ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI

DEĞİŞEN MADDE FONKSİYONUNU KONTROL ALTINA ALMADA ÇOK BOYUTLU MADDE TEPKİ KURAMI MODELLERİNİN KULLANIMI

DOKTORA TEZİ

HATİCE GÜRDİL

ANKARA
EKİM, 2023


ANKARA ÜNİVERSİTESİ
1946

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI

DEĞİŞEN MADDE FONKSİYONUNU KONTROL ALTINA ALMADA ÇOK BOYUTLU MADDE TEPKİ KURAMI MODELLERİNİN KULLANIMI

DOKTORA TEZİ

HATİCE GÜRDİL

DANIŞMAN: DOÇ. DR. ERGÜL DEMİR

ANKARA
EKİM, 2023

Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Hatice Gürdil adlı öğrencinin hazırladığı "Değişen Madde Fonksiyonunu Kontrol Altına Almada Çok Boyutlu Madde Tepki Kuramı Modellerinin Kullanımı" başlıklı bu çalışma Eğitim Bilimleri Anabilim Dalı / Eğitimde Ölçme ve Değerlendirme Doktora Programı'nda jüri üyelerince oy birliği ile **Doktora Tezi** Olarak Kabul Edilmiştir.

Jüri Üyeleri		İmza
Başkan	Prof. Dr. Duygu ANIL	……...
Üye	Doç. Dr. Ergül DEMİR (Danışman)	……...
Üye	Prof. Dr. Burcu ATAR	……...
Üye	Doç. Dr. C. Deha DOĞAN	……...
Üye	Doç. Dr. Seher YALÇIN	……...

ONAY

Bu tez Ankara Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca jüri üyeleri tarafından ../11/2023 tarihinde, Enstitü Yönetim Kurulunca …/02/2023 tarihinde kabul edilmiştir.

……...

Prof. Dr. Mehmet İkbal YETİŞİR

Eğitim Bilimleri Enstitüsü Müdürü

ETİK İLKELERE UYGUNLUK BİLDİRİMİ

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

Hatice GÜRDİL

ÖZET

DEĞİŞEN MADDE FONKSİYONUNU KONTROL ALTINA ALMADA ÇOK BOYUTLU MADDE TEPKİ KURAMI MODELLERİNİN KULLANIMI

GÜRDİL, Hatice
Doktora Tezi, Eğitim Bilimleri Anabilim Dalı
Tez Danışmanı: Doç. Dr. Ergül DEMİR
Ekim, 2023, xvi + 150 Sayfa

Bu çalışmanın amacı, farklı simülasyon koşularında Değişen Madde Fonksiyonu (DMF) gösteren maddeler nedeniyle tek boyutluluğun sağlanamadığı testlerde, Tek Boyutlu (TB) ve Çok Boyutlu Madde Tepki Kuramı (ÇBMTK) kestirimlerinin DMF'yi kontrol altına almaya yönelik etkilerinin incelenmesidir. Çalışmanın amacı doğrultusunda boyut sayısı (2, 3), DMF yüzdesi (%10, %20, %30), DMF büyüklüğü (küçük, büyük) ve DMF türü (tek biçimli, tek biçimli olmayan) olmak üzere döngüsel olarak 2x3x2x2=24 farklı deneysel koşul için 24x100 tekrar=2400 farklı veri seti üretilmiştir. Boyut sayısının, DMF yüzdesinin, DMF büyüklüğünün ve DMF türünün değişimlenmesine bağlı olarak; DMF belirleme tekniklerinin (Lojistik Regresyon-LR, Simultaneous item bias test-SIBTEST, Lord'un χ^2) performanslarını değerlendirmek üzereTip I hata ve istatistiksel güç oranlarındaki farklılaşmalar incelenmiştir. Verilerin üretiminde, Tek ve Çok Boyutlu MTK'ya dayalı yetenek parametrelerinin kestiriminde, DMF analizlerinin yapılmasında, Tip I hata ve istatistiksel güç oranlarındaki farklılaşmaların incelenmesinde ve grafiklerin çiziminde R programlama dili kullanılmıştır. Simülasyon koşullarına bağlı olarak üretilen veriler için MTK varsayımları kapsamında boyutluluk ve model-veri uyumu analizleri yapılmıştır. Buna ek olarak simülasyon verilerinin geçerliği için DMF içerecek şekilde üretilen verilerin istenilen DMF yüzdesi, büyüklüğü ve biçimini sağlama durumları incelenmiş, istatistiksel ve mantıksal kontrolleri yapılmıştır. TB ve TBO DMF türleri için ayrı ayrı olmak üzere, simülasyon koşullarının her birinde Tek ve Çok Boyulu MTK kestirimlerine yönelik DMF analizlerinin Tip I hata ve istatistiksel güç oranlarındaki farklılaşmaları incelenmiştir. Farklılaşmaların belirlenmesinde Faktöriyel ANOVA

analizi yapılmıştır. Analiz sonucunda TB ve TBO DMF türü için, çok boyutlu kestirim yöntemi ile tek boyutlu kestirim yöntemine göre daha düşük Tip 1 hata değerleri ve daha yüksek istatistiksel güç değerleri elde edilmiş; elde edilen farkların manidar ve yüksek düzey etkiye sahip olduğu sonucuna ulaşılmıştır. Elde edilen bu sonuç çok boyutlu kestirimlerin, tek boyutlu kestirimlere göre düzeltilmiş puan elde edilmesinde daha etkili olduğu şeklinde yorumlanabilmektedir. Gerçek veri analizinde de tek ve çok boyutlu kestirimlere ait Tip I hata ve istatistiksel güç oranları arasında, çok boyutlu kestirim lehine manidar farklılıklar olduğu sonucuna ulaşılmıştır. Elde edilen sonuçlar birlikte değerlendirildiğinde çok boyutlu kestirimlerin, tek boyutlu kestirimlere göre düzeltilmiş puan elde edilmesinde etkili olduğu görülmüştür. Diğer bir ifadeyle çok boyutlu kestirimlerin DMF'yi kontrol altına almada, tek boyutlu kestirimlerden daha etkili olduğu ve bu sonucun gerçek verilerle desteklendiği söylenebilmektedir. Bu kapsamda DMF kaynaklı çok boyutlu olan uygulanmış bir testte madde çıkarmadan çok boyutlu kestirim yapılarak düzeltilmiş toplam puan elde edilmesi önerilmektedir. Buna ek olarak TB ve TBO DMF türü için; DMF büyüklüğünün artması; TB DMF türü için ise DMF yüzdesi artması çok boyutlu kestirim Tip I hata oranları üzerinde daha fazla düzeltme sağlamaktadır. Hem TB hem de TBO DMF türü için; DMF büyüklüğü ve boyut sayısı arttıkça; TBO DMF türü içinse DMF yüzdesi arttıkça çok boyutlu kestirim istatistiksel güç oranları üzerinde daha fazla düzeltme sağlamaktadır. Bu sonuçlara ek olarak TB ve TBO DMF türü için; çok boyutlu kestirim sonucu tüm DMF belirleme tekniklerinde tek boyutlu kestirimden daha düşük Tip I hata ve daha yüksek istatistiksel güç oranları elde edilmiştir. Bir diğer sonuç olarak; tek boyutlu kestirimde belirlenen koşulların alt düzeyleri arasındaki farklılıklar belirginken çok boyutlu kestim sonucunda farklılıklar azalarak düzeyler birbirine benzer çalışmaktadır.

