1, 1:2) DMF gösteren madde sayısı nasıl değişmektedir?

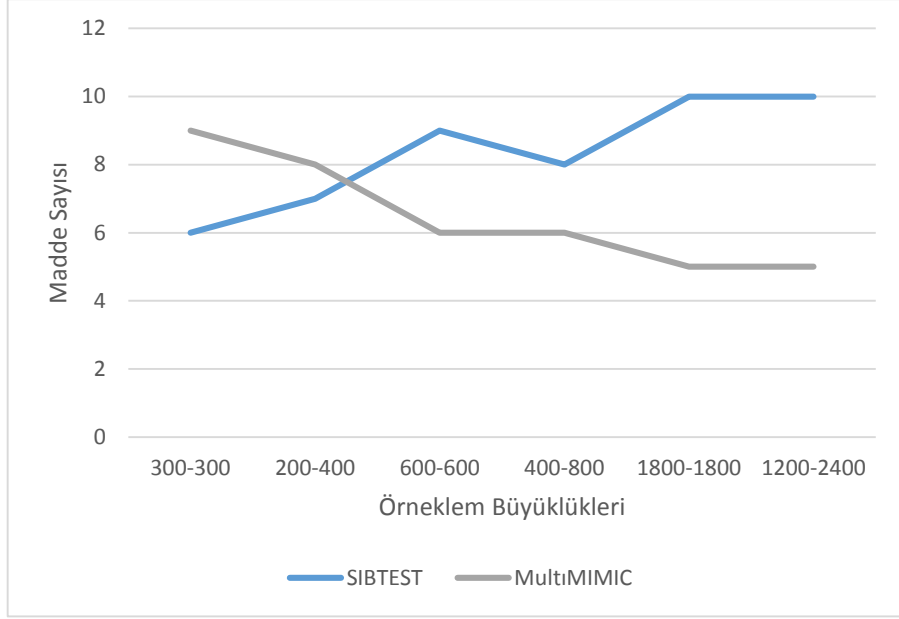
Bu alt probleme cevap bulabilmek için belirlenen koşullara uygun olarak gerçekleştirilen DMF analizleri neticesinde DMF gösteren madde sayıları Tablo 20'de verilmiştir.

Tablo 20

PISA Verisi DMF Analizi Sonuçları

Örneklem Büyüklüğü (Odak-Referans)	DMF Büyüklüğü	SIBTEST	Multi-MIMIC
600 (300-300)	B	3	
	C	3	
	Toplam	6	9
600 (200-400)	B	5	
	C	2	
	Toplam	7	8
1200 (600-600)	B	5	
	C	4	
	Toplam	9	6
1200 (400-800)	B	4	
	C	4	
	Toplam	8	6
3600 (1800-1800)	B	5	
	C	5	
	Toplam	10	5
3600 (1200-2400)	B	6	
	C	4	
	Toplam	10	5

PISA 2012 uygulamasından seçilen 60 maddeye DMF analizleri yapıldığında, farklı örneklem büyüklükleri ve örneklem büyüklükleri oranına göre, öğrencilerin cinsiyeti açısından DMF gösteren madde sayıları Tablo 20’de verilmiştir. SIBTEST yöntemi ile belirlenen DMF gösteren madde sayıları orta düzeyde DMF gösteren (B) ve yüksek düzeyde DMF gösteren (C) olarak ikiye ayrılmaktadır. Düşük düzeyde DMF gösteren madde sayıları ise ihmal edilebilir düzey olduğundan dikkate alınmamıştır. Multi-MIMIC yöntemi ile belirlenen DMF’li maddeler ise düzeylere göre ayrılmamaktadır. Bu sebeple DMF gösteren madde sayıları tek bir sütun olarak belirtilmiştir. Tablo incelendiğinde SIBTEST yönteminde DMF gösteren madde sayısının örneklem büyüklüğü arttıkça arttığı, Multi-MIMIC yönteminde ise örneklem büyüklüğü arttıkça azaldığı görülmektedir. Örneğin örneklem büyüklüğünün 600 olduğu durumlarda SIBTEST yöntemi ile 6 ve 7 madde belirlenmişken, örneklem büyüklüğünün 3600 olması ile 10 maddenin DMF gösterdiği belirlenmiştir. Multi-MIMIC yönteminde ise örneklem büyüklüğü 600 olduğunda 9 olarak belirlenen DMF’li madde sayısı örneklem büyüklüğü 3600 olduğunda 5 e kadar düşmüştür. Ayrıca örneklem büyüklüğünün 1200 ve 3600 olduğu koşullarda Multi-MIMIC yönteminde hep daha az sayıda DMF’li madde tespit edilmiştir. Şekil 25’te bu iki farklı yöntem ile değişen örneklem büyüklükleri ve oranlarına göre DMF’li madde sayısının nasıl değiştiği verilmiştir.



Şekil 25. DMF gösteren madde sayıları

SIBTEST ve Multi-MIMIC yöntemleri kullanılarak belirlenen DMF gösteren madde sayıları arasında yöntemle göre fark olup olmadığını belirlemek için parametrik olmayan testlerden Mann-Whitney U testi kullanılmıştır. Bu test sonucuna göre 6 koşulda belirlenen DMF'li madde sayısı ortalamaları SIBTEST ile 11,67, Multi-MIMIC yöntemi ile 6,50 olmuştur. Ayrıca analiz neticesinde Z istatistiği -2,822 olarak manidar bulunmuştur ($p=0,002$). Bu sonuçtan yola çıkarak bu iki yöntemin belirlediği DMF'li madde sayıları arasında farklılık olduğu söylenebilir.

Bölüm 5

Sonuç, Tartışma ve Öneriler

Bu bölümde araştırma bulgularından elde edilen sonuçların özeti, tartışması ve hepsi sonunda ulaşılan öneriler yer almaktadır.

Sonuçlar

Araştırmada iki boyutlu ve üç boyutlu olarak üretilen 2 kategorili veri setlerine geleneksel SIBTEST ve çok boyutlu çoklu nedenler çoklu göstergeler (Multi-MIMIC) yöntemleri kullanılarak gerçekleştirilen DMF analizleri neticesinde ulaşılan 1. tip hata ve güç değerleri karşılaştırılmıştır. Gerçek veri seti (PISA 2012 Türkiye Örnekleme) ile gerçekleştirilen analizlerde DMF yöntemlerine bağlı olarak belirlenen DMF'li madde sayıları karşılaştırılmıştır.

İki Boyutlu Yapılar. Yapay veri seti için yapılan analizlerde testin iki boyutlu basit ve iki boyutlu karmaşık yapıda olması ile değişen DMF düzeyleri, örneklem büyüklükleri ve odak ve referans grup oranları ile 1. tip hatanın ve DMF belirleme gücünün nasıl değiştiğine ilişkin sonuçlar aşağıdaki maddelerde verilmiştir.

1. Araştırmada örneklem büyüklükleri 600, 1200, 2400, 3600, 4800 ve 6000 olarak değişmektedir. 1. tip hatanın örneklem büyüklüğü arttıkça düştüğü sonucuna ulaşılmıştır.
2. Odak ve referans grup oranları 1:1 ve 1:2 olarak değişmektedir. Odak ve referans grup oranındaki değişimin 1. tip hata ve güç oranları üzerinde anlamlı bir etkisi olmamıştır.
3. DMF etki düzeyi arttıkça 1. tip hatanın azaldığı tespit edilmiştir. DMF etki düzeyi 1. tip hatanın yanında istatistiksel gücün de anlamlı olarak farklılaşmasına sebep olmuştur. Buna göre DMF'nin orta düzeyden yüksek düzeye çıkması güç oranlarını düşürmektedir. Özetle DMF etki düzeyinin orta düzeyde olduğu durumda yöntemlerin istatistiksel güç oranları daha yüksektir.
4. DMF analizlerinde SIBTEST ve Multi-MIMIC yöntemleri kullanılmıştır. Bu yöntemlerden testin boyut sayısını dikkate alan çok boyutlu model olan Multi-MIMIC ile elde edilen bulgularda 1. tip hatanın önemli oranda azaldığı belirlenmiştir. Ayrıca Multi-MIMIC yöntemi ile elde edilen bulgular incelendiğinde güç oranlarının da arttığı belirlenmiştir. Bu da Multi-MIMIC yöntemi ile

gerçekleştirilen analizlerin DMF belirleme gücünün daha yüksek ve 1. tip hatasının daha düşük olduğunu göstermektedir. Bu sonuç DMF analizlerinde çok boyutlu verilerde çok boyutlu DMF yöntemlerini kullanmayı öneren çalışmaların sonuçlarıyla paralellik göstermektedir.

5. Testin basit veya karmaşık yapıda olması da koşullardan biridir. Testin karmaşık yapıda olmasının 1. tip hatayı ve güç oranlarını arttırdığı belirlenmiştir.
6. Bulgular incelendiğinde, en düşük 1. tip hata iki boyutlu basit yapıya sahip yüksek DMF etki düzeyine ve yüksek örneklem büyüklüklerine sahip koşullarda Multi-MIMIC yöntemi ile elde edilmiştir. En yüksek güç oranına ise karmaşık yapıya sahip orta DMF etki düzeyine sahip testlerde ulaşılmıştır.

Üç Boyutlu Yapılar. Yapay veri seti için yapılan analizlerde testin üç boyutlu basit ve üç boyutlu karmaşık yapıda olması ile değişen DMF düzeyleri, örneklem büyüklükleri ve odak ve referans grup oranları ile 1. tip hatanın ve DMF belirleme gücünün nasıl değiştiğine ilişkin sonuçlar aşağıdaki maddelerde verilmiştir.

7. Araştırmada örneklem büyüklükleri 600, 1200, 2400, 3600, 4800 ve 6000 olarak değişmektedir. 1. tip hatanın örneklem büyüklüğü arttıkça düştüğü sonucuna ulaşılmıştır.
8. Odak ve referans grup oranları 1:1 ve 1:2 olarak değişmektedir. Odak ve referans grup oranındaki değişimin 1. tip hata ve güç oranları üzerinde anlamlı bir etkisi olmamıştır.
9. DMF etki düzeyinin orta ve yüksek olarak değişmesi 1. tip hatayı manidar düzeyde etkilemektedir. Buna göre DMF etki düzeyi arttıkça 1. tip hatanın azaldığı belirlenmiştir. DMF etki düzeyi 1. tip hatanın yanında istatistiksel gücün de anlamlı olarak farklılaşmasına sebep olmuştur. DMF'nin orta düzeyden yüksek düzeye çıkması güç oranlarını azaltmaktadır. Buna göre, DMF etki düzeyinin orta düzeyde olduğu durumda yöntemlerin güç belirleme performansları daha yüksektir.
10. DMF analizlerinde SIBTEST ve Multi-MIMIC yöntemleri kullanılmıştır. Bu yöntemlerden testin boyut sayısını dikkate alan çok boyutlu model olan Multi-MIMIC ile elde edilen bulgularda 1. tip hatanın önemli oranda azaldığı belirlenmiştir. Ayrıca Multi-MIMIC yönteminden elde edilen bulgular incelendiğinde güç oranlarının da arttığı belirlenmiştir. Bu da Multi-MIMIC

yöntemi ile gerçekleştirilen analizlerin DMF belirleme gücünün daha yüksek ve 1. tip hatanın daha düşük olduğunu göstermektedir. Bu DMF analizlerinde çok boyutlu verilerde çok boyutlu DMF yöntemlerini kullanmayı öneren çalışmaların sonuçlarıyla paralellik göstermektedir.

11. Testin basit veya karmaşık yapıda olması da koşullardan biridir. Testin karmaşık yapıda olmasının 1. tip hatayı arttırdığı belirlense de bu artışın manidar düzeyde olmadığı sonucuna ulaşılmıştır. Ancak güç oranlarındaki yükselme ise anlamlıdır.

12. Bulgular incelendiğinde, en düşük 1. tip hata üç boyutlu basit yapıya yüksek DMF etki düzeyine ve yüksek örneklem büyüklüklerine sahip koşullarda Multi-MIMIC yöntemi ile elde edilmiştir. En yüksek güç oranına ise karmaşık yapıya orta DMF etki düzeyine sahip testlerde Multi-MIMIC yöntemi kullanılarak ulaşılmıştır.

Gerçek Veri Seti. iki boyutlu gerçek veri seti tek boyutlu ve iki boyutlu DMF belirleme yöntemleri ile analiz edildiğinde, öğrencilerin cinsiyeti açısından, farklı örneklem büyüklüklerine, odak grup ve referans grup oranlarına göre DMF gösteren madde sayısının nasıl değiştiğini gösteren bulgular ile elde edilen sonuçlar ise aşağıda sıralanmıştır.

13. PISA 2012 Türkiye örnekleminde elde edilen gerçek veri setleri 600, 1200 ve 3600 kişilik örneklemden oluşmaktadır. Örneklem büyüklüğüne göre SIBTEST ile belirlenen DMF'li madde sayısı artarken, Multi-MIMIC yöntemi ile azalma gözlenmiştir.

14. Odak ve referans grup oranları da yapay veri setinde olduğu gibi 1:1 ve 1:2 olarak değişmektedir. Odak ve referans grup oranındaki değişim de DMF gösteren madde sayısını manidar düzeyde etkilememiştir.

15. DMF belirlemek için kullanılan tek boyutlu SIBTEST yöntemleri ile çok boyutlu Multi-MIMIC yöntemleri karşılaştırıldığında DMF gösteren madde sayısının manidar düzeyde değiştiği belirlenmiştir. Buna göre SIBTEST yöntemi ile belirlenen DMF gösteren madde sayısı Multi-MIMIC yöntemi ile belirlenen DMF gösteren madde sayısından daha fazladır.

16. Diğer çalışmalara paralel olarak örneklem büyüklüğüne bağlı olarak SIBTEST yönteminde DMF gösteren madde sayısı artış göstermiştir. Çok boyutlu Multi-MIMIC yönteminde ise örneklem büyüklüğü arttıkça DMF gösteren madde

sayısında azalma gözlenmiştir. Yine tek boyutlu ve çok boyutlu modellerde DMF gösteren madde sayısını araştıran benzer çalışmalar ile kıyaslandığında verinin yapısına uygun olarak seçilen DMF belirleme yöntemi neticesinde DMF gösteren madde sayısında azalma olduğu tespit edilmiştir.

Tartışma

Bu çalışmada yapay olarak üretilen iki boyutlu ve üç boyutlu veri setlerinde testin yapısı (basit-karmaşık), DMF etki düzeyi (orta-yüksek), örneklem büyüklükleri (600-1200-2400-3600-4800-6000) ve odak ve referans grup oranı (1:1 ve 1:2) değiştirilerek koşullar oluşturulmuştur.