Anahtar Sözcükler: TBMTK yetenek parametresi kestirimi, ÇBMTK yetenek parametresi kestirimi, DMF Analizleri, MCMC, R programlama dili

ABSTRACT

THE USE OF MULTIDIMENSIONAL ITEM RESPONSE THEORY MODELS IN CONTROLLING DIFFERENTIAL ITEM FUNCTIONING

GÜRDİL, Hatice

Doctoral Dissertation, Department of Educational Sciences

Supervisor: Assoc. Prof. Dr. Ergül DEMİR

October, 2023, xvi + 150 Pages

The purpose of this study is to investigate the effects of Unidimensional (UD) and Multidimensional Item Response Theory (MIRT) Estimations on controlling Variability of Differential Item Functioning (DIF) in tests where unidimensionality cannot be ensured due to tests with DIF in different simulation conditions. In line with the aim of the study, 24 different experimental conditions, alternatingly for the number of dimensions (2, 3), DIF percentage (10%, 20%, 30%), DIF size (small, large), and DIF type (Uniform_UDIF, Nonuniform_NUDIF) 2x3x2x2 = 24x100 repetitions = 2400 different data sets were generated. Depending on the changes in the number of dimensions, DIF percentage, DIF size, and DIF type; differences in the performance of DIF determination techniques (Logistic Regression-LR, Simultaneous item bias test-SIBTEST, Lord's $\chi 2$) in terms of Type I error and statistical power rates were examined. In the generation of data, estimating ability parameters based on Unidimensional and Multidimensional IRT, examining differences in Type I error and statistical power rates for performing DIF analyses, and drawing graphs R programming language were used. In line with IRT assumptions, dimensionality and model-data fit analyses were conducted for data generated under simulation conditions. In addition, for the validity of simulation data, cases providing the desired DIF percentage, DIF size, and DIF type -produced data were examined, and statistical and logical checks were performed. UDIF and NUDIF types, differences in Type I error and statistical power rates of DIF analyses for Unidimensional and Multidimensional IRT estimates in each simulation condition were separately examined. Factorial ANOVA analysis was used to determine the differences. As a result of the analysis, for UDIF and NUDIF types, lower Type 1 error values and

higher statistical power values were obtained for the multidimensional estimation method compared to the unidimensional estimation method, and it was concluded that the obtained differences were significant and had a high level of effect. These results can be interpreted as indicating that multidimensional estimations are more effective in obtaining corrected scores compared to unidimensional estimations. In real data analysis, it was concluded that there were significant differences in favor of multidimensional estimations in terms of Type I error and statistical power rates for both unidimensional and multidimensional estimations. When the results are evaluated together, it is seen that multidimensional estimations are effective in controlling DIF compared to unidimensional estimations, and this result is supported by real data. In this context, in a test with multidimensional DIF, it is recommended to obtain corrected total scores by making multidimensional estimations without removing items. In addition, for UDIF and NUDIF DIF types; Increasing DIF size; For UDIF DIF type, increasing DIF percentage provides more correction in multidimensional estimation Type I error rates. For both UDIF and NUDIF DIF types; Increasing DIF size and the number of dimensions, and for NUDIF DIF type, increasing DIF percentage provide more correction in multidimensional estimation statistical power rates. In addition to these results, for UDIF and NUDIF DIF types; lower Type I error and higher statistical power rates were obtained in multidimensional estimation results for all DIF determination techniques compared to unidimensional estimation. Another result is that while the differences between the sub-levels determined in unidimensional estimation are prominent, the differences decrease and the levels work more similarly in the multidimensional estimation results.

Keywords: Unidimensional IRT ability parameter estimation, Multidimensional IRT ability parameter estimation, DIF analyses, MCMC, R programming language.

ÖNSÖZ

Yüksek lisans ve doktora eğitimim süresince bilgi ve deneyimlerinden yararlandığım, öğrencilerine yeni ufuklar açan, akademik hayatımın ilk basamaklarında büyük emek ve pay sahibi olan değerli danışman hocam sayın Doç. Dr. Ergül DEMİR'e,

Doktora çalışmam süresince bilgi ve deneyimlerinden yararlandığım TİK üyeleri sayın Prof. Dr. Duygu ANIL ve sayın Doç. Dr. Seher YALÇIN'a,

Değerli görüş ve önerileriyle tezime katkı sağlayan değerli jüri üyeleri ve saygıdeğer hocalarım Prof. Dr. Burcu ATAR'a ve Doç. Dr. Deha DOĞAN'a,

Yüksek lisans ve doktora eğitimim boyunca akademik gelişimime katkı ve teşvikte bulunan kıymetli hocalarıma,

Bu süreçteki yol arkadaşım, dert ortağım Yeşim Beril SOĞUKSU'ya,

Akademik hayatımı keyifli kılan çok değerli arkadaşlarım Ceren TUNABOYLU DEMİR, Filiz İSNAÇ, Yasemin Duygu ESEN, Hatice Özlem ANADOL, Fatma ÇOŞKUN, Ayşe BİLİCİOĞLU GÜNEŞ, Emre KUCAM, Serpil DURMUŞ AKGEYİK, Muharrem ŞENGÜL, Eray SELÇUK ve Ömer KAMIŞ'a,

Tezden yorulup vazgeçmek üzere olduğumda sürekli "Hadi az kaldı" diyen, en büyük motivasyon kaynağım Neslihan ESER'e,

Eğitim-öğretim hayatımın ilk aşamalarından itibaren bugünlere gelmeme kadar en büyük emeğe sahip olan, süreç içerisinde her türlü desteği veren, en kıymetlilerim annem Elif GÜRDİL, babam Sabri GÜRDİL, kardeşlerim Serap İNCİ, Mehtap GÜRDİL ve Filiz KILIÇ'a ve eniştelerim Tuncay İNCİ ve Mehmet KILIÇ'a teşekkür ederim.

Bu çalışmayı yiğenlerim Semih İNCİ, Elif Gönül İNCİ, Gökçe EĞERCİ, Sabri EĞERCİ ve Bahtiyar Çınar KILIÇ'a atfediyorum.

Hatice GÜRDİL

Kasım, 2023

İÇİNDEKİLER

Sayfa

ETİK İLKELERE UYGUNLUK BİLDİRİMİ ... iii
ÖZET ... iv
ABSTRACT ... vi
ÖNSÖZ ... viii
İÇİNDEKİLER ... ix
TABLOLAR DİZİNİ ... xiii
ŞEKİLLER DİZİNİ ... xv
KISALTMALAR ... xvi
BÖLÜM 1 ... 1
GİRİŞ ... 1
Problem ... 1
Amaç ... 8
Önem ... 9
Sınırlılıklar ... 10
BÖLÜM 2 ... 11
KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR ... 11
Ek Madde Tepki Kuramı ... 11
Çok Boyutlu Madde Tepki Kuramı ... 12
ÇBMTK'da Boyutluluk ... 13
Basit Yapı (Çoklu Tek Boyutlu Model) - Maddeler Arası. ... 13
Karmaşık Yapı - Maddeler İçi. ... 14
Karışık Model: Basit ve Karmaşık Yapı. ... 15
Çok Boyutluluk İçin Genişletilmiş Lojistik Modeller ... 18
Telafisel Olan Modeller için Genişletilmiş Tek Boyutlu Modeller ... 18
Çok Boyutluluk için Genişletilmiş Rasch Modeli. ... 19
Çok Boyutluluk için Genişletilmiş 2PLM ... 19
Çok Boyutluluk için Genişletilmiş 3PLM ... 20
Çok Boyutluluk İçin Genişletilmiş Normal - Ogive Modeller. ... 21
Çok Boyutlu Kısmi Modeli (ÇBKPM) ve Çok Boyutlu Genelleştirilmiş Kısmi Modeli (ÇBGKPM). ... 22
Çok Boyutlu Aşamalı Tepki Modeli (ÇBATM). ... 22
Telafisel Olmayan Modeller için Genişletilmiş Tek Boyutlu Modeller ... 23
Çok Boyutlu Madde Tepki Kuramı Modelinin Parametreleri ... 24
Madde Karakteristik Yüzeyi (MKY) ... 24
Uzunluk - Çok Boyutlu Madde Ayırıcılığı (ÇBMA). ... 25

Orijine Olan Uzaklık - Çok Boyutlu Madde Güçlüğü (ÇBMG). 26
Yön (αi). 27
Bilgi Fonksiyonu 27
Test Fonksiyonu 28
Çok Boyutlu Madde Tepki Kuramı Modelinin Parametre Kestirim Yöntemleri 28
Madde Parametre Kestirimi 28
Marjinal Maksimum Likelihood Tekniği (MML). 29
Bock Aitkin Expectation - Maximization (BA-EM). 29
Bayesyan Kestirim Teknikler 30
Yetenek Parametresi Kestirimleri 30
Maksimum Olabilirlik (ML). 30
Maksimum Sonsal (MAP_Maximum a Posteriori). 31
Beklenilen Sonsal (EAP_ Expected a posteriori) Yöntemler. 31
Değişen Madde Fonksiyonu 31
Lojistik Regresyon (LR) Yöntemi 34
SIBTEST 34
Lord'un χ^2 35
Çok Boyutlu Değişen Madde Fonksiyonu 35
Çok Boyutlu SIBTEST 36
Çok Boyutlu Veriler için MIMIC 36
Çok Boyutlu Veriler için MTK - LR 37
Çok Boyutlu Veriler için Lojistik Regresyon 37
İlgili Araştırmalar 38
Çok Boyutlu Madde Tepki Modeli ile İlgili Araştırmalar 38
Değişen Madde Fonksiyon ile İlgili Araştırmalar 43
ÇBMTK ve Değişen Madde Fonksiyonun Birlikte Ele Alındığı İlgili Araştırmalar . 44
İlgili Çalışmaların Genel Değerlendirmesi 46
BÖLÜM 3 49
YÖNTEM 49
Araştırmanın Modeli 49
Simülasyon Deseni 49
Sabit Tutulan Koşullar 50
Örneklem Büyüklüğü. 50
Tepki Kategori Sayısı 51
Toplam Madde Sayısı. 51
Madde Ayırıcılığı. 51
Madde Güçlüğü. 52
Birey Yetenek Parametreleri. 52
Boyutlar Arası Korelâsyon. 52

Veri Üretiminde Kullanılan Teknik. 52
Veri Üretiminde Kullanılan Model. 53
Değişimleme Yapılan Koşullar 53
Boyut Sayısı. 53
DMF Gösteren Madde Oranı. 53
DMF Türü. 54
DMF Büyüklüğü. 54
Verilerin Üretimi 54
DMF Gösteren Verilerin Oluşturulması 55
Verilere DMF Yüklenmesi. 55
100 Tekrarlı Veri Setlerinin Oluşturulması. 56
Simülasyona Dayalı Verilerin Geçerlik İncelemeleri 56
Veri Setlerine ait Model-Veri Uyumu İncelemeleri 56
Boyutluluk Kapsamında Yapılan İncelemeler 57
DMF Türlerinin Kontrol Edilmesi 61
DMF Büyüklüklerinin Kontrol Edilmesi 61
Verilerin Analizi 62
Gerçek Veri Çalışması 64
Değerlendirme Ölçütleri 69
Ana Etkiler ile Etkileşimler için Faktöriyel ANOVA. 71
BÖLÜM 4 73
BULGULAR VE YORUMLAR 73
Tek ve Çok Boyutlu Kestirimlere Yönelik DMF'li Madde Yüzdeleri, Boyut Sayısı, DMF Büyüklüğü ve DMF Türüne Göre Tip I Hata ve İstatistiksel Güç Oranları Ortalamalarının İncelenmesi 73
TB DMF Türü için; Tek Boyutlu ve Çok Boyutlu Kestirimlerinin, DMF Belirleme Tekniği Boyut Sayısı, DMF Yüzdesi, DMF Büyüklüğü Koşullarındaki Tip I Hata Oranlarının İncelenmesi 78
TB DMF Türü için; Tek Boyutlu ve Çok Boyutlu Kestirimlerinin, DMF Belirleme Tekniği, Boyut Sayısı, DMF Yüzdesi, DMF Büyüklüğü Koşullarındaki İstatistiksel Güç Oranlarının İncelenmesi 89
TBO DMF Türü için; Tek Boyutlu ve Çok Boyutlu Kestirimlerinin, DMF Belirleme Tekniği, Boyut Sayısı, DMF Yüzdesi, DMF Büyüklüğü Koşullarındaki Tip I Hata Oranlarının İncelenmesi 97
TBO DMF Türü için; Tek Boyutlu ve Çok Boyutlu Kestirimlerin, DMF Belirleme Tekniği, Boyut Sayısı, DMF Yüzdesi, DMF Büyüklüğü Koşullarındaki İstatistiksel Güç Oranlarının İncelenmesi 106
Tek ve Çok Boyutlu Madde Tepki Kuramı Yetenek Parametre Kestirimlerine Ait Uyum İstatistiklerinin Değişiminin İncelenmesi 116

Gerçek Veri Setinde Tek Boyutlu ve Çok Boyutlu Kestirimlerin Lojistik Regresyon, SIBTEST ve Lord'un χ^2 Tekniklerinin Tip I Hata ve İstatistiksel Güç Oranlarının İncelenmesi 121
BÖLÜM 5 123
SONUÇ VE ÖNERİLER 123
Sonuçlar 123
Tek Boyutlu ve Çok Boyutlu Kestirimlere Yönelik Yapılan DMF Analizlerine Ait Tip I hata ve İstatistiksel Güç Oranlarının İncelenmesi 123
Tek ve Çok Boyutlu Kestirimlere Yönelik Farklı Koşullar Altında Tip I Hata ve İstatistiksel Güç Oranlarının İncelenmesi 124
Tek ve Çok Boyutlu Kestirimlere Yönelik Yetenek Parametre Kestirimlerine Ait Uyum İstatistiklerinin İncelenmesi 125
Öneriler 126
Uygulayıcılara Öneriler 126
Araştırmacılara Öneriler 126
KAYNAKLAR 127
EKLER 145
EK 1. Simülasyon Koşullarına Yönelik Veri Üretimi 146
EK 2. Çok Boyutlu Kestirimden Elde Edilen Birinci Yetenek Parametresi Değerleri 147
EK 3. LR, SIBTEST ve Lord'un $\chi 2$ Yöntemleri DMF İncelemesi Sonuçları 148
BENZERLİK BİLDİRİMİ 149
ÖZGEÇMİŞ 150