Önceki çalışmalar incelendiğinde temel boyutun yanında etkisi gözlenen ikincil boyutun DMF'ye sebep olduğu vurgulanmıştır (De Ayala, 2009; Walker, Zhang ve Surber 2008; Mazor ve Ripkey, 1996; Camili, 1992). Bu sebeple çok boyutlu modellerde tek boyutluluk varsayımı içeren klasik yöntemler ile DMF analizi yapmanın 1. tip hatayı arttırdığı ve güç oranlarını düşürdüğü ifade edilmiştir (Ackerman ve Girdl 2003; Mazor ve Ripkey, 1996; Stout, Li, Nandakumar ve Bolt 2006; Huang 2010; Ling Liaw, 2015; Liu, Yin, Xin, Shao ve Yuan, 2019). Klasik yöntemler yerine testin boyut sayısını dikkate alan, yetenek düzeylerini beklenen ölçüde ayırıştırarak daha düşük hata ve yüksek güç oranı ile DMF analizi gerçekleştirmeye uygun yeni modeller önerilmiştir. Stout, Li, Nandakumar ve Bolt tarafından iki boyutlu testler için MULTISIB, Snow ve Oshima (2009) tarafından MULTI-DFIT, Bulut, Lee ve Suh (2016) MULTI-MIMIC ve çok boyutlu geliştirilmiş MTK-OO yöntemleri bu modellere örnek olarak gösterilebilir.

Mazor ve Ripkey' e (1996) göre testlerin toplam puanları dikkate alınarak sınıflandırılan bireyler için DMF analizi yaparken tek boyutlu varsayımını kullanmanın 1. tip hatayı arttırdığı ve güç oranlarını düşürdüğünü belirtmişlerdir. Yapılan çalışmalar incelendiğinde bu iki araştırmacıyı destekler nitelikte birçok çalışma görülmüştür. Buna göre çok boyutlu veri setlerine çok boyutlu DMF yöntemi uygulamanın 1. tip hatayı azalttığı ve güç oranlarını yükselttiği görülmüştür (Jin ve Chen, 2020; Park, Ahn ve Zopluoğlu, 2020; Bulut ve Suh, 2017; Bulut, Lee ve Suh, 2016; Ling Liaw 2015; Huang 2010; Snow ve Oshima, 2009; Mazor ve Ripkey, 1996). Bu araştırma da DMF analizleri için klasik yöntemler yerine faktör analizi temelli çok boyutlu model olan Multi-MIMIC yöntemi uygulanmış ve netice olarak

çok daha düşük 1. tip hata ve daha yüksek güç oranları elde edilmiştir. Bu yönüyle araştırma diğer çalışmalar ile paralellik göstermektedir.

Yapay veri seti için oluşturulan koşullar dikkate alındığında; örneklem büyüklüğündeki artışın hem tek boyutlu klasik yöntemlerde hem de çok boyutlu modellerde 1. tip hatada düşüşe sebep olduğu ve güç oranında artış gözlemlendiği tespit edilmiştir (Özdoğan, 2017; Finch, 2012; Ross, 2008; Barnett ve Ercikan, 2006; Russell, 2005). Bu çalışmada elde edilen bulgular incelendiğinde örneklem artışı ile birlikte 1. tip hatada düşüş olduğu gözlemlenmiştir. Bunun, örneklem büyüklüğü artışı ile RMSE değerlerinin düşmesinden ve daha doğru analizler için uygun altyapının ortaya çıkmasından kaynaklandığı düşünülmektedir. Bazı çalışmalarda örneklem büyüklüğü artışı ile RMSE'nin de arttığı ifade edilmektedir (Rustam, Naga ve Supriyati, 2019; Zhang, 2012; Finch, 2010; Bolt ve Lall, 2003). Ancak diğer çalışmalar ile kıyaslandığında örneklem büyüklüğü ve güç oranları arasında bu paralellik yoktur. Güç oranlarının 0,90 nın üzerinde olduğu birçok koşul dikkate alındığında yüksek güç oranlarında değişim olmaması beklenen bir durumdur. Odak ve referans grup oranları dikkate alındığında anlamlı farklılık gözlemlenmemiştir. Ancak burada belirtilen oranlar 1:1 ve 1:2 olarak belirlenmiştir. Bu durum Bulut ve Suh (2017) tarafından gerçekleştirilen çalışma ile paralellik göstermektedir. Bu oranların çeşitlendirilmesi ve keskin biçimde iki grup lehine de arttırılması yeterli örneklem büyüklüklerinde 1. tip hata ve güç oranlarını etkileyebilir.

Testteki DMF etki düzeyinin orta ve yüksek düzeyde olduğu koşul ele alındığında hem iki boyutlu hem de üç boyutlu yapay veri seti için DMF etki düzeyinin artmasının 1. tip hatayı azalttığı fakat güç oranlarında düşüşe sebep olduğu görülmüştür. Diğer çalışmalar ile kıyaslandığında Huang (2010), Walker ve Göçer Şahin (2016) ve Liu, Yin, Xin, Shao ve Yuan, (2019) ile benzer sonuçlara ulaşıldığı görülmektedir.

Testin basit veya karmaşık yapıda olma durumu incelendiğinde iki boyutlu veri setinde basit yapıdan karmaşık yapıya geçişin 1. tip hatayı ve güç oranlarını arttırdığı belirlenmiştir. Ancak bu durum üç boyutlu veri setinde gözlemlenmemiştir. Bunun ilk nedeninin madde sayısı ile ilgili olduğu düşünülmektedir. İki boyutlu veri setinde iki boyut için ayrı ayrı 20 olmak üzere 40 madde bulunurken, üç boyutlu veri setinde ilk boyut için 14 ikinci ve üçüncü boyut için 13'er madde bulunmaktadır. Madde sayısındaki azalmanın model veri uyumunu ve birey madde korelasyonlarını

azalttığı ifade edilmiştir (de la Torre ve Patz, 2005). İkinci nedenin ise boyut sayısındaki artış ile birlikte maddelerin iki boyutlu modellerdekinin aksine öncül yetenek yerine karmaşık modellerde diğer yetenek düzeylerine yaklaşması ve bu durumun da DMF ye etkisi olduğu düşünülmektedir. Walker, Zhang ve Surber' e (2008) göre boyut sayısının artması ile koordinat düzleminde yetenek düzeyleri ve maddeler arasındaki açının artması DMF olarak algılanmakta, iki boyutlu verilerde boyut etkisinden arındırılrsa da üç veya daha fazla boyut içeren karmaşık yapıları modellerde DMF hatasında artış olabilmektedir. Bu sonuç ile Snow ve Oshima (2009) ve Camili (1992) çalışmalarından elde edilen sonuç paralellik göstermektedir.

Gerçek veri seti için gerçekleştirilen DMF analizlerinde örneklem büyüklüğü arttıkça SIBTEST yöntemi ile DMF gösteren madde sayısı artmıştır. Bu bulgu Yörü ve Atar (2019), Arıkan, Uğurlu ve Atar (2016), Furlow, Ross ve Gagne (2009), Gierl, Jodoin ve Ackerman' in (2000) çalışmaları ile paralellik göstermektedir. Analiz yöntemine göre DMF gösteren madde sayıları dikkate alındığında Ling Liaw (2015) ve Huang (2010)'in çalışmalarında çok boyutlu modeller ile gerçekleştirilen DMF analizlerinde klasik DMF belirleme yöntemlerine göre daha az sayıda DMF gösteren madde bulunmuştur. Bulgular ile karşılaştırıldığında da Multi-MIMIC yönteminin SIBTEST ile manidar düzeyde farklılaşması bu çalışmaları destekler niteliktedir. Ancak bu farklılaşmanın daha yüksek örneklem büyüklüklerinde gözlenmesi beklenmektedir.

Huang (2010) DMF gösteren madde sayısındaki azalışın temel sebebinin çok boyutlu yapılarda farklı özellikleri ölçen maddeler boyut etkisinden arındırılmadan (tek boyutlu klasik DMF yöntemleri ile) DMF analizi gerçekleştirildiğinde maddelerin farklı özellikleri ölçmesinin DMF ye sebep olduğu ve DMF gösteren madde sayısını arttırdığını iddia etmiştir.

Özetle bir testin boyut sayısını dikkate alarak ve buna en uygun analiz yöntemini seçerek, yetenek düzeylerinden kaynaklanması muhtemel madde farklılaşmasının önüne geçilebilir ve yanlılığa sebep olabilecek maddeler daha kolay belirlenebilir. Ancak bu analiz yöntemlerini kullanırken testin yapısının, örneklem büyüklüğünün ve DMF etki düzeyinin de doğru bir şekilde belirlenmesi gerekir. Sonuçlardan hareketle, eğer iki boyutlu veri seti kullanılacaksa daha doğru ve daha az hatalı sonuçlar için testin basit yapıları, en az 2400 örneklem büyüklüğünde boyut

sayısını dikkate alan Multi-MIMIC yöntemi ile analiz edilmesi önerilmektedir. Bu araştırmaya göre, eğer üç boyutlu veri seti ile DMF analizleri gerçekleştirilecekse, testin, basit yapılı, en az 2400 örneklem büyüklüğünde Multi-MIMIC yönteminin tercih edilmesinin hataları düşüreceği ve doğru bulgular elde edileceğini göstermektedir.

Uygulayıcılara Yönelik Öneriler

1. Araştırmada yapay veri seti için örneklem büyüklüğü 600, 1200, 2400 3600, 4800 ve 6000 olarak belirlenmiştir. Ancak elde edilen sonuçlara göre 1. tip hata 2400 ve sonraki örneklerde düşüş göstermektedir. Bu sebeple örneklem büyüklüğünün 2400 altında olmaması önerilmektedir.
2. 1. tip hatalar karşılaştırıldığında en düşük hataların basit yapılı testlerde elde edildiği belirlenmiştir. Daha düşük hata için basit yapılı testlerin tercih edilmesi önerilmektedir.
3. Tüm koşullarda en düşük 1. tip hatalara ve en yüksek güç oranlarına çok boyutlu yapıya uygun olan Multi-MIMIC yöntemi kullanılarak ulaşılmıştır. Araştırmacıların analizlerinde verinin yapısına uygun yöntemi seçmeleri önerilmektedir.
4. Güç oranları karşılaştırıldığında daha yüksek güç oranlarına karmaşık yapılı yüksek DMF etki düzeyine sahip koşullarda rastlanmıştır. Ayrıca özellikle üç boyutlu yapılarda kullanılan yöntemin de güç oranlarını etkilediği belirlenmiştir. Karmaşık yapıya sahip üç boyutlu testlerde DMF analizi için Multi-MIMIC yönteminin kullanılması önerilmektedir.

İleri Araştırmalara Yönelik Öneriler

1. Testler iki boyutlu ve üç boyutlu olarak basit veya karmaşık yapıda üretilmiştir. Boyut sayısı artırılarak boyutlar arası etkinin DMF ye etkisi, 1. tip hatayı ve güç oranlarını nasıl etkilediği belirlenebilir. Ayrıca basit ve karmaşık yapının yanında yarı karmaşık yapı da araştırmaya dâhil edilerek karmaşıklık düzeyinin etkisi karşılaştırılabilir.
2. Bu çalışmanın yapay veri seti için örneklem büyüklüğü 600, 1200, 2400 3600, 4800 ve 6000 olarak belirlenmiştir. Çalışma için çok daha küçük örneklem

büyükliklerinde 1. tip hatanın ve güç oranlarının nasıl değiştiği incelenebilir. Ayrıca sonuçlar incelendiğinde yüksek örneklem büyüklükleri birbirinden manidar düzeyde ayrılmadığı için örneklem gruplarının sayısı azaltılabilir.

3. Çalışmada odak ve referans grup oranı 1:1 ve 1:2 ile sınırlı tutulmuştur. Gelecek çalışmalarda örneklem büyüklüğü iki grup lehine de yüksek oranda değişen oranlar tercih edilerek 1. tip hatayı ve güç oranlarını nasıl etkileyeceği araştırılabilir.
4. Bu çalışmanın yapay veri seti 40 madde olarak sabit tutulmuştur. Madde sayısının iki boyutlu ve üç boyutlu yapay veri setlerinde 40 olması yerine üç boyutlu veri setinde 60 olarak belirlenebilir. Böylelikle tüm boyutlar için 20 madde sağlanmış olur. Madde sayılarının değişkenlik gösterdiği yeni koşullar ile test uzunluğunun etkisi araştırılabilir.
5. Çalışmada çok boyutluluk için geliştirilmiş iki kategorili iki parametrelili lojistik model kullanılmıştır. Sonraki çalışmalarda şans başarısının da dâhil olduğu üç parametrelili lojistik model kullanılabilir.
6. DMF belirleme yöntemleri olarak SIBTEST ve Multi-MIMIC kullanılmıştır. Diğer DMF belirleme yöntemleri ile yöntemler arasındaki farklılıklar incelenebilir. Ancak MULTISIB ve çok boyutlu genişletilmiş MTK-OO fonksiyonları için analiz koşulları ve varsayımlarına dikkat edilmelidir.
7. Gerçek veri seti PISA 2012 Türkiye örnekleminde elde edilmiştir. Testin iki boyutlu olması için 30 matematik okuryazarlığı sorusuna 30 okuduğunu anlama becerisini ölçen sorular dâhil edilmiş ve analizler neticesinde bu testin 2 boyutlu bir testte olması gereken özellikleri gösterdiği belirlenmiştir. Madde sayısı tüm koşullarda 60 olarak sabit tutulmuştur. Test uzunluğu değiştirilerek DMF'ye etkisi incelenebilir. Testin boyut sayısı da artırılarak ve azaltılarak karşılaştırmalar yapılabilir.
8. DMF gösteren maddeler cinsiyete göre belirlenmiş, 600, 1200 ve 3600 kişilik örneklem büyüklükleri kullanılmıştır. Sonraki araştırmalarda daha fazla örneklem sınıflaması ve daha büyük grupla cinsiyet yerine farklı örtük özellikler ile DMF gösteren maddeler belirlenebilir.
9. Ayrıca odak ve referans grup oranları da yapay veride olduğu gibi 1:1 ve 1:2 olarak değiştirilmiştir. Yine yapay veri seti için önerildiği gibi örneklem

büyükliđünün iki grup lehine de yüksek oranda deđişen oranlar tercih edilerek DMF gösteren madde sayısına etkisi incelenebilir.

10. Gerçek veri seti için veri setlerine uygun farklı DMF belirleme yöntemleri kullanılarak aralarındaki ilişki ve deđişkenlik belirlenebilir.

11. Gerçek veri setinde DMF analizi bireylerin cinsiyet deđişkenine göre gerçekleştirilmiştir. Bireylerin cinsiyet dışında okul türü, bölge, sosyo ekonomik düzey gibi gözlenen deđişkenler ile veya örtük özellikler kullanılarak çok boyutlu DMF analizlerine etkisi karşılaştırılabilir.

Kaynaklar

- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29(1), 67-91.
- Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, 7(4), 255-278.
- Ackerman, T. A. (1996). Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement*, 20(4), 311-329.
- 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, 37–53.
- Adams, R. J. ve Rowe, K. J. (1988). Item bias. İinde J.P Keeves (Ed.), Educational research, methodology, and measurement: An international handbook. Oxford: Pergamon Press.
- AERA, A. (1999). NCME. *Standards for Educational and Psychological Testing*. American Educational Research Association. 2nd ed. Washington DC.
- Akın Arıkan, ., Uęurlu, S. & Atar, B. (2016). A DIF and bias study by using MIMIC, SIBTEST, logistic regression and Mantel-Haenszel methods [in Turkish]. Hacettepe University Journal of Education, *Hacettepe Üniversitesi Eęitim Fakóltesi Dergisi*, 31(1), 34-52.
- Anastasi, A. ve Urbina, S. (1988). *Psychological testing* (7th ed.). USA: Macmillan Pub. Co. Inc.
- Angoff, W. H. (1988). *Validity: An evolving concept*.
- Anıl, D., Güzeller, C. O., okluk, Ö., & Şekercioęlu, G. (2010). Level determination exam (SBS-2008) the determination of the validity and reliability of 7th grade mathematics sub-test. *Procedia-Social and Behavioral Sciences*, 2(2), 5292-5298.

- Anıl, D., Özkan Y. Ö., Demir, E. (2015) PISA 2012 Araştırması Ulusal Nihai Rapor ISBN: 978-975-11-3899-6, <http://odsgm.meb.gov.tr/> den alınmıştır.
- APA, A. (1999). NCME. *Standards for educational and psychological testing*. AERA.
- Arslan, S. (2014). Çoklu Göstergeler Çoklu Nedenler ve Lojistik Regresyon Yöntemlerinin Değişen Madde Fonksiyonunu Belirleme Performansları, Yayınlanmış Yüksek Lisans Tezi, Hacettepe Üniversitesi. <http://tez2.yok.gov.tr/adresinden edinilmiştir>
- Atalay, K., Gök, B., Kelecioğlu, H. & Arsan, N. (2012). Değişen madde fonksiyonunun belirlenmesinde kullanılan farklı yöntemlerin karşılaştırılması: Bir simülasyon çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 43, 270- 281.
- Atalay Kabasakal, K. Arsan, N. Gök, B. ve Kelecioğlu, H. (2014). Değişen Madde Fonksiyonunun Belirlenmesinde MTK Olabilirlik Oranı SIBTEST ve Mantel-Haenszel Yöntemlerinin Performanslarının (I. tip Hata ve Güç) Karşılaştırılması, *Kuram ve Uygulamada Eğitim Bilimleri*, 14(6), 2175-2193
- Başusta, N. B. (2013). PISA 2006 Fen Başarı Testinin Madde Yanlılığının Kültür ve Dil Açısından İncelenmesi (Doktora Tezi). Hacettepe Üniversitesi, Ankara, <http://tez2.yok.gov.tr/adresinden edinilmiştir>.
- Baykul, Y. (2000). Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması. Ankara: ÖSYM Yayınları.
- Baykul, Y. (2010). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: Pegem Akademi.
- Bertrand, R. & Boiteau, N. (2003). Comparing the stability of IRT-Based and non IRT based DIF methods in different cultural contexts using TIMSS data *Eric Document Reproduction* No. ED 476 924).
- Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability*. In Lord, F.M. & Novick, M.R. (Eds.), *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Bolt, D. (2000). A SIBTEST approach to testing DIF hypothesis using experimentally designed test items. *Journal of Educational Measurement*, 37, 307-327.

- Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov Chain Monte Carlo. *Applied Psychological Measurement, 27*(6), 395-414.
- Bolt, D., & Stout, W. (1996). Differential item functioning: Its multidimensional model and resulting SIBTEST detection procedure. *Behaviormetrika, 23*(1), 67-95.
- Bulut, O. (2013). *Between-person and within-person subscore reliability: Comparison of unidimensional and multidimensional IRT models*. Unpublished Doctoral Dissertation. University of Minnesota Faculty of the Graduate School.
- Bulut, O. and Suh, Y. (2017) Detecting Multidimensional Differential Item Functioning with the Multiple Indicators Multiple Causes Model, the Item Response Theory Likelihood Ratio Test, and Logistic Regression. *Front. Educ. 2*:51. DOI: 10,3389/feduc.2017.00051
- 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. [https://doi.org/10,1177/014662169201600203](https://doi.org/10.1177/014662169201600203)
- Camili, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. London: Sage.
- Carmines, E. G., & Zeller, R. A. (1982). Reliability and Validity Assessment. 5th printing. *Beverly Hills: Sage Publications Inc.*
- Carter, N. T. (2011). Applications of differential functioning methods to the generalized graded unfolding model (Doctoral dissertation). Bowling Green State University, ABD.
- 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.
- Clauser, B. E. and Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice, 17*, 31-44.
- Chen, H. F., & Jin, K. Y. (2018). Applying logistic regression to detect differential item functioning in multidimensional data. *Frontiers in Psychology, 9*, 1302.

- Clauser, B. E., Nungester, R. J., Mazor, K., & Ripkey, D. (1996). A comparison of alternative matching strategies for DIF detection in tests that are multidimensional. *Journal of Educational Measurement*, 33(2), 202-214.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, A. S., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement*, 42, 133–148.
- Cole, N.S., and Moss, P.A. (1989). *Bias in Test Use in Educational Measurement* (Ed. Robert L. Linn), 201-218. Collier MacMillian: London
- Crocker, L. ve Algina, J. (1986). *Introduction to classical and modern test theory*. USA: Rinehart and Winston Inc.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Journal of Educational Measurement* (2), 443-507. Washington, DC: American Council on Education.
- Crocker, L. ve Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Orlando: Harcourt Brace Jovanovich, Inc.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Ohio: Cengage Learning.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Cronbach, L. J. ve Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Cureton, E. E. (1951). *Validity*. In E. F. Lindquist (Ed.), *Educational measurement* (1st ed., pp. 621- 694). Washington, DC: American Council on Education. -
- De Ayala, R. J. (1993). An introduction to polytomous item response theory models. *Measurement & Evaluation in Counseling & Development*, 25(4).
- 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.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford.

- De la Torre, J., & Patz, R. L. (2005). Making the most of what we have: A practical application of multidimensional item response theory in test scoring. *Journal of Educational and Behavioral Statistics*, 30(3), 295-311.
- Dođan, N., & Sevindik, H. (2011). İlköđretim 6. sınıflar için uygulanan Seviye Belirleme Sınavı'nın uygunluk geđerliđi. *Eđitim ve Bilim*, 36(160).
- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Erlbaum.
- Embretson (Whitely), S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93,179-197.
- Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum Publishers.
- Erdem, B. (2015). Ortaöđretime geđište kullanılan ortak sınavların deđişen madde fonksiyonu ađısından kitapçık türlerine göre farklı yöntemlerle incelenmesi.Yayınlanmış Yüksek Lisans Tezi, Hacettepe Üniveristesi, Ankara. <http://tez2.yok.gov.tr/adresinden edinilmiştir>.
- Erdem Keklik, D. (2012). İki kategorili maddelerde tek biçimli deđişen madde fonksiyonu belirleme tekniklerinin karşılaştıırılması: bir simülasyon çalışması (Doctoral dissertation, Doktora Tezi, Ankara Üniversitesi, Eđitim Bilimleri Enstitüsü). <http://tez2.yok.gov.tr/adresinden edinilmiştir>.
- Fang, T. (1999). Detecting DIF in polytomous item responses. Unpublished doctor dissertation, University of Ottawa, Canada.
- Ferrando, P. J., & Lorenzo-Seva U. (2016). A note on improving EAP trait estimation in oblique factor-analytic and item response theory models. *Psicologica*, 37, 235-247.
- Ferrando, P. J., & Lorenzo-Seva U. (2017). Assessing score determinacy, measurement quality, and closeness to unidimensionality in exploratory item factor analysis. *Educational and Psychological Measurement*, 0013164417719308

- Finch, 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*, 278-295.
- Finch, W. H. & French, B. F. (2007). Detection of crossing differential item functioning: a comparison of four methods. *Educational and Psychological Measurement, 67*, 565-582.
- Finch, H. (2010). Item parameter estimation for MIRT model: Bias and precision of confirmatory factor analysis based models. *Applied Psychological Measurement, 34*(1), 10-26.
- Frasier, C. (1988). NOHARM [Computer program]. New South Wales, Australia: University of New England, Center for Behavioral Studies.
- Furlow, F. C., Ross, T.R & Gagne, P. (2009). The impact of multidimensionality on the detection of differential bundle functioning using simultaneous item bias test. *Applied Psychological Measurement, 33-6*, 441-464, Sage Publications.
- Gierl, M. J., Jodoin, M. G., & Ackerman, T. A. (2000). Performance of Mantel-Haenszel, Simultaneous Item Bias Test, and Logistic Regression when the proportion of DIF items is large. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Guion, R. M. (1977). Content validity—The source of my discontent. *Applied Psychological Measurement, 1*(1), 1-10,
- Gulliksen, H. (1950). The reliability of Speeded Test, *Psychometrika, 15*, 259-269
- Guttman, L. (1944). A basis for scaling qualitative data. *American sociological review, 9*(2), 139-150,
- Hambleton, R. K. & Swaminathan, H. (1985). *Item Response Theory: Principles and Application*. Boston: Kluwer Academic Publishers Group.
- Hambleton, K. R. ve Rogers, J. H. (1989). Detecting potentially biased test items: comparison of IRT area and MH methods. *Applied Measurement in Education, 2* (4), 313–334.
- Hambleton, R.K., Swaminathan, H. ve Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. London: Sage Publication.

- Harman, H. H. (1962). *Modern Factor Analysis*, 2nd Edition. University of Chicago Press, Chicago.
- Harold J. (1989). The determination of cultural item bias in the California Achievement Tests Arizona Üniversitesi; Doktora tezi.
- Hartig, J., & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*, 35(2009), 57–63.
- Harvey, R. J., & Hammer, A. L. (1999). Item response theory. *The Counseling Psychologist* 27, 353-383.
- Higaldo, M. D. ve Lopez-Pina, J. A. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement*, 64, 903-915.
- Holland, J. L. (1985). *The self-directed search*. Psychological Assessment Resources.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.
- Horst, P. (1966). *Psychological Measurement and Prediction*. Belmont: Wadsworth Pub. Co.
- Huang, X. (2010). Differential item functioning: The consequence of language, curriculum, or culture? (Doctoral dissertation, UC Berkeley).
- Jin, K. Y., & Chen, H. F. (2020). MIMIC approach to assessing differential item functioning with control of extreme response style. *Behavior Research Methods*, 52(1), 23-35.
- Jodoin, M. G. and Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14, 329–349.
- Johnson Frotman, K., A. (2007). The evaluation of new criteria for polytomous DIF in the DFIT framework (Doctoral Dissertation). Illinois Institute of Technology, Chicago.

- Kane, M. T. (2006), Validation. In Brennan R. (Ed.), *Journal of Educational Measurement* (4), 17-64, Westport, CT: Praeger Publishers.
- Kelley, T. L. (1935). Essential Traits of Mental Life, Harvard Studies in Education, vol. 26. Harvard University Press, Cambridge.
- Kim, S. H. and Cohen, A. S. (1998). Detection of differential itemfunctioning under the graded response model with the likelihood ratio test. *Applied Psychological Measurement*, 22, 345–355.
- 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.
- Kim, Y.H, & Jang E. E. (2009). Differential functioning of reading subskills on the OSSLT for L1 and ELL students: A multidimensionality model-based DBF/DIF approach. *Language Learning*, 59:4, pp.825-865
- Keleciođlu, H. ve Őahin, S. G. (2014). GeçmiŐten g¼n¼m¼ze geęerlik, *Eđitimde ve Psikolojide Ölęme ve Deęerlendirme Dergisi*, cilt 5 sayı 2 kiŐ 2014 1-11, issn 1309-6575.
- Klein, A., and Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika* 65, 457–474. doi:10,1007/BF02296338
- Koęar, H., (2014). Madde tepki kuramının farklı uygulamalarından elde edilen parametrelerin ve model uyumlarının örneklem büyüklüęü ve test uzunluęu aęısından karŐılaŐtırılması. Hacettepe Üniversitesi Eđitim Bilimleri Enstit¼s¼, Ankara. <http://tez2.yok.gov.tr/adresinden edinilmiŐtir>.
- Kromrey, J. D., Parshall, C. G., Chason, W. M., & Yi, Q. (1999). Generating item responses based on multidimensional item response theory. Retrieved July 25, 2005, from <http://www2.sas.com/proceedings/sugi24/Posters/p241-24.pdf>
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Houghton Mifflin Co.

- Lee, S., Bulut, O., and Suh, Y. (2017). Multidimensional extension of multiple indicators multiple causes models to detect DIF. *Educational and Psychological Measurement* doi:10.1177/0013164416651116
- Lee, J. (2012). Multidimensional item response theory: An investigation of interaction effects between factors on item parameter recovery using Markov Chain Monte Carlo. Unpublished Doctoral Dissertation. Michigan State University Measurement and Quantitative Methods.
- Liaw, Y. L. (2015). When can multidimensional item response theory (MIRT) models be a solution for differential item functioning (DIF)? A Monte Carlo Simulation Study (Doctoral dissertation).
- Linden, W., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.
- Lissitz, W.R. ve Samuelsen, K. (2007a). A suggested change in terminology and emphasis regarding validity and education, *Educational Researcher*, 36(8), 437-448.
- Liu, Y., Yin, H., Xin, T., Shao, L., & Yuan, L. (2019). A Comparison of Differential Item Functioning Detection Methods in Cognitive Diagnostic Models. *Frontiers in Psychology*, 10, 11-37.
- Loevinger (1957). Objective tests as instruments of psychological theory. *Psychological Reports, Monograph Supplement*, 3, 635-694.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley Publishing Company
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Magnusson, D. (1968). *Test theory*. Massachusetts. Addison-Wesley.
- Mantel, N. & Haenszel, W. M. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institute*, 22, 719-748.