TABLOLAR DİZİNİ

Tablo **Sayfa**

Tablo 1. Simülasyon Deseni 1 50
Tablo 2. Simülasyon Deseninde Belirlenen Boyut Sayısına Uyum Sonuçları 58
Tablo 3. Doğru ve Yanlış Olarak DMF'li BelirlenenMadde Sayıları ve Oranları 60
Tablo 4. Simülasyon Deseninde Belirlenen DMF Büyüklüklerinin İncelenmesi 61
Tablo 5. Modellerin Karşılaştırma Sonuçları 67
Tablo 6. Bradley'in Tip I Hata Oranlarına İlişkin Sınıflandırması 69
Tablo 7. DMF'li Madde Yüzdeleri, Boyut Sayısı, DMF Büyüklüğü ve DMF Türüne Göre Tip I Hata ve İstatistiksel Güç Oranları Ortalamaları 74
Tablo 8. DMF'li Madde Yüzdeleri, Boyut Sayısı, DMF Büyüklüğü ve DMF Türüne Göre Tip I Hata ve Güç Oranları Ortalamaları 76
Tablo 9. TB DMF Türü için Faktöriyel ANOVA Sonucu (Tip I hata Oranları) 78
Tablo 10. TB DMF Türü İçin Tip I Hata Oranlarının Kestirim Yöntemine Göre İkili Karşılaştırmaları 80
Tablo 11. TB DMF Türü İçin Tip I Hata Oranlarının Metotlara Göre İkili Karşılaştırmaları 81
Tablo 12. TB DMF Türü İçin Tip I Hata Oranlarının DMF Yüzdelerine Göre İkili Karşılaştırmaları 82
Tablo 13. TB DMF Türü İçin Tip I Hata Oranlarının 'Kestirim' x 'Teknik' İkili Etkileşime Göre İkili Karşılaştırmaları 83
Tablo 14. TB DMF Türü İçin Tip I Hata Oranlarının 'Kestirim' x 'DMF Yüzdesi' İkili Etkileşime Göre İkili Karşılaştırmaları 85
Tablo 15. TB DMF Türü İçin Tip I Hata Oranlarının 'DMF Büyüklüğü' İkili Etkileşime Göre İkili Karşılaştırmaları 87
Tablo 16. TB DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları) 89
Tablo 17. TB DMF Türü İçin İstatistiksel Güç Oranlarının Kestirim Yöntemine Göre İkili Karşılaştırmaları 90
Tablo 18. TB DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'Teknik' İkili Etkileşime Göre İkili Karşılaştırmaları 91

Tablo 19. TB DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'Boyut Sayısı' İkili Etkileşime Göre İkili Karşılaştırmaları 93
Tablo 20. TB DMF Türü İçin Tip I Hata Oranlarının 'DMF Büyüklüğü' İkili Etkileşime Göre İkili Karşılaştırmaları 95
Tablo 21. TBO DMF Türü için Faktöriyel ANOVA Sonucu (Tip I hata Oranları) 97
Tablo 22. TBO DMF Türü İçin Tip I Hata Oranlarının Kestirim Yöntemine Göre İkili Karşılaştırmaları 99
Tablo 23. TBO DMF Türü İçin Tip I Hata Oranlarının DMF Yüzdelerine Göre İkili Karşılaştırmaları 100
Tablo 24. TBO DMF Türü İçin Tip I Hata Oranlarının 'Kestirim' x 'Teknik' İkili Etkileşime Göre İkili Karşılaştırmaları 101
Tablo 25. TBO DMF Türü İçin Tip I Hata Oranlarının 'Kestirim' x 'DMF Yüzdesi' İkili Etkileşime Göre İkili Karşılaştırmaları 103
Tablo 26. TBO DMF Türü İçin Tip I Hata Oranlarının 'DMF Büyüklüğü' İkili Etkileşime Göre İkili Karşılaştırmaları 105
Tablo 27. TBO DMF Türü için Faktöriyel ANOVA Sonucu (İstatistiksel Güç Oranları) 107
Tablo 28. TBO DMF Türü İçin İstatistiksel Güç Oranlarının Kestirim Yöntemine Göre İkili Karşılaştırmaları 108
Tablo 29. TBO DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'Teknik' İkili Etkileşime Göre İkili Karşılaştırmaları 109
Tablo 30. TBO DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'DMF Yüzdesi' İkili Etkileşime Göre İkili Karşılaştırmaları 111
Tablo 31. TBO DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'Boyut Sayısı' İkili Etkileşime Göre İkili Karşılaştırmaları 113
Tablo 32. TBO DMF Türü İçin Tip I Hata Oranlarının 'DMF Büyüklüğü' İkili Etkileşime Göre İkili Karşılaştırmaları 114
Tablo 33. Tüm Koşular için Tek Boyutlu Kestirimlere Yönelik Uyum İstatistikleri Ortalamaları 116
Tablo 34. Tüm Koşular için Çok Boyutlu Kestirimlere Yönelik Uyum İstatistikleri Ortalamaları 119

ŞEKİLLER DİZİNİ

Şekil **Sayfa**

Şekil 1. Basit Yapılı Çok Boyutlu Model 14
Şekil 2. Karmaşık Yapılı Çok Boyutlu Model Örneği 14
Şekil 3. Karışık Model: Basit ve Karmaşık Yapı Birlikte Örneği 15
Şekil 4. Madde Karakteristik Yüzeyi Örneği 25
Şekil 5. Tek Biçimli DMF Örneği 32
Şekil 6. Tek Biçimli Olmayan Ve Çapraz Tek Biçimli Olmayan DMF Örneği 33
Şekil 7. Paralel Analiz Yamaç-Birikinti Grafiği 66
Şekil 8. Kestirim Ana Etkisi için Tip I Hata Oranları 80
Şekil 9. Teknik Ana Etkisi için Tip I Hata Oranları 81
Şekil 10. DMF Yüzdesi Ana Etkisi için Tip I Hata Oranları 82
Şekil 11. Kestirim ile Teknik Etkileşimi İçin Tip I Hata Oranları 84
Şekil 12. Kestirim ile DMF Yüzdesi Etkileşimi İçin Tip I Hata Oranları 86
Şekil 13. Kestirim ile DMF Büyüklüğü Etkileşimi İçin Tip I Hata Oranları 88
Şekil 14. Kestirim Ana Etkisi için Güç İstatistiksel Güç Oranları 91
Şekil 15. Kestirim ile Teknik Etkileşimi için İstatistiksel Güç Oranları 93
Şekil 16. Kestirim ile Boyut Sayısı Etkileşimi için İstatistiksel Güç Oranları 94
Şekil 17. Kestirim ile DMF Büyüklüğü Etkileşimi için İstatistiksel Güç Oranları 96
Şekil 18. Kestirim Ana Etkisi için Tip I Hata Oranları 99
Şekil 19. DMF Yüzdesi Ana Etkisi için Tip I Hata Oranları 100
Şekil 20. Kestirim ile Teknik Etkileşimi İçin Tip I Hata Oranları 102
Şekil 21. Kestirim ile DMF Yüzdesi Etkileşimi İçin Tip I Hata Oranları 104
Şekil 22. Kestirim ile DMF Büyüklüğü Etkileşimi İçin Tip I Hata Oranları 106
Şekil 23. Kestirim Ana Etkisi için İstatistiksel Güç Oranları 109
Şekil 24. Kestirim ile Teknik Etkileşimi için İstatistiksel Güç Oranları 110
Şekil 25. Kestirim ile DMF Yüzdesi Etkileşimi için İstatistiksel Güç Oranları 112
Şekil 26. Kestirim ile Boyut Sayısı Etkileşimi için İstatistiksel Güç Oranları 114
Şekil 27. Kestirim ile DMF Büyüklüğü Etkileşimi için İstatistiksel Güç Oranları 115