- Mazor, K. M., Hambleton, R. K., & Clauser, B. E. (1998). Multidimensional DIF analyses: The effects of matching on unidimensional subtest scores. *Applied Psychological Measurement, 22*(4), 357-367.
- McDonald, R. P. (1982). Linear versus models in item response theory. *Applied Psychological Measurement, 6*, 379-396.
- McKinley, R. L., & Reckase, M. D. (1982). *The Use of the General Rasch Model with Multidimensional Item Response Data* (No. RR-82-1-ONR). American Coll Testing Program Iowa City Ia.
- MEB (2007). *PISA 2006 Uluslararası Öğrenci Değerlendirme Programı Ulusal Ön Rapor*. MEB, Ankara.
- Mellenberg, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics, 105*-118.
- Mellenberg, G. J. (1983). *Conditional Item Bias Methods*. In S. H. Irvine and W. J. Barry (Eds), *Human Assessment and Cultural Factors* (pp. 293-302). New York: Plenum Press.
- Mellenberg, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research: Applications of Item Response Theory, 13*(2), 123-144.
- Mellor, T. L. (1995). A comparison of four differential item functioning methods for polytomously scored items. Unpublished doctor dissertation. The university of Texas Austin.
- 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.
- Messick, S. (1989). *Validity*. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., 13–103). New York, NY: American Council on Education and Macmillan.
- Miller, M. D., & Oshima, T. C. (1992). Effect of sample size, number of biased items, and magnitude of bias on a two-stage item bias estimation method. *Applied Psychological Measurement, 16*(4), 381-388.

- Millsap, R.E. & Everson, H.T. (1993). Methodology Review: Statistical Approaches for Assessing Measurement Bias. *Applied Psychological Measurement*, 17(4), 297–334.
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62(3), 229-258.
- Muthen, L. K., and Muthen, B. O. (1998). *Mplus: Statistical Analysis with Latent Variables User's Guide*. Los Angeles, CA: Muthen & Muthen.
- Narayanan, P., and Swaminathan, H. (1996). Identification of Items that show Nonuniform DIF. *Applied Psychological Measurement*, 20(3), 257–274.
- Oshima, T. C., Raju, N. S., & Flowers, C. P. (1997). Development and Demonstration of Multidimensional IRT-Based Internal Measures of Differential Functioning of items and Tests. *Journal of Educational Measurement*, 34(3), 253-272.
- Osterlind, S. & Everson H. (2009). *Differential item functioning*. Londra: Sage Publications.
- Öğretmen, T. , Doğan, N. (2004). OKÖSYS matematik alt testine ait maddelerin yanlılık analizi. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 5,8.
- Özdoğan, D. (2017). Çok Boyutlu Testlerde Değişen Madde Öbeği Fonksiyonunun SIBTEST Yöntemiyle İncelenmesi.
- Paek, I., & Cai, L. (2014). A comparison of item parameter standard error estimation procedures for unidimensional and multidimensional item response theory modeling. *Educational and Psychological Measurement*, 74(1), 58-76.
- Park, S. E., Ahn, S., & Zopluoglu, C. (2020). Differential Item Functioning Effect Size from the Multigroup Confirmatory Factor Analysis for a Meta-Analysis: A Simulation Study. *Educational and Psychological Measurement*, 0013164420925885.
- Pedrajita, J. Q. & Talisayon, V. M. (2009). Identifying biased test items by differential item functioning analysis using contingency table approaches: A comparative study. *Education Quarterly*, 67(1), 21-43.

- PISA 2006 Uluslararası Öğrenci Başarılarını Değerlendirme Programı Ulusal Ön Rapor (2007) <http://earged.meb.gov.tr/dosyalar> adresinden indirilmiştir.
- Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, 53(4), 495-502.
- Raju, N. S., Oshima, T.C., & Walach, A.H. (2009). *DFIT8* [Computer software]. St. Paul, MN: Assessment Systems Corporation.
- Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9:401–412
- Reckase, M. D. Ackerman, T. A. ve Carlson, J. E. (1988). Building a unidimensional test using multidimensional items. *Journal of Educational Measurement*, 25, 193-203.
- Reckase, M. D. (2007). Multidimensional item response theory. C. R. Rao, S. Sinharay, (Ed), *Handbook of statistics 26: psychometrics* (607-642). North Holland: Elsevier B. V.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.
- Romberg, T. A., Zarinnia, E. A., & Williams, S. (1989). The influence of mandated testing on mathematics instruction: Grade eight teachers' perceptions. T. Romberg & L. Wilson (1992, September). Alignment of tests with the standards. *Arithmetic Teacher*, 40(1), 18-22.
- Roussos, L. A ve Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel type I error performance. *Journal of Educational Measurement*, 33, 215-230,
- Rulon, P. J. (1946). On the validity of educational tests. *Harvard Educational Review*, 16, 290-296.
- Russell, S. S. (2005). Estimates of Type I error and power for indices of differential bundle and test functioning. Unpublished Doctoral Dissertation. Bowling Green State University, Ohio.
- Rustam, A., Naga, D. S., & Supriyati, Y. (2019). A Comparison of Mantel-Haenszel and Standardization Methods: Detecting Differential Item Functioning. *Jurnal Matematika dan Pembelajaran*, 7(1), 16-31.

- Samuelsen, K. M. (2005). Examining differential item functioning from a latent class perspective (Doctoral Dissertation). Department of Measurement, Statistics and Evaluation, University of Maryland, ABD.
- Shealy, R. T., & Stout, W. F. (1993a). *An item response theory model for test bias and differential test functioning*. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp.197-239). Hillsdale, NJ: Erlbaum.
- Shealy, R., & Stout, W. F. (1993b). 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*, 159–194.
- Shepard, L. A. (1993). Chapter 9: Evaluating test validity. *Review of research in education*, *19*(1), 405-450,
- Sheng Y. and Wikle C. K. (2007). Comparing multi unidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, *68*(3), 413-430,
- Shih, C.-L., & Wang, W.-C. (2009). Differential item functioning detection using the multiple indicators, multiple causes MIMIC method with a pure short anchor. *Applied Psychological Measurement*, *33*, 184-199.
- Snow, T.K. & Oshima, T.C. (2009). A comparison of unidimensional and three-dimensional differential item functioning analysis using two dimensional data. *Educational and Psychological Measurement* *69*:5, 732-747, Sage Publications
- Spearman, C. (1904). Measurement of association, Part II. Correction of 'systematic deviations'. *Am J Psychological*, *15*, 88-101.
- Steinberg, L., & Thissen, D. (2006). Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods*, *11*, 402–415. DOI: 10.1037/1082-989X.11.4.402
- Stout, W., & Roussos, L. (1995). *SIBTEST manual*. Urbana: University of Illinois, Department of Statistics, Statistical Laboratory for Educational and Psychological Measurement.

- Stout, W., Li, H. H., Nandakumar, R., & Bolt, D. (1997). MULTISIB: A procedure to investigate DIF when a test is intentionally two-dimensional. *Applied Psychological Measurement*, 21(3), 195-213.
- Suh, Y., and Cho, S.-J. (2014). Chi-square difference tests for detecting functioning in a multidimensional IRT model: A Monte Carlo study. *Appl. Psychol. Measure.* 38, 359–375. doi:10,1177/0146621614523116
- Sünbül, Ö. (2011). Çeşitli boyutluluk özelliğine sahip yapılarda, madde parametrelerinin değişmezliğinin klasik test teorisi, tek boyutlu madde tepki kuramı ve çok boyutlu madde tepki kuramı çerçevesinde incelenmesi, Yayınlanmamış Doktora Tezi. Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü. <http://tez2.yok.gov.tr/adresinden edinilmiştir>.
- Swaminathan, H. & Rogers, H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*, 27(4), 361-370,
- Sympson, J. B. (1978). *A model for testing with multidimensional items*. In Weiss DJ (ED) Proceedings of the 1977 Computerized Adaptive Testing Conference, University of Minnesota, Minneapolis.
- Tabachnick, B.G. and Fidell, L.S. (2013) Using Multivariate Statistics. Pearson, Boston.
- Thissen, D., Steinberg, L. ve Wainer, H. (1993). *Detection of Differential item functioning using the parameters of item response models*. In P. W. Holland ve H. Wainer (Eds.), *Differential item functioning* (pp. 67–114). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Thissen, D. (2001). IRTLRF v.2.0,b: *Software for the Computation of the Statistics Involved in Item Response Theory Likelihood-Ratio tests for Differential Item Functioning*.
- Thorndike R. L.ve Hagen, E. (1961). *Measurement and Evaluation in Psychology and Education*. New York: John Wiley and sons.
- Tittle, C.K. (1988). *Test Bias in Educational Teach, Methodology, and measurement: An International Handbook* (Eds. J. Keeves), 392-398.Pergamon Press: UK.

- Toland, M. D., Bowen, P., & Dueber, D. M. (2020). Multidimensional Item Response Theory Analysis of Work-Related, After-Hours Contact, and Psychosocial Measurement Scales for Construction Professionals. *Journal of Construction Engineering and Management*, 146(7), 04020062.
- Tunç, E. B. (2016). İki Ve Çok Kategorili Puanlanan Maddelerde Değişen Madde Fonksiyonlarının Karşılaştırılması, (Doktora Tezi), Ankara Üniversitesi, Ankara. <http://tez2.yok.gov.tr/adresinden edinilmiştir>.
- Turgut, M. F. ve Baykul, Y. (2013). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Yayıncılık.
- Usta, H. G. (2014). PISA 2003 ve PISA 2012 Matematik okuryazarlığı üzerine uluslararası bir karşılaştırma: Türkiye ve Finlandiya (PISA 2003 and PISA 2012 mathematics literacy on an international comparison: Turkey and Finland). Unpublished master's thesis. Ankara University, Ankara.
- Uyar, Ş. (2015). Gözlenen gruplara ve örtük sınıflara göre belirlenen değişen madde fonksiyonunun karşılaştırılması. Yayınlanmış Doktora tezi, Hacettepe Üniversitesi, Ankara, <http://tez2.yok.gov.tr/adresinden edinilmiştir>.
- Vaughn, B. K., & Wang, Q. (2010). DIF trees: Using classification trees to detect differential item functioning. *Educational and Psychological Measurement*, 70, 941-952.
- 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, 147–163.
- 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.
- Walker, C. M., Zhang, B., & Surber, J. (2008). Using a multidimensional differential item functioning framework to determine if reading ability affects student performance in mathematics. *Applied Measurement in Education*, 21(2), 162-181. DOI: 10.1080/08957340801926201
- Walker C. M. & Göçer Şahin, S. (2016). Using a multidimensional IRT framework to better understand differential item functioning (DIF): A tale of three DIF

detection procedures. *Educational and Psychological Measurement* 1-26, Sage Publications

- Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). Unidimensional IRT estimates the comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimates. *Applied Psychological Measurement*, 12(3), 239-252.
- Whitaker, T. A., Fitzpatrick, S. J., Williams, N. J., & Dodd, B. G. (2003). IRTGEN: A SAS macro program to generate known trait scores and item responses for commonly used item response theory models. *Applied Psychological Measurement*, 27, 299-300,
- Wyse, A. E. and Mapuranga, R. (2009). Differential item functioning analysis using Rasch item information functions. *International Journal of Testing*, 9(4), 333-357.
- Yıldırım, O. (2019). Detecting Gender Differences in PISA 2012 Mathematics Test with Differential Item Functioning. *International Education Studies*, 12(8), 59-71.
- Ayva-Yörü, F. G., & Atar, H. Y. (2019). Determination of differential item functioning (DIF) according to SIBTEST, Lord's χ^2 , Raju's area measurement and Breslow-Day methods. *Journal of Pedagogical Research*, 3(3), 139.
- Yu, C., & Muthen, B. (2002, April). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, L.A.
- Zenisky, A.L., Hambleton, R.K. ve 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.
- Zhang, A., Xie, X., & Li, F. (2010). *Parameters estimation for multidimensional item response theory: an effective method of determining dimensions and Bayesian method parameters estimation*. Anti-Counterfeiting Security and Identification in Communication (ASID).

- Zhang, J. (2012). Calibration of response data using MIRT models with simple and mixed structures. *Applied Psychological Measurement, 36*(5), 375-398.
- Zumbo, B. D., & Thomas, D. R. (1996). A measure of DIF effect size using logistic regression procedures. Paper presented at the National Board of Medical Examiners, Philadelphia, PA.
- Zumbo, B. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

EKLER DİZİNİ

EK-A1: Madde Parametreleri İki Boyutlu Basit Yapılı Test

Madde	a1	a2	D	MDISC	Açı
1	1,270	0,000	-0,579	1,270	0,000
2	1,082	0,010	0,422	1,082	0,009
3	1,688	0,030	-0,109	1,688	0,018
4	1,332	0,036	-0,533	1,332	0,027
5	1,000	0,036	-0,233	1,001	0,036
6	0,962	0,043	-0,123	0,963	0,045
7	0,934	0,050	-0,726	0,935	0,053
8	1,213	0,076	0,415	1,215	0,063
9	1,130	0,081	0,074	1,133	0,072
10	0,811	0,065	-0,147	0,814	0,080
11	1,339	0,120	-1,699	1,344	0,089
12	1,097	0,108	0,614	1,102	0,098
13	0,833	0,090	-0,218	0,838	0,108
14	0,925	0,108	0,434	0,931	0,116
15	0,946	0,119	-0,960	0,953	0,125
16	1,515	0,205	-0,306	1,529	0,134
17	1,060	0,153	0,309	1,071	0,143
18	0,796	0,122	-0,267	0,805	0,152
19	0,819	0,133	1,270	0,830	0,161
20	1,227	0,211	1,082	1,245	0,170
21	0,434	1,194	-0,579	1,270	0,349
22	0,351	1,023	0,422	1,082	0,331
23	0,519	1,607	-0,109	1,689	0,312
24	0,386	1,275	-0,533	1,332	0,294
25	0,272	0,963	-0,233	1,001	0,275
26	0,245	0,931	-0,123	0,963	0,257
27	0,221	0,908	-0,726	0,935	0,239
28	0,266	1,186	0,415	1,215	0,221
29	0,227	1,110	0,074	1,133	0,202
30	0,149	0,800	-0,147	0,814	0,184
31	0,221	1,326	-1,699	1,344	0,165
32	0,161	1,091	0,614	1,103	0,147
33	0,107	0,831	-0,218	0,838	0,128
34	0,102	0,926	0,434	0,932	0,110
35	0,087	0,949	-0,960	0,953	0,091
36	0,112	1,525	-0,306	1,529	0,073
37	0,059	1,069	0,309	1,071	0,055
38	0,030	0,805	-0,267	0,806	0,037
39	0,015	0,829	1,270	0,829	0,018
40	0,000	1,245	1,082	1,245	0,000