KISALTMALAR

MTK	Madde Tepki Kuramı
TBMTK	Tek Boyutlu Madde Tepki Kuramı
ÇBMTK	Çok Boyutlu Madde Tepki Kuramı
1 PLM	1 Parametreli Lojistik Model
2 PLM	2 Parametreli Lojistik Model
3 PLM	3 Parametreli Lojistik Model
DMF	Değişen Madde Fonksiyonu
TB	Tek Biçimli
TBO	Tek Biçimli Olmayan

BÖLÜM 1

GİRİŞ

Bu bölümde araştırmanın problemine, alt problemlerine, amacına, önemine ve sınırlılıklarına yer verilmiştir.

Problem

Testler belirli bir amaç doğrultusunda hazırlanmakta ve testlerden elde edilen puanlara göre birtakım çıkarımlar yapılmaktadır. Çok önemli kararların alınabildiği bu çıkarımların farklı test katılımcısı gruplarında geçerli olması için testte adaletin sağlanması gerekmektedir. Educational Testing Service-ETS (2014) tarafından "test puanlarına göre yapılan çıkarımların farklı test katılımcısı grupları için geçerli olma derecesi" olarak tanımlanan test için adalet (fairness) kavramı, ölçme açısından seçme ve sınıflandırma kararları için bireylerin değerlendirilmesinde kullanılan test ya da madde puanlarına ilişkin anlamların aralığı şeklinde açıklanabilir. Adalet kavramı ilkelerinden birini de test ve madde düzeyinde yanlılık oluşturmaktadır (AERA, APA ve NCME, 2014). Testte adaletsizlik kaynağı olarak eğitim imkânları, öğretmen deneyimi ve öğretmenin yeterlik durumu gösterilebilmekle birlikte en yaygın kaynak olarak test maddelerinin dezavantajlı grupların sınırlı maruz kaldığı dil ve kapsama ilişkin bilgiler içermesi olasılığı olarak gösterilmektedir. Testi alan belirli alt gruplara yönelik bu adaletsizlik yanlılık olarak isimlendirilir (Penfield ve Camilli, 2007).

DMF çalışmaları öncelikle test düzeyinde ele alınmakla birlikte test düzeyinde DMF incelenmesinin ve istatistiksel olarak modellenmesinin zor olması nedeniyle madde düzeyinde DMF incelemesi yapılmaya başlanmıştır (Dorans ve Holland, 1993). 1980-1990 arası Değişen Madde Fonksiyonu (DMF) kavramına ve ölçümüne yönelik ilgi artmıştır. DMF, 2014 yılında APA standartlarında "aynı yetenek seviyesinde olan farklı test katılımcı gruplarının farklı doğru yanıt frekanslarına sahip olma derecesinin istatistiksel bir göstergesi" şeklinde tanımlanmıştır (AERA, APA ve NCME, 2014). Bu nedenle, DMF çalışmaları ile aynı yetenek düzeyinde yer alan farklı gruplardaki adayların madde yanıtlardaki farklılık değerlendirilmektedir (Dorans ve Holland, 1993). DMF, aynı

yetenek düzeyinde ancak farklı gruplardaki bireylerin madde performanslarının aynı olmadığını ortaya koyar (Steinberg ve Thissen, 2006). Adaletsiz bir değerlendirme, bireylerin veya grupların bazılarının ölçülen yapı üzerindeki duruşunu sistematik olarak yanlış temsil eder. Buradan anlaşıldığı üzere adillik, geçerlik kavramı ile yakından ilişkilidir.

Geçerlik, test puanlarından kaynaklanan eylemler ve sonuçların, değerlendirme yaklaşımlarının teorik temelleri ve deneysel kanıtlarla birlikte uygunluğunu ve yeterliliğini destekleyen birleşik bir değerlendirmedir. Test puanlarının geçerliğini değerlendirmek için çoklu kanıt bulmak amacıyla farklı geçerlik kanıtları toplanmalıdır. Bu kanıtlar test içeriği, yanıt süreçleri, içyapı, diğer değişkenlerle ilişkiler ve testin sonuçları ile ilgili kanıtlar şeklinde beş başlık altında verilmiştir (AERA, APA ve NCME, 2014). DMF incelemeleri de içyapı kapsamındaki geçerlik kanıtlarından biridir. DMF çalışmalarında aynı yetenek düzeyindeki bireylerin cinsiyet, sosyo-ekonomik farklılıklar, bilgiye ulaşma imkânları, bireyin ilgi alanları, yaşanılan coğrafya, etnik/ırk, dil, yaş gibi DMF yaratabilecek olası kaynaklara göre maddenin performansının benzer çalışıp çalışmadığı incelenmekte bu kapsamda yapı geçerliğine hitap etmektedir.

Bir testin ölçmeyi amaçladığı yapı alanının önemli yönlerini yakalayamama derecesi eksik yapılandırmaya işaret eder. Test puanlarının anlamını bozan yabancı faktörler ilgisiz varyans olarak adlandırılmaktadır. DMF çalışmalarıyla grup farklılıklarına bağlı olarak yapı ile ilgisiz varyans, istatistiksel olarak tespit edilebilmekte; testten elde edilen puanların farklı gruplarda aynı yapıyı yansıtması sağlanmaktadır (Camilli ve Shepard, 1994; Ellis ve Raju, 2003; Sireci ve Allalouf, 2003).

Yanlılık, bir gruba diğerine göre daha fazla zarar verme durummunu ortaya koymaktadır (Shepard vd., 1985). Bu tanım performanstaki farklılığın basit bir gözleminden ziyade bir değerlendirme duygusu içerir ve bir grup üzerindeki adaletsiz etkisine atıfta bulunur (Angoff, 1973). Bir geçerlik sorunu olan yanlılık, farklı alt gruplardaki bireylerin test puanlarının, bulundukları gruba bağlı olarak sistematik hata içermesidir. Bir testin tarafsızlığını, geçerliğini ve buna bağlı olarak güvenirliğini olumsuz yönde etkileyen yanlılığın belirlenmesi, DMF incelemeleri ile başlamaktadır.

DMF incelemelerinde, DMF türlerindeki farklılıkları dikkate alan farklı DMF yöntemleri mevcuttur. Bu durumda DMF analizlerinin yürütülmesi ve yorumlanması sürecinde DMF türleri arasındaki ayrım önemli hale gelmektedir (Penfield ve Camilli, 2007). Bununla birlikte hem tek biçimli hem de tek biçimli olmayan DMF, maddeye verilen yanıtın sadece yeteneğe değil aynı zamanda grup üyeliği ile ilişkili ikinci bir

faktöre bağlılığına işaret eder. Bu nedenle testin geçerliğini tehdit eder. DMF'nin varlığı madde performansını etkileyen ikinci bir boyutun varlığı olarak çok boyutluluk açısından da modellenebilir.

Tek boyutlu bir testte, test ile ölçülmek istenen bir yetenek vardır. Bu yeteneğe aynı düzeyde sahip olan bireylerin doğru yanıtlama olasılıklarının benzer olması beklenir. Bununla birlikte, test maddeleri doğru olarak geliştirilmediğinde teste ölçülmek istenen dışında farklı boyutlar karışabilmektedir. Bu durumda testle ölçülmek istenen yetenek bakımından aynı düzeydeki bireylerin doğru yanıtlama olasılıkları da farklılaşabilmektedir. Diğer bir ifadeyle aynı yetenek düzeyindeki bireylerin doğru yanıtlama olasılıklarına ait madde fonksiyonları farklılaşmaktadır. Bu maddeler için madde fonksiyonları değişeceğinden ilgili maddeler DMF gösteren maddeler olarak düşünülebilmektedir.

Testle ölçülmek istenen yetenek birinci boyutu oluşturmaktadır. Farklı yetenekler ise bu boyut içerisinde yer alamayacağı için test içerisinde farklı boyutlar olarak konumlanır. Maddelerin doğru yanıtlanabilmesi için bireylerin testle ölçülmek istenen birinci boyuttaki yetenek haricinde, diğer boyutları oluşturan farklı yeteneklere daha sahip olmaları gerekebilmektedir (Angoff, 1973). Tek boyutlu testlerde bazı maddeler DMF içerebilir. Bu durum testlerin tek boyutluluğun bozulacağı anlamına gelmez; DMF her zaman tek boyutluluk ihlali değildir. Diğer bir ifadeyle tek boyutluluk sağlansa bile madde DMF gösterebilir; tek boyutlu maddelerde DMF olmayacağı söylenemez. Bununla birlikte, testin DMF gösteren maddeler içermesi, bazı durumlarda tek boyutluluk varsayımını ihlal edebilir. Araştırmacılar tarafından yirmi yıldan uzun bir süredir DMF'nin önde gelen nedeni olarak bazı test maddelerinin çok boyutlu olması gösterilmektedir (Ackerman, 1992; Roussos ve Stout, 1996; Russell, 2005; Shealy ve Stout, 1993). Bu durum DMF'den ziyade boyutluluk meselesidir. Çok boyutlu bir testte ikincil yetenek, ölçülmesi amaçlanan birincil yetenekle ilgisiz olduğunda DMF istenmeyen boyut olmaktadır. Bu durumda tek boyutluluk varsayımı ihlal olduğu için Tek Boyutlu Madde Tepki Kuramı (TBMTK) modellerinin kullanılması uygun olmamaktadır. Kullanılması halinde birincil yeteneğin veya özelliğin doğru temsil edilememesine; madde ve parametre kestirimlerinin yanlış yapılmasına ve hatalı test puanlarının elde edilmesine neden olur. Bu durum özellikle yüksek riskli sınavlar olarak ifade edilen, önemli kararların alındığı sınavlarda ya da bireylere özel öğretimsel müdahelelerin sağlanacağı yeterlik testlerinde, bireyler hakkında hatalı kararlar alınması sorununu ortaya çıkarır.

Tek boyutluluk varsayımı ihlal edildiğinde testi alanlar tek bir özellik üzerinde sıralanamamaktadır. Çünkü çok boyutlu örtük uzay ile cevaplayıcının tek bir yeteneği birebir eşleştirilememektedir. Bu gibi durumlarda TBMTK modelleri yerine bunların çok boyutlu veri setleri için genişletilmiş hali olan Çok Boyutlu Madde Tepki Kuramı (ÇBMTK) modellerinin kullanılması uygundur (Embertson ve Reise, 2000). ÇBMTK'da da TBMTK'da olduğu gibi test maddesi esastır.

ÇBMTK, iki ya da daha fazla gizil değişken ile bireyin belli bir test maddesine doğru yanıt verme olasılığı arasındaki ilişkiyi matematiksel bir modelle açıklayan modellerdir. Diğer bir ifadeyle, ÇBMTK'nın temelinde tek bir maddede çoklu bilişsel yeteneklerin olması yatmaktadır. TBMTK'dan farklı olarak her bir boyut için birey yeteneği farklı bir değer aldığından, bireye ait yetenekler vektör halinde tanımlanmaktadır. ÇBMTK'da önemli bir aşama boyut belirlemektir (Reckase, 2009). Ölçmede boyut, bireylerin sıralanabildiği ölçek olarak tanımlanmaktadır. Jacoby'e (1991) göre bir madde kümesinin boyut olarak ele alınabilmesi için, bu kümenin bireyler arasında değişkenlik kaynağı oluşturacak şekilde bir niteliğe sahip olması gerekmektedir. Bununla birlikte maddeler ortak bir karakteristik özelliğe sahip olmalıdır.

Çok boyutlu testlerde ikinci boyut istenmeyen boyut, diğer bir ifadeyle DMF olarak belirlenebilmektedir. Buna ek olarak istenmeyen boyut sayısı ikiden fazla olabilmektedir. Bu boyutlar birinci (temel) boyuttan sonraki boyutları temsil etmektedir. Yapı iki boyutlu ise birinci boyut kalibre edildikten sonra birinci boyuta göre ikinci boyut farklılaşmasına bakıldığı için birinci boyutta DMF görülmez (Mair, 2018). Bahsedilen bu duruma örnek olarak, anadilinden farklı bir dilde matematik sınavına giren bir öğrenci ele alınabilir. Bu öğrenci, dil anlama ihtiyacı yüksek olan bir matematik maddesine yanıt vermekte zorlanabilir. Çünkü burada maddeye doğru yanıt verme olasılığı üzerinde sadece matematik becerisi değil, dil anlama becerisi de ikinci bir boyut olarak etkili olmaktadır. Bu durumda bu öğrenci ile aynı yetenek düzeyine sahip anadilinde sınav olan bir öğrencinin bu soruyu doğru yanıtlama olasılığının daha yüksek olması olasıdır. Bu durumda dil anlama becerisi ikincil bir boyut olarak test puanına karışabilmektedir. Testle ölçülmesi amaçlanmayan bu ikincil boyutun dikkate alınmayarak testin tek boyutlu olarak düşünülmesi ve tek boyutlu kestirimler üzerinden ana dilinde sınav olmayan öğrencinin ilk boyuttaki yeteneğinin saptanmaya çalışılması, bu yeteneğin yanlış tahmin edilmesine neden olabilir. İkincil yetenek olarak "dil anlama becerisi" de doğru yanıtlama olasılığı üzerinde etkili olacağından; asıl ölçülmek istenen "matematik becerisi" yeteneği daha düşük tahmin edilebilir. Diğer bir ifade ile testle ölçülmek istenen yetenek ile ilgisi

olmayan ikincil yetenekler, bir grup için yanlılık oluşturabilir. Bu sınavın bilgisayar üzerinde yapıldığı ve teknoloji kullanma becerisi ihtiyacı yüksek olduğu düşünüldüğünde, teknoloji becerisi de üçüncü bir boyut olarak test puanları üzerinde etkili olmaktadır. Bu bağlamda tek boyutluluk varsayımı ihlal olduğu için TBMTK modellerinin kullanılması uygun olmaz. Kullanılması halinde birincil yeteneğin veya özelliğin doğru temsil edilememesine; madde ve parametre kestirimlerinin yanlış yapılmasına neden olmaktadır (Mair, 2018).