EK-A2: Madde Parametreleri İki Boyutlu Karmaşık Yapılı Test

Madde	a1	a2	D	MDISC	Açı
1	1,194	0,434	-0,579	1,270	0,349
2	1,008	0,393	0,422	1,082	0,372
3	1,559	0,648	-0,109	1,688	0,394
4	1,219	0,539	-0,533	1,333	0,416
5	0,906	0,425	-0,233	1,001	0,439
6	0,863	0,428	-0,123	0,963	0,460
7	0,828	0,434	-0,726	0,935	0,483
8	1,063	0,589	0,415	1,215	0,506
9	0,979	0,571	0,074	1,133	0,528
10	0,694	0,426	-0,147	0,814	0,551
11	1,13	0,729	-1,699	1,345	0,573
12	0,913	0,618	0,614	1,102	0,595
13	0,683	0,485	-0,218	0,838	0,617
14	0,747	0,556	0,434	0,931	0,640
15	0,752	0,586	-0,96	0,953	0,662
16	1,184	0,967	-0,306	1,529	0,685
17	0,814	0,696	0,309	1,071	0,707
18	0,601	0,537	-0,267	0,806	0,729
19	0,606	0,566	1,27	0,829	0,751
20	0,89	0,871	1,082	1,245	0,775
21	0,708	0,724	-0,68	1,013	0,774
22	0,791	0,846	-1,74	1,158	0,752
23	0,654	0,732	1,689	0,982	0,729
24	1,084	1,269	0,421	1,669	0,707
25	0,664	0,813	-0,849	1,050	0,685
26	0,513	0,658	-0,447	0,834	0,662
27	0,673	0,904	-0,912	1,127	0,640
28	0,467	0,658	0,12	0,807	0,617
29	0,571	0,843	-1,287	1,018	0,595
30	0,437	0,678	0,425	0,807	0,573
31	0,745	1,213	0,699	1,424	0,551
32	0,426	0,731	1,98	0,846	0,528
33	0,656	1,184	0,464	1,354	0,506
34	0,633	1,206	-0,605	1,362	0,483
35	0,443	0,892	-1,103	0,996	0,461
36	0,427	0,911	1,332	1,006	0,438
37	0,565	1,278	-0,19	1,397	0,416
38	0,455	1,096	-0,986	1,187	0,393
39	0,419	1,075	-0,093	1,154	0,372
40	0,546	1,5	1,001	1,596	0,349

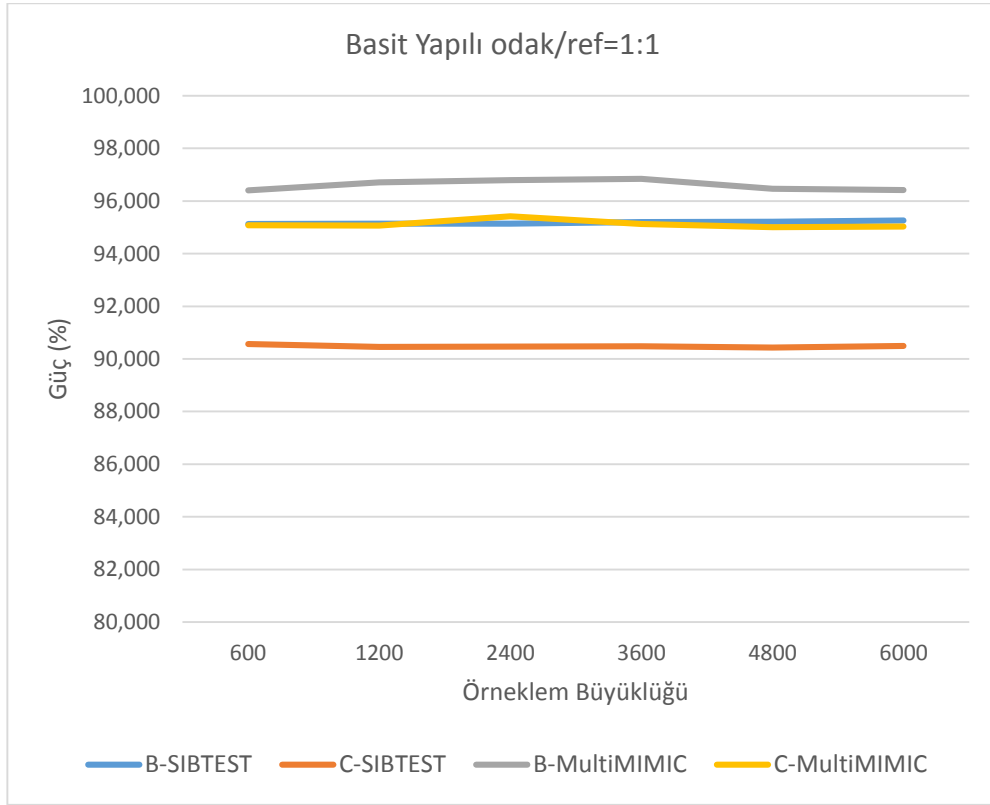
EK-A3: Madde Parametreleri Üç Boyutlu Basit Yapılı Test

Madde	a1	a2	a3	D	MDISC	Açı
1	1,270	0,001	0,083	-0,580	1,273	0,065
2	1,082	0,012	0,002	0,401	1,082	0,011
3	1,688	0,037	0,049	-0,185	1,689	0,036
4	1,332	0,036	0,069	-0,533	1,334	0,058
5	1,180	0,023	0,094	-0,566	1,184	0,082
6	1,108	0,035	0,001	0,466	1,109	0,032
7	1,700	0,035	0,051	-0,222	1,701	0,036
8	1,442	0,023	0,079	-0,545	1,444	0,057
9	1,001	0,036	0,078	-0,245	1,005	0,086
10	0,962	0,043	0,099	-0,195	0,968	0,112
11	0,934	0,050	0,099	-0,736	0,941	0,118
12	1,213	0,076	0,101	0,405	1,220	0,104
13	1,130	0,081	0,095	0,074	1,137	0,110
14	0,811	0,065	0,089	-0,157	0,818	0,135
15	0,266	1,186	0,320	0,715	1,257	0,337
16	0,227	1,110	0,115	0,074	1,139	0,225
17	0,149	0,820	0,201	-0,247	0,857	0,296
18	0,255	1,198	0,345	0,754	1,272	0,344
19	0,234	1,131	0,121	0,099	1,161	0,229
20	0,153	0,881	0,232	-0,266	0,924	0,306
21	0,221	1,316	0,106	-1,599	1,339	0,184
22	0,161	1,021	0,197	0,514	1,052	0,244
23	0,107	0,881	0,189	-0,219	0,907	0,242
24	0,102	0,936	0,150	0,464	0,953	0,191
25	0,087	0,959	0,078	-0,103	0,966	0,121
26	0,112	1,565	0,101	-0,340	1,572	0,096
27	0,059	1,069	0,089	0,320	1,074	0,100
28	0,034	0,194	1,112	-0,279	1,129	0,175
29	0,151	0,223	1,008	0,422	1,043	0,261
30	0,199	0,201	1,607	-0,222	1,632	0,174
31	0,044	0,208	1,009	-0,688	1,031	0,208
32	0,185	0,208	1,118	0,532	1,152	0,244
33	0,201	0,261	1,307	-0,199	1,348	0,247
34	0,108	0,228	1,295	-0,503	1,319	0,192
35	0,272	0,301	0,963	-0,343	1,045	0,399
36	0,245	0,176	0,831	-0,123	0,884	0,348
37	0,221	0,221	0,908	-0,726	0,960	0,332
38	0,281	0,332	1,036	-0,533	1,124	0,397
39	0,181	0,236	0,978	-0,845	1,022	0,295
40	0,122	0,103	0,999	-0,586	1,012	0,158

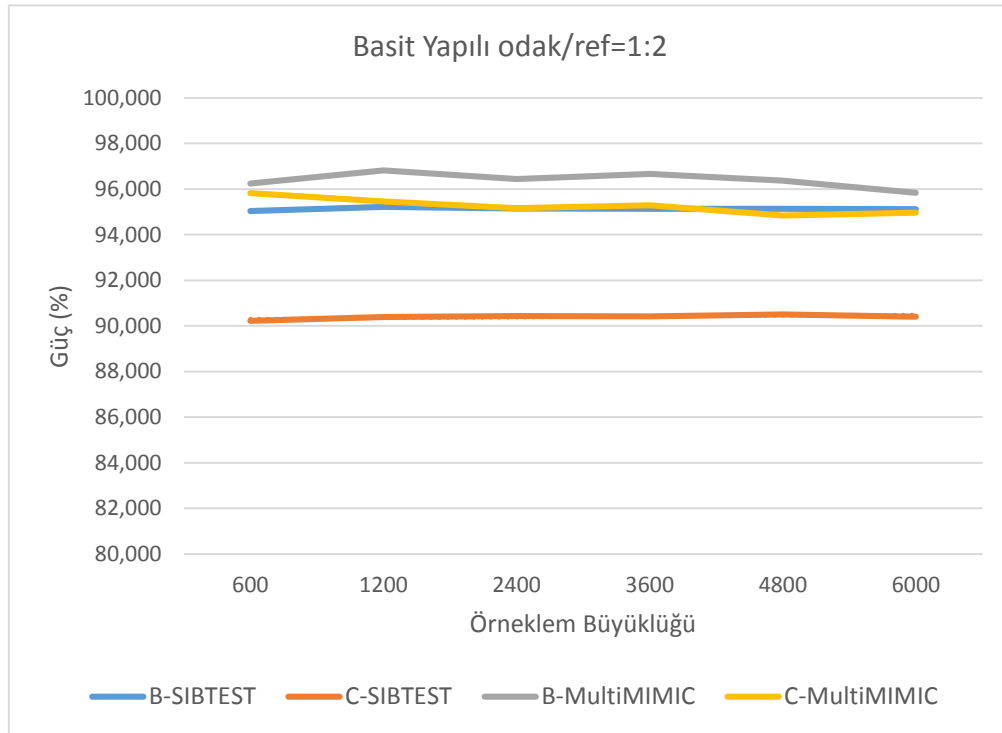
EK-A4: Madde Parametreleri Üç Boyutlu Karmaşık Yapılı Test

Madde	a1	a2	a3	D	MDISC	Açı
1	1,270	0,401	0,283	-0,830	1,362	0,369
2	1,082	0,612	0,202	0,421	1,259	0,537
3	1,688	0,037	0,549	-0,285	1,775	0,315
4	1,332	0,636	0,369	-0,523	1,521	0,504
5	1,001	0,436	0,078	-0,244	1,095	0,417
6	1,213	0,576	0,101	0,825	1,347	0,449
7	1,130	0,561	0,595	0,674	1,395	0,626
8	1,366	0,421	0,243	-0,920	1,450	0,342
9	1,182	0,513	0,222	0,433	1,308	0,442
10	1,458	0,337	0,549	-0,244	1,594	0,416
11	1,032	0,645	0,369	-0,523	1,272	0,624
12	0,962	0,043	0,499	-0,185	1,085	0,480
13	0,934	0,350	0,599	-0,736	1,163	0,639
14	0,811	0,365	0,289	-0,537	0,935	0,521
15	0,466	1,186	0,220	0,715	1,293	0,410
16	0,247	1,110	0,515	0,044	1,248	0,475
17	0,249	0,820	0,301	-0,257	0,908	0,445
18	0,432	1,144	0,254	0,722	1,249	0,413
19	0,347	1,020	0,526	0,233	1,199	0,553
20	0,244	1,029	0,501	-1,256	1,170	0,496
21	0,821	1,316	0,506	-1,499	1,632	0,632
22	0,461	1,021	0,337	0,514	1,170	0,510
23	0,407	0,881	0,389	-0,219	1,046	0,569
24	0,402	1,036	0,150	0,634	1,121	0,393
25	0,097	1,139	0,578	-1,124	1,281	0,475
26	0,512	1,262	0,301	-0,440	1,395	0,440
27	0,359	1,069	0,789	0,320	1,376	0,681
28	0,334	0,264	1,112	-0,379	1,191	0,366
29	0,151	0,623	1,108	1,152	1,280	0,525
30	0,679	0,271	1,607	1,232	1,765	0,427
31	0,580	0,240	1,295	-0,553	1,439	0,451
32	0,344	0,565	1,132	-0,399	1,311	0,529
33	0,222	0,658	1,078	1,352	1,282	0,572
34	0,655	0,233	1,307	1,279	1,480	0,489
35	0,272	0,601	0,963	-0,343	1,167	0,601
36	0,345	0,571	0,931	0,823	1,145	0,622
37	0,525	0,231	0,908	-1,226	1,074	0,563
38	0,281	0,732	1,036	-1,533	1,299	0,648
39	0,781	0,236	0,978	-1,845	1,274	0,695
40	0,522	0,603	0,999	-1,586	1,278	0,674