Alanyazında DMF belirlenmesinde hem TBMTK hem de ÇBMTK'ya dayalı yöntemlerin kullanıldığı çalışmalar (Atar, 2007; Bolt ve Stout, 1996; Huang, 2010; Liaw, 2015) yer almaktadır. Bu çalışmalarda DMF incelemelerinde birden fazla DMF tekniği kullanılarak maddenin DMF içerdiğine karrar verilmesi önerilmektedir. Bilindiği üzere DMF istatistiksel bir çıkarımdır. Sadece istatistiksel yanlılığa bakılarak maddenin yanlı olduğuna karar verilemez. DMF çok boyutluluğa göre ikinci faktörün varlığı olup asıl önemli olan bu faktörün test yapısıyla kasıtlı olarak ilişkili olup olmadığının belirlenmesi, bu farklılığın gerçek bir farklılık mı yoksa maddenin içeriğinden kaynaklı bir farklılık mı olduğuna karar verilmesidir.

DMF gösteren maddelerin yanlı olarak belirlenmesi için bu maddelerin uzmanlar tarafından incelendiği panel çalışmaları yapılmalıdır (Karami, 2012). Uzman görüşüne bağlı olarak, DMF'nin varlığı istenmeyen madde içeriğine ya da diğer bazı istenmeyen madde özelliklerine atfedilebiliyorsa maddenin yanlı olduğu söylenebilmektedir. Bu durum maddenin yanlı olması için DMF'nin var olmasının yeterli olmadığını göstermektedir (Penfield ve Camilli, 2007). Yanlı olduğuna karar verilen maddelerin uzmanlar tarafından yeniden değerlendirilmesi, gözden geçirilip düzeltilmesi ya da testten çıkarılması önerilmektedir (Penfield ve Camilli, 2007; Osterlind ve Everson, 2009). DMF düzeltmelerinde önerilen bir yöntem olarak arındırma (item purification) da DMF gösteren maddenin çıkarılarak geri kalan maddeler üzerinden yetenek kestirimi önerilmekte (Mair, 2018), ancak bu durum kapsam daralmasına yol açabilmektedir. Özellikle madde havuzu küçükse ve DMF içeren madde sayısı çok ise bu durum ideal bir çözüm olmayacaktır. Magis ve De Boeck (2011) da yaptıkları simülasyon çalışmasında DMF gösteren madde sayısı ve DMF miktarı arttıkça Tip I hata oranlarının önemli ölçüde büyüdüğünü, özellikle büyük örneklemlerde bu büyümenin çok fazla olduğunu ve arındırma işleminin bile bu büyümeyi kontrol altına alamadığını belirtmişlerdir. Bu bağlamda, genel vurgu test geliştirme aşamasında madde yanlılığın dikkate alınması ve yanlılık içermeyecek maddeler yazılmasıdır. Önerilen bir diğer durum ise ÇBDMF

modellerinin kullanılmasıdır. Buna ek olarak; ikincil boyutta düşük ve orta düzeyde a parametresinden kaynaklı DMF gösteren maddeler içeren iki boyutlu bir testte ÇBMTK modelleri kullanılarak birincil yetenek tahmininin kalibre edebilmesini (Liaw, 2015) öneren bir çalışma da mevcuttur. Buna ek olarak yapılan yüksek riskli sınavlar sonucu gerçekleştirilen DMF analizleriyle DMF gösteren maddelerin saptandığı, izleyen uzman görüşü çalışmalarıyla da bazı maddelerin yanlı olduğuna karar verildiği çalışmalar (Akalın ve Koç, 2014; Arslan, 2020; Satıcı ve Özer Özkan, 2017; Özyıldırım Gümüş, 2018; Yıldırım, 2012) mevcuttur. Bu çalışmalarda maddelerin yanlı olduğu tespit edilmesine rağmen sınavı alanlar bu maddelerin de dahil olduğu test puanı üzerinden puan almış ve sıralanmıştır. Maddelerin düzeltilmesi ise ancak ileri dönem sınavlar için bir düzeltme sağlayabilir. Mevcut puanlar üzerinde yapılan yanlışlar için herhangi bir düzeltme sağlamaz. DMF'nin sistematik bir hata olduğu ve geçerliği düşürdüğü bilinmektedir. Bu durumda testi alanlar bile bile geçerliği düşük bir şekilde sıralanabilmekte ve haklarında yanlı kararlar verilebilmektedir. Diğer bir ifadeyle test sonuçlarına sistematik hata karışmaktadır.

Test sonuçlarına sabit, sistematik ve tesadüfi olmak üzere üç tür hata karışabilmektedir. Sabit ve sistematik hatalar belirlenebildiği için bu hatalar düzeltilerek düzeltilmiş puan elde edilebilmektedir. DMF de sistematik bir hata olmasına rağmen puanların düzeltilmesi yoluna gidilmediği görülmektedir. Bu nedenle bu çalışmanın problemini ÇBMTK kullanılarak DMF'nin düzeltilip düzeltilemeyeceği oluşturmaktadır.

Alanyazında yapılan incelemeler sonucunda, son yıllarda ÇBMTK (Ackerman; 1989; Ackerman; 2005; Bolt ve Lall, 2003; Bobcock, 2009; De la Torre ve Patz, 2005; Kirisci, Hsu ve Yu, 2001; Kreiter, 1993; Levy, 2006; Luecht ve Miller, 1992; Spray vd., 1990; Soysal, 2017) yönteminin sıklıkla inclelendiği görülmektedir. Bununla birlikte TBMTK ve ÇBMTK'nın birlikte ele alındığı (Ansley ve Forsyth, 1985; Reckase vd., 1988; Zeng, 1989), parametrik ve parametrik olmayan yöntemlerin (Koğar, 2015; Şengül ve Tavşancıl, 2015), tek boyutlu ve çok boyutlu madde tepki kuramı modellerinin karşılaştırıldığı (Çakıcı Eser, 2015; Koğar, 2015; Köse, 2010; Liaw, 2015; Seungho Yang, 2007; Sheng, 2005; Spencer, 2004; Sünbül, 2011; Özer-Özkan, 2012) çalışmalar da yer almaktadır. Yapılan parametre kestirim çalışmalarında birey ve madde parametrelerinin kestirim kararlığı; örneklem büyüklüğü, test uzunluğu, madde oranı, boyut sayısı, boyutlar arası korelâsyon, yetenek dağılımı gibi özelliklerin değişimlenmesi ile belirlenmeye çalışılmıştır.