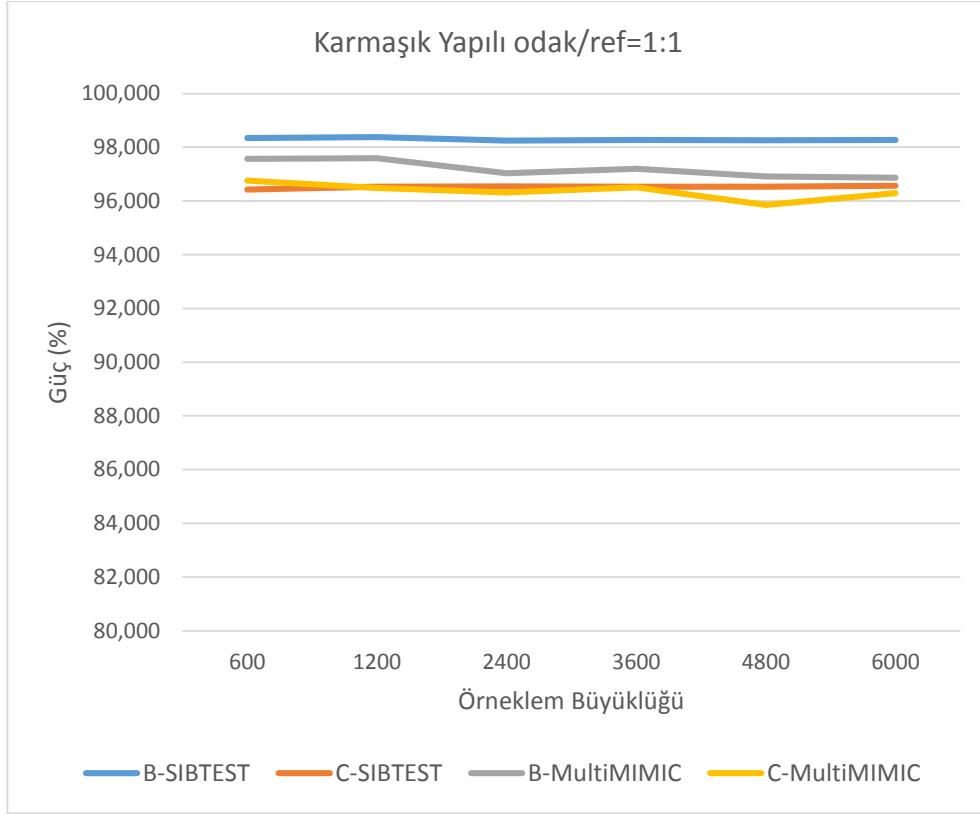
EK-B: Güç Oranlarına ait Grafikler



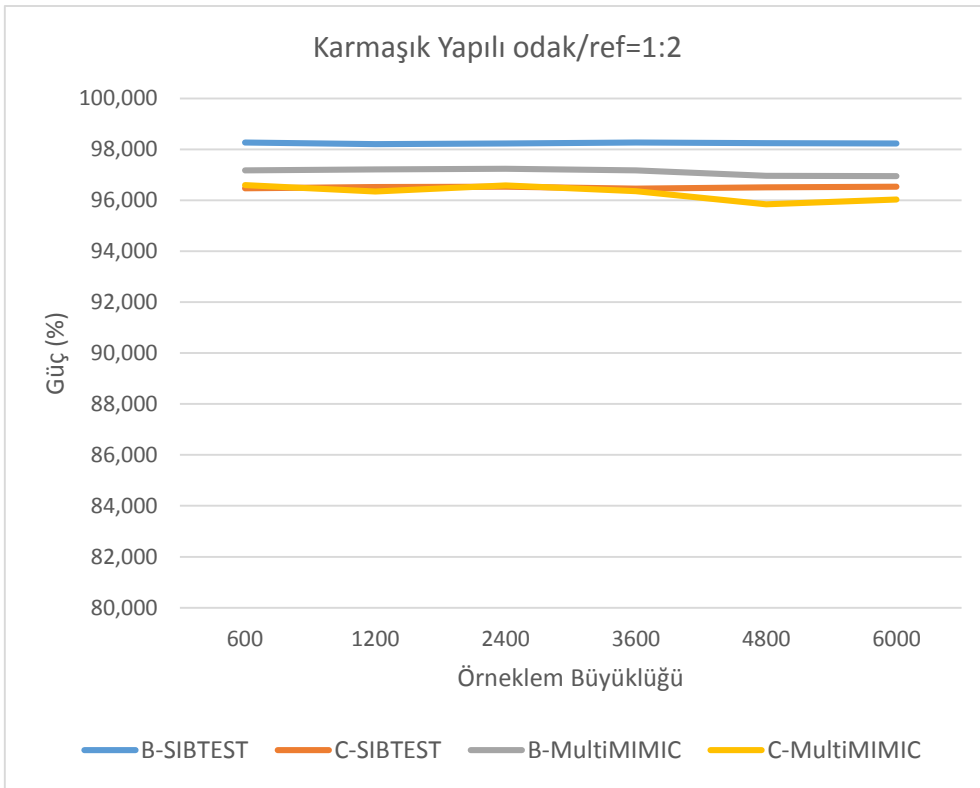
Şekil 26. İki boyutlu basit yapılu DMF (1:1) için güç oranları



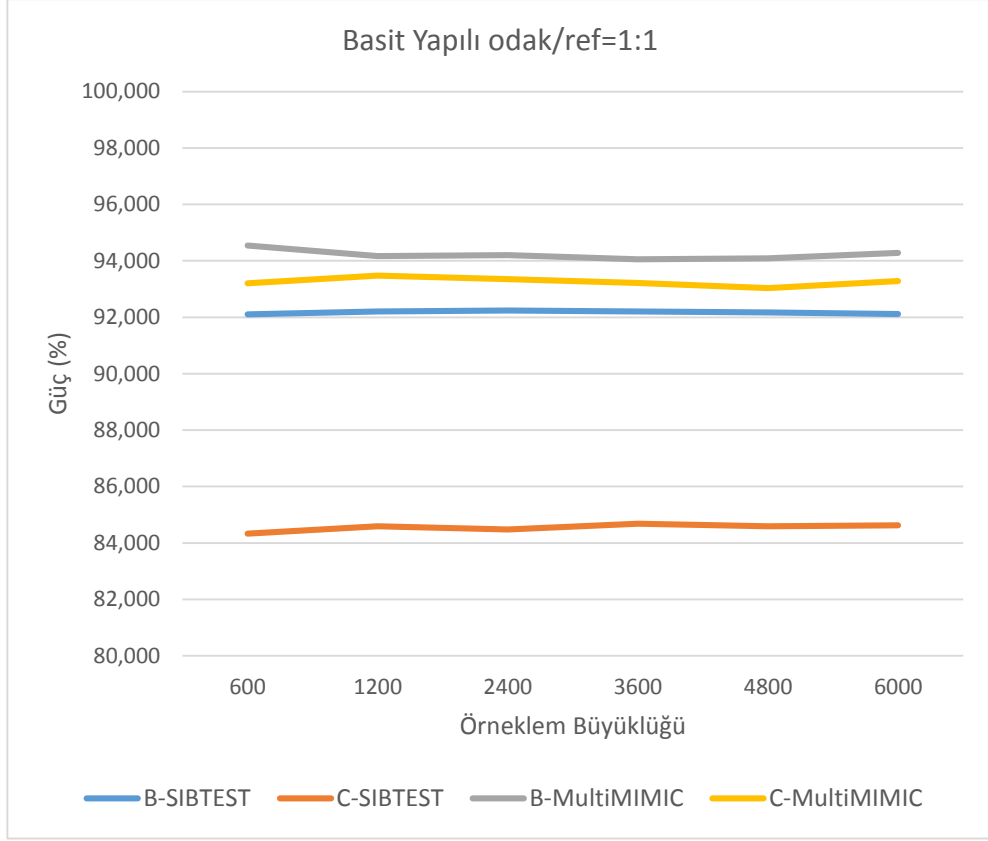
Şekil 27. İki boyutlu basit yapılu DMF (1:2) için güç oranları



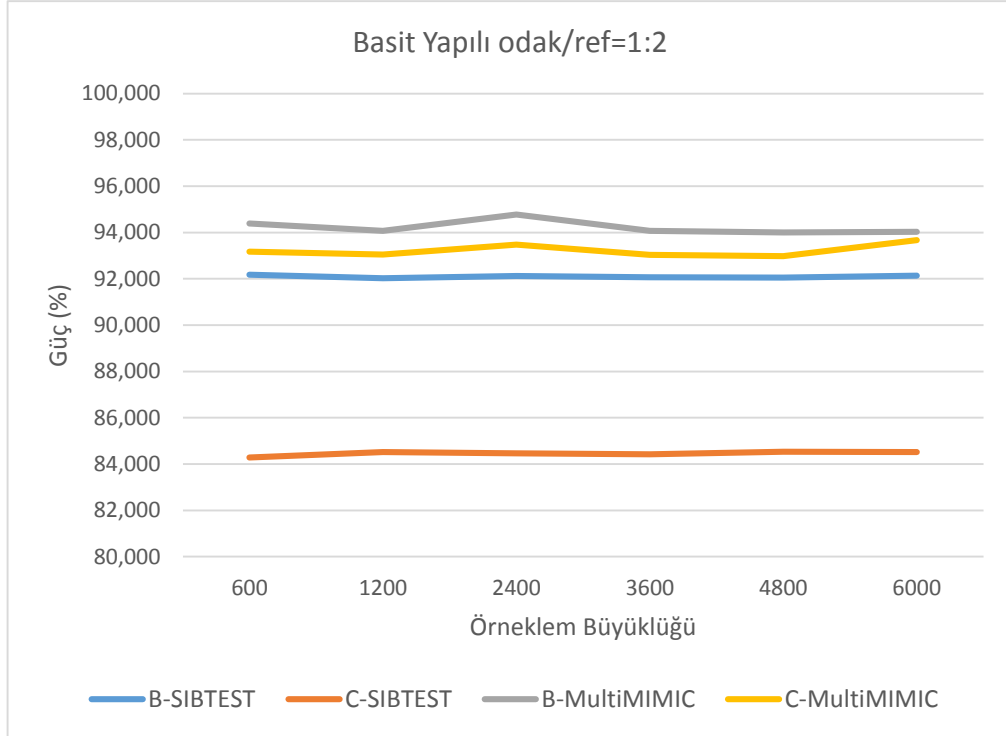
Şekil 28. İki boyutlu karmaşık yapı DMF (1:1) için güç oranları



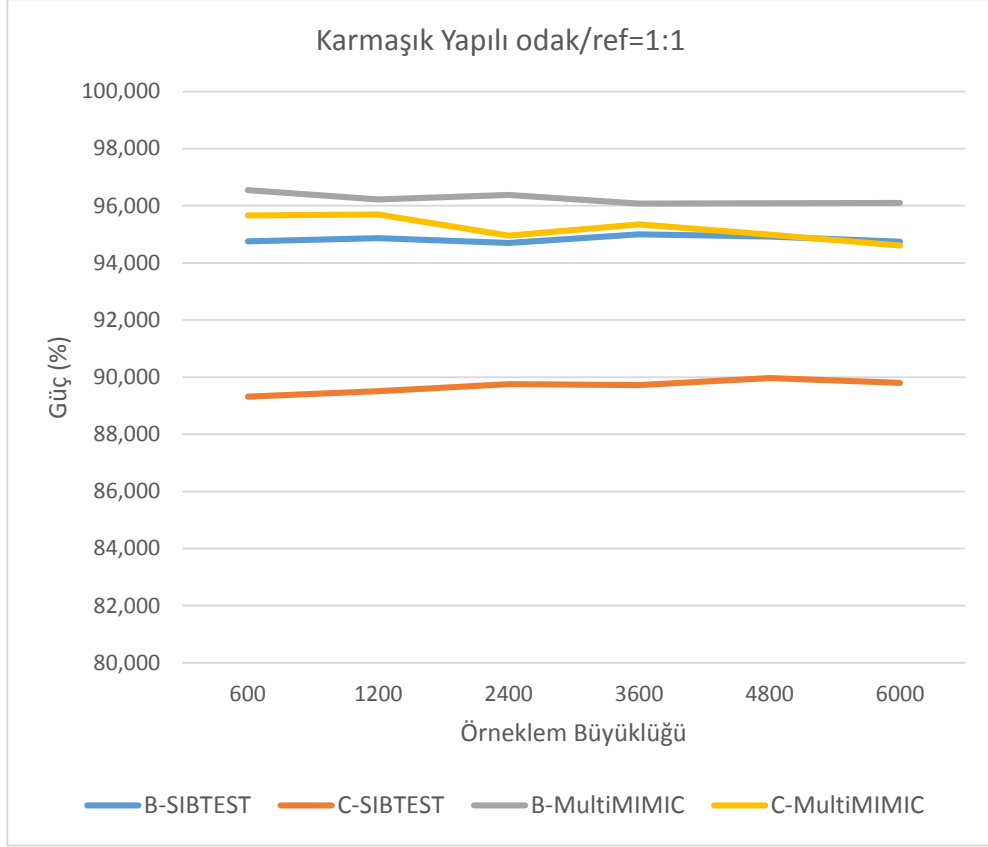
Şekil 29. İki boyutlu karmaşık yapı DMF (1:2) için güç oranları



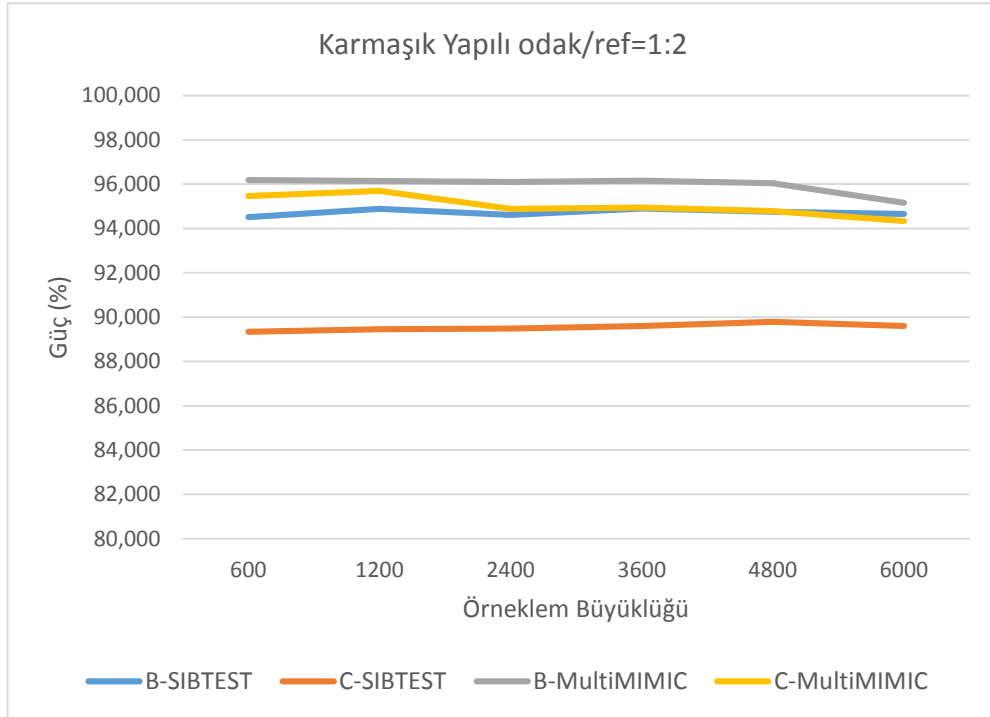
Şekil 30. Üç boyutlu basit yapılı DMF (1:1) için güç oranları



Şekil 31. Üç boyutlu basit yapılı DMF (1:2) için güç oranları



Şekil 32. Üç boyutlu karmaşık yapı DMF (1:1) için güç oranları



Şekil 33. Üç boyutlu karmaşık yapı DMF (1:2) için güç oranları

EK-C: Gerçek Veri Seti İçin DMF Analizleri Çıktıları

N= 600 Odak/Referans=1 (300-300)												
Madde No	SIBTEST			MultimIMIC			SIBTEST			MultimIMIC		
	Beta Est	P Value	Madde No	Beta Est	P Value	Madde No	Beta Est	P Value	Beta Est	P Value	P Value	
1	0.066	0.518	31	0.047	0.247	31	-0.05	0.922	-0.074	0.736	0.736	
2	0.065	0.043*	32	0.146	0.015*	32	-0.096	0.546	-0.065	0.428	0.428	
3	0.086	0.331	33	0.027	0.069	33	-0.013	0.835	-0.025	0.674	0.674	
4	-0.036	0.65	34	-0.116	0.41	34	-0.017	0.45	-0.022	0.252	0.252	
5	0.062	0.423	35	0.017	0.28	35	-0.056	0.448	-0.135	0.347	0.347	
6	0.115	0.387	36	0.062	0.366	36	-0.098	0.023*	-0.07	0.121	0.121	
7	0.057	0.518	37	0.079	0.302	37	0.061	0.056	0.069	0.029*	0.029*	
8	-0.048	0.597	38	-0.036	0.736	38	0.076	0.263	0.043	0.208	0.208	
9	0.039	0.918	39	0.014	0.125	39	-0.012	0.33	0.058	0.287	0.287	
10	0.05	0.525	40	0.102	0.469	40	-0.109	0.415	-0.09	0.369	0.369	
11	0.091	0.087	41	0.278	0.247	41	0.072	0.261	0.062	0.222	0.222	
12	0.073	0.041*	42	0.144	0.031*	42	-0.052	0.141	-0.269	0.169	0.169	
13	-0.056	0.717	43	-0.259	0.909	43	-0.049	0.047*	-0.025	0.046*	0.046*	
14	0.156	0.325	44	0.062	0.535	44	-0.038	0.065	-0.033	0.079	0.079	
15	0.013	0.291	45	0.017	0.203	45	0.05	0.774	0.169	0.717	0.717	
16	-0.043	0.041*	46	0.043	0.315	46	0.005	0.62	0.042	0.634	0.634	
17	0.088	0.051	47	0.05	0.169	47	0.102	0.708	0.153	0.756	0.756	
18	0.04	0.067	48	0.099	0.247	48	-0.041	0.869	-0.157	0.908	0.908	
19	0.078	0.268	49	0.202	0.271	49	-0.071	0.001*	-0.069	0.091	0.091	
20	-0.009	0.383	50	-0.19	0.464	50	-0.071	0.77	0.236	0.874	0.874	
21	0.063	0.718	51	0.049	0.814	51	-0.121	0.015*	-0.255	0.069	0.069	
22	0.087	0.213	52	0.074	0.169	52	-0.085	0.596	0.013	0.774	0.774	
23	0.101	0.008*	53	0.032	0.197	53	-0.03	0.329	0.066	0.426	0.426	
24	-0.09	0.379	54	-0.123	0.474	54	0.086	0.058	0.008	0.022*	0.022*	
25	0.049	1.013	55	0.047	0.935	55	0.014	0.412	0.047	0.528	0.528	
26	0.075	0.482	56	-0.102	0.436	56	-0.023	0.503	-0.069	0.563	0.563	
27	-0.087	0.257	57	-0.099	0.674	57	-0.028	0.489	-0.088	0.332	0.332	
28	0.036	0.173	58	-0.374	0.144	58	-0.069	0.026*	-0.074	0.117	0.117	
29	-0.143	0.048*	59	-0.063	0.02*	59	-0.087	0.269	-0.039	0.396	0.396	
30	0.098	0.479	60	-0.014	0.474	60	-0.062	0.236	-0.015	0.233	0.233	

N= 600 Odak/Referans=1/2 (200-400)

Madde No	SIBTEST		MultiMIMIC		SIBTEST		MultiMIMIC	
	Beta Est	P Value	Beta Est	P Value	Beta Est	P Value	Beta Est	P Value
1	0.045	0.385	0.065	0.347	-0.076	0.753	-0.025	0.832
2	0.051	0.045*	0.148	0.036*	-0.12	0.586	-0.021	0.476
3	0.063	0.148	-0.036	0.175	-0.02	0.875	0.036	0.723
4	-0.015	0.513	-0.112	0.47	-0.024	0.259	0.021	0.345
5	0.058	0.202	0.089	0.402	-0.038	0.434	0.088	0.336
6	0.094	0.25	0.055	0.453	-0.071	0.049*	0.105	0.075
7	0.032	0.345	0.069	0.565	0.096	0.039*	0.092	0.012*
8	-0.029	0.46	-0.074	0.813	0.074	0.247	0.084	0.106
9	0.008	0.769	0.085	0.517	-0.088	0.328	0.074	0.26
10	0.034	0.4	0.156	0.616	-0.051	0.474	-0.036	0.346
11	0.032	0.274	0.236	0.111	0.05	0.265	0.039	0.191
12	0.068	0.028*	0.147	0.029*	-0.028	0.024*	-0.37	0.174
13	-0.075	0.051	-0.236	0.8	-0.058	0.087	0.074	0.039*
14	0.107	0.222	-0.047	0.447	-0.029	0.143	0.006	0.136
15	0.069	0.188	0.016	0.3755	0.003	0.727	0.225	0.713
16	0.008	0.202	0.125	0.2	0.067	0.555	0.074	0.592
17	0.087	0.758	0.365	0.078	-0.038	0.686	-0.102	0.674
18	0.053	0.049*	0.074	0.114	-0.029	0.569	-0.136	0.866
19	-0.06	0.101	-0.095	0.156	-0.065	0.014*	-0.047	0.042*
20	-0.026	0.229	-0.105	0.377	-0.061	0.778	0.296	0.769
21	0.06	0.565	0.032	0.679	-0.047	0.023*	-0.026	0.083
22	0.085	0.183	0.074	0.156	-0.039	0.7	0.035	0.663
23	-0.069	0.035*	-0.01	0.048*	0.094	0.125	0.074	0.409
24	-0.071	0.698	-0.136	0.452	0.006	0.093	0.014	0.174
25	0.026	0.983	-0.003	1.038	0.027	0.42	0.076	0.449
26	0.049	0.448	-0.143	0.556	-0.04	0.511	0.005	0.049*
27	0.061	0.827	0.057	0.456	0.085	0.058	0.093	0.23
28	0.027	0.143	-0.321	0.247	-0.084	0.046*	0.147	0.117
29	0.117	0.048*	0.036	0.048*	-0.075	0.346	0.074	0.219
30	0.098	0.449	0.008	0.517	-0.088	0.128	0.056	0.182

N= 1200 Odak/Referans=1 (600-600)

Madde No	SIBTEST			MultiMIMIC			SIBTEST			MultiMIMIC		
	Beta Est	P Value	Madde No	Beta Est	P Value	Madde No	Beta Est	P Value	Madde No	Beta Est	P Value	Madde No
1	0.051	0.455	31	0.061	0.156	31	-0.061	0.899	31	-0.074	0.703	31
2	0.063	0.048*	32	0.169	0.049*	32	-0.108	0.518	32	-0.056	0.34	32
3	0.072	0.236	33	0.032	0.094	33	-0.036	0.874	33	-0.015	0.592	33
4	-0.014	0.592	34	-0.173	0.427	34	-0.031	0.922	34	-0.043	0.168	34
5	0.057	0.365	35	0.074	0.336	35	-0.074	0.42	35	-0.074	0.288	35
6	0.103	0.328	36	0.039	0.374	36	-0.097	0.031*	36	0.059	0.059	36
7	0.043	0.459	37	0.025	0.56	37	0.078	0.048*	37	0.047	0.011*	37
8	-0.036	0.574	38	-0.074	0.768	38	-0.082	0.233	38	0.038	0.118	38
9	0.028	0.859	39	0.069	0.474	39	-0.095	0.302	39	0.047	0.204	39
10	0.041	0.475	40	0.115	0.569	40	-0.058	0.458	40	-0.096	0.269	40
11	0.054	0.336	41	0.374	0.074	41	0.069	0.225	41	0.013	0.135	41
12	0.064	0.009*	42	0.146	0.362	42	-0.035	0.125	42	-0.374	0.145	42
13	-0.045	0.67	43	0.274	0.745	43	-0.028	0.039*	43	0.022	0.058	43
14	0.103	0.278	44	0.098	0.456	44	-0.036	0.852	44	-0.025	0.067	44
15	0.029	0.244	45	0.031	0.082	45	0.009	0.713	45	0.69	0.647	45
16	0.004	0.254	46	0.101	0.171	46	0.089	0.584	46	0.034	0.506	46
17	0.049	0.014*	47	0.065	0.231	47	-0.032	0.672	47	-0.145	0.663	47
18	0.031	0.327	48	0.074	0.001*	48	0.023	0.859	48	0.136	0.898	48
19	0.075	0.225	49	0.124	0.132	49	-0.078	0.02*	49	-0.039	0.017*	49
20	-0.002	0.285	50	-0.369	0.326	50	-0.055	0.768	50	0.247	0.756	50
21	0.069	0.67	51	0.061	0.639	51	-0.101	0.004*	51	-0.015	0.044*	51
22	0.019	0.115	52	0.066	0.174	52	-0.022	0.678	52	0.036	0.675	52
23	0.075	0.045*	53	0.014	0.171	53	-0.079	0.426	53	-0.047	0.474	53
24	-0.083	0.281	54	-0.114	0.435	54	0.022	0.316	54	0.021	0.169	54
25	0.026	0.912	55	0.027	0.974	55	0.101	0.393	55	0.093	0.474	55
26	0.061	0.384	56	-0.38	0.426	56	-0.022	0.469	56	-0.047	0.369	56
27	0.073	0.759	57	-0.085	0.57	57	-0.067	0.086	57	-0.136	0.25	57
28	0.028	0.075	58	-0.358	0.174	58	-0.096	0.028*	58	0.074	0.068	58
29	-0.129	0.008*	59	-0.074	0.045*	59	-0.058	0.326	59	-0.015	0.296	59
30	0.082	0.381	60	-0.014	0.469	60	-0.069	0.648	60	0.029	0.167	60

N= 1200 Odak/Referans=1/2 (400-800)

Madde No	SIBTEST			MultIMIMIC			SIBTEST			MultIMIMIC		
	Beta Est	P Value	Madde No	Beta Est	P Value	Madde No	Beta Est	P Value	Madde No	Beta Est	P Value	
1	0.066	0.518	31	0.047	0.247	31	-0.05	0.922	31	-0.074	0.736	
2	0.065	0.043*	32	0.146	0.015*	32	-0.096	0.546	32	-0.065	0.428	
3	0.086	0.331	33	0.027	0.069	33	-0.013	0.835	33	-0.025	0.674	
4	-0.036	0.65	34	-0.116	0.41	34	-0.017	0.45	34	-0.022	0.252	
5	0.062	0.423	35	0.017	0.28	35	-0.056	0.448	35	-0.135	0.347	
6	0.115	0.387	36	0.062	0.366	36	-0.098	0.023*	36	-0.07	0.121	
7	0.057	0.518	37	0.079	0.302	37	0.061	0.056	37	0.069	0.029*	
8	-0.048	0.597	38	-0.036	0.736	38	0.076	0.263	38	0.043	0.208	
9	0.039	0.918	39	0.014	0.125	39	-0.012	0.33	39	0.058	0.287	
10	0.05	0.525	40	0.102	0.469	40	-0.109	0.415	40	-0.09	0.369	
11	0.091	0.087	41	0.278	0.247	41	0.072	0.261	41	0.062	0.222	
12	0.073	0.041*	42	0.144	0.031*	42	-0.052	0.141	42	-0.269	0.169	
13	-0.056	0.717	43	-0.259	0.909	43	-0.049	0.047*	43	-0.025	0.046*	
14	0.156	0.325	44	0.062	0.535	44	-0.038	0.065	44	-0.033	0.079	
15	0.013	0.291	45	0.017	0.203	45	0.05	0.774	45	0.169	0.717	
16	-0.043	0.041*	46	0.043	0.315	46	0.005	0.62	46	0.042	0.634	
17	0.088	0.051	47	0.05	0.169	47	0.102	0.708	47	0.153	0.756	
18	0.04	0.067	48	0.099	0.247	48	-0.041	0.869	48	-0.157	0.908	
19	0.078	0.268	49	0.202	0.271	49	-0.071	0.001*	49	-0.069	0.091	
20	-0.009	0.383	50	-0.19	0.464	50	-0.071	0.77	50	0.236	0.874	
21	0.063	0.718	51	0.049	0.814	51	-0.121	0.015*	51	-0.255	0.069	
22	0.087	0.213	52	0.074	0.169	52	-0.085	0.596	52	0.013	0.774	
23	0.101	0.008*	53	0.032	0.197	53	-0.03	0.329	53	0.066	0.426	
24	-0.09	0.379	54	-0.123	0.474	54	0.086	0.058	54	0.008	0.022*	
25	0.049	1.013	55	0.047	0.935	55	0.014	0.412	55	0.047	0.528	
26	0.075	0.482	56	-0.102	0.436	56	-0.023	0.503	56	-0.069	0.563	
27	-0.087	0.257	57	-0.099	0.674	57	-0.028	0.489	57	-0.088	0.332	
28	0.036	0.173	58	-0.374	0.144	58	-0.069	0.026*	58	-0.074	0.117	
29	-0.143	0.048*	59	-0.063	0.02*	59	-0.087	0.269	59	-0.039	0.396	
30	0.098	0.479	60	-0.014	0.474	60	-0.062	0.236	60	-0.015	0.233	

N= 3600 Odak/Referans=1 (1800-1800)

Madde No	SIBTEST			MultitMIMIC			SIBTEST			MultitMIMIC		
	Beta Est	P Value	Madde No	Beta Est	P Value	Madde No	Beta Est	P Value	Madde No	Beta Est	P Value	
1	0.075	0.647	31	0.081	0.189	31	-0.041	0.146	31	-0.084	0.094	
2	0.121	0.031*	32	0.196	0.021*	32	-0.079	0.574	32	-0.074	0.058	
3	0.123	0.465	33	0.037	0.087	33	0.069	0.136	33	-0.066	0.075	
4	-0.047	0.178	34	-0.169	0.274	34	0.078	0.175	34	-0.087	0.101	
5	0.075	0.553	35	0.174	0.205	35	-0.038	0.753	35	-0.074	0.195	
6	0.126	0.515	36	0.069	0.336	36	-0.113	0.04*	36	0.045	0.069	
7	0.064	0.046*	37	0.174	0.474	37	-0.025	0.016*	37	0.047	0.036*	
8	-0.053	0.758	38	-0.036	0.669	38	0.083	0.329	38	-0.001	0.074	
9	0.033	0.277	39	0.102	0.345	39	-0.088	0.398	39	-0.007	0.136	
10	0.074	0.614	40	0.147	0.363	40	0.029	0.032*	40	0.114	0.174	
11	0.055	0.476	41	0.336	0.174	41	-0.066	0.335	41	-0.063	0.563	
12	0.074	0.034*	42	0.174	0.041*	42	0.064	0.287	42	0.074	0.079	
13	-0.066	0.178	43	-0.269	0.12	43	-0.099	0.019*	43	-0.065	0.136	
14	0.121	0.415	44	0.112	0.089	44	-0.045	0.169	44	-0.069	0.101	
15	0.074	0.38	45	0.062	0.075	45	-0.069	0.745	45	0.145	0.547	
16	0.04	0.009*	46	0.091	0.081	46	0.012	0.636	46	0.026	0.463	
17	0.068	0.021*	47	0.096	0.081	47	-0.127	0.774	47	-0.024	0.574	
18	0.027	0.036*	48	0.174	0.087	48	-0.048	0.11	48	-0.174	0.12	
19	0.098	0.356	49	0.196	0.123	49	-0.039	0.024*	49	-0.036	0.169	
20	-0.039	0.173	50	-0.356	0.374	50	-0.078	0.796	50	0.274	0.605	
21	0.09	0.207	51	0.087	0.672	51	-0.112	0.042*	51	-0.069	0.111	
22	0.057	0.135	52	0.069	0.145	52	0.099	0.716	52	0.037	0.59	
23	0.076	0.034*	53	0.074	0.0028*	53	-0.037	0.078	53	0.014	0.204	
24	-0.115	0.306	54	-0.06	0.332	54	-0.098	0.236	54	-0.059	0.097	
25	0.076	0.174	55	0.144	0.711	55	0.002	0.437	55	0.032	0.241	
26	-0.097	0.406	56	-0.195	0.213	56	-0.036	0.563	56	0.001	0.463	
27	0.105	0.728	57	-0.125	0.12	57	-0.04	0.494	57	0.027	0.202	
28	0.053	0.097	58	-0.36	0.091	58	-0.074	0.046*	58	-0.026	0.023*	
29	0.101	0.038*	59	-0.09	0.051	59	-0.036	0.005*	59	-0.047	0.238	
30	0.117	0.406	60	-0.03	0.095	60	-0.074	0.774	60	-0.069	0.132	

N= 3600 Odak/Referans=1/2 (1200-2400)

Madde No	SIBTEST		MultitMIMIC		SIBTEST		MultitMIMIC	
	Beta Est	P Value	Beta Est	P Value	Beta Est	P Value	Beta Est	P Value
1	0.068	0.447	0.063	0.078	-0.041	0.685	-0.074	0.274
2	0.098	0.019*	0.145	0.04*	-0.08	0.745	-0.081	0.225
3	0.097	0.274	0.039	0.062	-0.025	0.703	-0.047	0.489
4	-0.049	0.548	-0.174	0.121	-0.075	0.274	-0.095	0.101
5	0.074	0.357	0.095	0.312	-0.036	0.45	-0.106	0.226
6	0.136	0.316	0.069	0.351	-0.096	0.008*	-0.021	0.097
7	0.074	0.025*	0.096	0.171	-0.063	0.04*	-0.02	0.049*
8	-0.062	0.524	-0.042	0.205	0.094	0.214	-0.023	0.073
9	0.04	0.86	0.105	0.397	-0.077	0.274	-0.077	0.122
10	0.069	0.474	0.12	0.421	0.04	0.023*	-0.161	0.105
11	0.078	0.356	0.347	0.128	-0.054	0.225	-0.052	0.053
12	0.063	0.029*	0.122	0.036*	-0.054	0.077	-0.093	0.067
13	-0.07	0.669	-0.214	0.363	-0.077	0.007*	0.061	0.053
14	0.126	0.241	0.089	0.374	-0.017	0.118	-0.047	0.193
15	0.074	0.202	0.037	0.07	0.048	0.156	0.168	0.097
16	0.045	0.047*	0.036	0.116	0.016	0.174	0.021	0.056
17	0.069	0.051	0.054	0.174	-0.074	0.647	-0.147	0.274
18	0.057	0.049*	0.08	0.126	-0.014	0.254	-0.12	0.056
19	0.098	0.236	0.119	0.074	-0.025	0.027*	-0.049	0.056
20	-0.025	0.344	-0.124	0.427	-0.044	0.174	0.265	0.358
21	0.04	0.247	0.057	0.256	-0.102	0.038*	-0.061	0.021*
22	0.046	0.145	0.081	0.057	0.058	0.669	0.042	0.327
23	0.061	0.039*	0.069	0.109	-0.015	0.088	0.017	0.136
24	-0.174	0.374	-0.081	0.363	0.068	0.111	0.017	0.187
25	0.061	0.245	0.018	0.732	0.036	0.145	0.014	0.369
26	-0.083	0.432	-0.206	0.275	-0.074	0.502	-0.028	0.058
27	0.107	0.745	-0.117	0.081	-0.009	0.285	0.029	0.079
28	-0.06	0.157	-0.396	0.105	-0.065	0.018*	-0.012	0.075
29	0.11	0.038*	-0.098	0.027*	-0.067	0.061	-0.023	0.057
30	0.116	0.455	-0.037	0.263	-0.048	0.194	-0.052	0.069

EK-Ç: SAS Programı Kodları

```
data parms;
    INFILE "C:\bulut\parms.txt";
    INPUT a1 a2 d1;
run;
proc iml;
    call randseed(0);
    use parms;
    read all var {a1 a2} into a;
    read all var {d1} into d;

    mean = {0,0,50}; *change these values to transform theta1 and
theta2 to a different mean for the focal group;
    cov = {1 0,30,
           0,30 1}; *change the covariance to modify the
correlation between dimensions for the focal group;
    theta = RANDNORMAL(2000,mean,cov);

    z=j(2000,40);
    p=j(2000,40);
    u=j(2000,40);
    do j=1 to 2000;
        do i=1 to 40;
            p[j,i]= exp(a[i,]*theta`[,j]+d[i,])/
                (1+(exp(a[i,]*theta`[,j]+d[i,])));

            u[j,i]=ranuni(0);
            if p[j,i]<u[j,i] then z[j,i]=0;
            else if p[j,i]>=u[j,i] then z[j,i]=1;
        end;
    end;
    create response from z;
    append from z;
    names1={theta1 theta2};
    create theta from theta [colname=names1];
    append from theta;
    quit;

data keep_foc;
    set response;
    group = 1;
    file "C:\bulut\foc_data.txt";
    put (col1-col40) (1.0);
run;
proc iml;
    call randseed(0);
    use parms;
    read all var {a1 a2} into a;
    read all var {d1} into d;

    mean = {0,-0,50}; *change these values to transform theta1 and
theta2 to a different mean for the reference group;
    cov = {1 0,30,
           0,30 1}; *change the covariance to modify the
correlation between dimensions for the reference group;
```

```

theta = RANDNORMAL(4000,mean,cov);
z=j(4000,40);
p=j(4000,40);
u=j(4000,40);
do j=1 to 4000;
  do i=1 to 40;
    p[j,i]= exp(a[i,]*theta`[,j]+d[i,])/
      (1+(exp(a[i,]*theta`[,j]+d[i,])));

    u[j,i]=ranuni(0);
    if p[j,i]<u[j,i] then z[j,i]=0;
    else if p[j,i]>=u[j,i] then z[j,i]=1;
  end;
end;
create response from z;
append from z;
names1={theta1 theta2};
create theta from theta [colname=names1];
append from theta;
quit;
data keep_ref;
  set response;
  group = 0;
  file "C:\bulut\ref_data.txt";
  put (col1-col40) (1.0);
run;
data all;
  set keep_ref keep_foc;
  rename col1-col40 = i1-i40;
  total = sum(of col1-col40);
  subtot=sum(of col1 - col40);
  file "C:\bulut\all data.txt";
  put (col1-col40) (2.0) group total subtot;
run;

GETPOWER;
data allsib;
  infile 'C:\bulut\cond12\output\allsib.out' ignoredoseof;

if pbeta1 < 0,05 then powersib1=1; else powersib1=0;
if pbeta2 < 0,05 then powersib2=1; else powersib2=0;
if pbeta3 < 0,05 then powersib3=1; else powersib3=0;
if pbeta4 < 0,05 then powersib4=1; else powersib4=0;
if pbeta5 < 0,05 then powersib5=1; else powersib5=0;
if pbeta6 < 0,05 then powersib6=1; else powersib6=0;
if pbeta7 < 0,05 then powersib7=1; else powersib7=0;
if pbeta8 < 0,05 then powersib8=1; else powersib8=0;
if pbeta9 < 0,05 then powersib9=1; else powersib9=0;
if pbeta10 < 0,05 then powersib10=1; else powersib10=0;
if pbeta11 < 0,05 then powersib11=1; else powersib11=0;
if pbeta12 < 0,05 then powersib12=1; else powersib12=0;
if pbeta13 < 0,05 then powersib13=1; else powersib13=0;
if pbeta14 < 0,05 then powersib14=1; else powersib14=0;
if pbeta15 < 0,05 then powersib15=1; else powersib15=0;
if pbeta16 < 0,05 then powersib16=1; else powersib16=0;
if pbeta17 < 0,05 then powersib17=1; else powersib17=0;

```

```

if pbeta18 < 0,05 then powersib18=1; else powersib18=0;
if pbeta19 < 0,05 then powersib19=1; else powersib19=0;
if pbeta20 < 0,05 then powersib20=1; else powersib20=0;
if pbeta21 < 0,05 then powersib21=1; else powersib21=0;
if pbeta22 < 0,05 then powersib22=1; else powersib22=0;
if pbeta23 < 0,05 then powersib23=1; else powersib23=0;
if pbeta24 < 0,05 then powersib24=1; else powersib24=0;
if pbeta25 < 0,05 then powersib25=1; else powersib25=0;
if pbeta26 < 0,05 then powersib26=1; else powersib26=0;
if pbeta27 < 0,05 then powersib27=1; else powersib27=0;
if pbeta28 < 0,05 then powersib28=1; else powersib28=0;
if pbeta29 < 0,05 then powersib29=1; else powersib29=0;
if pbeta30 < 0,05 then powersib30=1; else powersib30=0;
if pbeta31 < 0,05 then powersib31=1; else powersib31=0;
if pbeta32 < 0,05 then powersib32=1; else powersib32=0;
if pbeta33 < 0,05 then powersib33=1; else powersib33=0;
if pbeta34 < 0,05 then powersib34=1; else powersib34=0;
if pbeta35 < 0,05 then powersib35=1; else powersib35=0;
if pbeta36 < 0,05 then powersib36=1; else powersib36=0;
if pbeta37 < 0,05 then powersib37=1; else powersib37=0;
if pbeta38 < 0,05 then powersib38=1; else powersib38=0;
if pbeta39 < 0,05 then powersib39=1; else powersib39=0;
if pbeta40 < 0,05 then powersib40=1; else powersib40=0;

filename out1 'C:\bulut\cond12\output\beta_stats';
file out1;
put (beta1-beta40) (10,3);
filename out2 'C:\bulut\cond12\output\pbeta_stats';
file out2;
put (pbeta1-pbeta40) (10,3);
run;
proc means data=all sib;
var beta1-beta40 powersib1-powersib40;
run;
FOR LOOPING;
options noprint nonotes;
%let nreps=100;
%let cond = 12;
%macro reps;
x "copy C:\bulut\empty.dat
C:\bulut\cond&cond\output\allsib.out";
* x "copy C:\ncme15\empty.dat
c:\ncme15\cond&cond\output\all_lr_chi.out";
*x "copy C:\ncme15\empty.dat
c:\ncme15\cond&cond\output\all_lr_rsq.out";
*x "copy C:\ncme15\empty.dat c:\ncme15\all_lr_chi.out";
* x "copy C:\ncme15\empty.dat c:\ncme15\all_lr_rsq.out";
%do k = 1 %to &nreps;
%put "starting loop &k";
%include 'C:\bulut\gen_mean_diff.sas';
x "C:\bulut\run_sib";
x "copy C:\bulut\ref_data.txt
C:\bulut\cond&cond\refdata\REF&k..dat";
x "copy C:\bulut\foc_data.txt
C:\bulut\cond&cond\focdata\FOC&k..dat";

```



```

        x "copy C:\bulut\sib.out
C:\bulut\cond&cond\output\sib&k..out";
        *x "copy c:\ncme15\lrchi.out
c:\ncme15\cond&cond\output\lrchi&k..out";
        * x "copy c:\ncme15\lr_rsqa.out
c:\ncme15\cond&cond\output\lr_rsqa&k..out";
        x "cd C:\bulut\cond&cond\output";
        x "copy allsib.out+sib&k..out allsib.out";
        *x "copy all_lr_chi.out+lrchi&k..out all_lr_chi.out";
        *x "copy all_lr_rsqa.out+lr_rsqa&k..out all_lr_rsqa.out";
        %include 'C:\bulut\getpower.sas';
    %put "ending loop &k";
%end;
%mend;

options noxwait;
%reps

```

EK-D: Multi-MIMIC Yöntemi için MPLUS Kodu

```
title: detecting dif with multidimensional mimic-interaction model
data:
file is bulutpısadata.dat;
format is free;
type is individual;
ngroups=i;
variable:
!60 dichotomous items and one group variable
names are item1-item60 z;
categorical are item1-item60;
model:
!latent trait i;
[thetai@i];
thetai
thetai@i;
thetai on z;
!uniform dif;
itemii on z;
zxthetal | z xwith thetal;
item on zxthetal;
!latent trait 2-3;
[theta2@0];
thetai@i; thetaii on z;
analysis: estimator = mlr;
type = random;...
```

EK-E: Etik Komisyonu Onay Bildirimi



T.C.
HACETTEPE ÜNİVERSİTESİ
Rektörlük

Tarih: 16.01.2019 18:35
Sayı: 35853172-300-E.00000413093

E0000413093

Sayı : 35853172-300
Konu : Bulut YILDIZTEKİN Hk.

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 21.12.2018 tarihli ve 51944218-300/00000380844 sayılı yazı.

Enstitünüz Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı doktora programı öğrencilerinden Bulut YILDIZTEKİN'in Prof. Dr. Hülya KELECİOĞLU danışmanlığında yürüttüğü "Test boyutluluğundan değişen madde fonksiyonuna etkisinin farklı koşullar altında incelenmesi" başlıklı tez çalışması, Üniversitemiz Senatosu Etik Komisyonunun 8 Ocak 2019 tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini saygılarımla rica ederim.

e-İmzalıdır
Prof. Dr. Rahime Meral NOHUTCU
Rektör Yardımcısı

Evrakın elektronik imzalı suretine <https://belgedogrulama.hacettepe.edu.tr> adresinden cd80248b-5759-47ad-abd4-ac057ce-4f27d kodu ile erişebilirsiniz.
Bu belge 5070 sayılı Elektronik İmza Kanunu'na uygun olarak Güvenli Elektronik İmza ile imzalanmıştır.

Hacettepe Üniversitesi Rektörlük 06100 Sıhhiye-Ankara
Telefon:0 (312) 305 3001-3002 Faks:0 (312) 311 9992 E-posta: yazim@hacettepe.edu.tr İnternet
Adresi: www.hacettepe.edu.tr

Doç. Dr. Didem İLFRİ