ÇBMTK'nın DMF belirleme çalışmalarında kullanıldığı çalışmalar incelendiğinde; DMF'nin nedenlerinin araştırıldığı çalışmaların (Ackerman, 1992; Roussos ve Stout, 1996; Russell, 2005; Shealy ve Stout, 1993); MIMIC, MTK-LR ve Lojistik Regresyon'un tek boyutlu formlarını birbirleriyle ve diğer DMF tespit yöntemleriyle karşılaştırdığı çalışmaların (Atar ve Kamata, 2011; Finch, 2005; Kan ve Bulut, 2014; Woods ve Grimm, 2011;) ve çok boyutlu testlerde DMF inceleme yöntemlerinden MIMIC, MTK-LR ve Lojistik Regresyon yaklaşımlarının karşılaştırıldığı bir çalışmanın (Bulut ve Suh, 2017) yer aldığı görülmektedir. Mevcut çalışma koşullarında boyut sayısı, DMF büyüklüğü ve DMF türü koşulları incelenmekte olup alan yazında yapılan çalışmalarda bu koşulların incelenmediği görülmüştür. Alan yazında ele alınan konudaki çalışmalarda daha çok madde parametreleri arasındaki farklılıklar, DMF belirlenmesinde farklı p değerlerinin kullanılması vb. koşullar ele alınmıştır (Liaw, 2015). Simülasyona dayalı olarak genellikle iki boyutlu veri setleri elde edildiği ve elde edilen veri setlerinde Tip I hata ve istatistiksel güç oranlarının kıyaslandığı görülmektedir. Yapılan çalışmalar incelendiğinde DMF belirleme süreçlerinde testin boyut yapısı üzerinde duran, çok boyutlu testlerin tek boyutlu olarak ele alınmasının yanıltıcı sonuçlar ortaya çıkmasına neden olduğunu gösteren bir çalışma (Liaw, 2015) olmakla birlikte; DMF içeren maddeler nedeniyle tek boyutluluğu sağlanmayan testlerde ÇBMTK modelleri ile bireylerin birinci yetenek kestirimlerinin elde edilmesi ve buna yönelik Tip I hata ve istatistiksel güç oranlarındaki düzeltmenin incelendiği bir araştırmaya ise rastlanmamıştır.

Tek boyutlu testlerin DMF gösteren maddeler içermesi hem testin tek boyutluluğunu hem de geçerlik ve güvenirliğini riske atmaktadır. Testle ölçülmesi planlanan yetenek birinci boyutta yer alacağı için DMF çok boyutluluğa neden olacaktır Angoff (1973). DMF büyüklüğü ve DMF gösteren madde yüzdesinin artması durumunda test yanlılığın yükselme olasılığı vardır. Bu durum tek boyutlu testte geçerlik sorunu oluşturabilmektedir. DMF tek boyutlu bir test için bir risk olabilmekte; testte ikinci ve üçüncü boyutun karıştığını gösterebilmektedir. Bu çalışma kapsamında, yanlılık sorununun ortadan kaldırılmasında ikinci ve üçüncü boyutların ölçülmek istenilen yapıyla ilgisiz olduğu durumlarda çok boyutlu kestirimler yapmanın birinci yetenek kestiriminin doğruluğu (accuracy) üzerindeki etkisi incelenmiştir. DMF gösteren maddeler nedeniyle tek boyutluluğu sağlanamayan bir testte; ÇBMTK kestirimi ile ikinci ve üçüncü boyutlar da modele dâhil edilebildiğinden bir düzeltme sağlayabileceği düşünülmektedir. Bu bağlamda çalışma kapsamında düzeltilmiş puan, ÇBMTK kestirimi yapıldığında elde

edilen birinci yetenek puanları olarak tanımlanmaktadır. ÇBMTK modeli kullanılarak, birden fazla yetenek eş zamanlı olarak tahmin edilebilmektedir. Bu bağlamda ÇBMTK kestirimiyle; DMF gösteren maddeleri testten çıkarmak yerine; sistematik hatadan arındırılmış; düzeltilmiş yetenek puanlarına bağlı olarak yapılan DMF analizi yöntemlerinden hatası daha düşük, istatistiksel gücü daha yüksek sonuçlar elde etme düşüncesi vardır. Çok boyutlu modeller yapıları gereği Tip I hatayı kontrol altına almada tek boyutlu modellere göre daha etkilidir. Bu kapsamda bu çalışmada problem durumu olarak ÇBMTK kestirim sonuçlarının incelenmesi; hangi koşullarda en iyi kontrolün sağlandığının ve ne düzeyde kontrol altına alındığının incelenmesi ele alınmıştır.

Amaç

Bu çalışmada DMF gösteren maddeler nedeniyle tek boyutluluğun sağlanamadığı testlerde, Tek Boyutlu ve Çok Boyutlu MTK kestirtimlerinin DMF'yi kontrol altına almaya yönelik etkilerinin farklı koşullar altında incelenmesi amaçlanmıştır. Bu amaç doğrultusunda DMF belirleme teknikleri (LR, SIBTEST, Lord' $\chi 2$), boyut sayısı (2, 3), DMF yüzdesi (%10, %20, %30), DMF büyüklüğü (.50, .75) ve DMF türü (TB, TBO) faktörlerine yönelik koşullar değişimlenmiştir. Belirtilen DMF belirleme tekniklerinin performanslarının Tip I hata ve istatistiksel güç oranları karşılaştırılmıştır. TB ve TBO DMF türleri için ayrı olmak üzere, değişimlenen koşulların etkileri, hipotez testleri ile test edilmiştir. Bu kapsamda aşağıdaki sorulara yanıt aranmıştır:

1. TB DMF Türü için kestirim (Tek Boyutlu, Çok boyutlu), DMF belirleme teknikleri (LR, SIBTEST, Lord χ^2), boyut sayısı (2,3), DMF yüzdesi (%10, %20, %30) ve DMF büyüklüğü (.50, .75) koşullarındaki
   a) Tip I hata oranları,
   b) İstatistiksel güç oranları nasıl değişmektedir?
2. TBO DMF Türü için Kestirim (Tek Boyutlu, Çok boyutlu), DMF belirleme teknikleri (LR, SIBTEST, Lord χ^2), Boyut sayısı (2,3), DMF yüzdesi (%10, %20, %30) ve DMF büyüklüğü (.50, .75) koşullarındaki
   a) Tip I hata oranları,
   b) İstatistiksel güç oranları nasıl değişmektedir?
3. Tek ve Çok Boyutlu MTK yöntemine dayalı olarak elde edilen yetenek parametre kestirimlerine ait uyum istatistikleri nasıl değişmektedir?

4. DMF gösteren maddeler içeren bir gerçek veri setinde, TBMTK ve ÇBMTK modelleri ile elde edilen birinci yetenek parametre kestirimlerine yönelik veri setlerinde Lojistik Regresyon, SIBTEST ve Lord'un χ^2 tekniklerinin;
    a. Problem Tip I hata oranları nasıl değişmektedir?
    b. Güç oranları nasıl değişmektedir?

Önem

Test maddelerinden bazılarının bağlamının, testi alan odak gruba tanıdık gelmediği düşünüldüğünde; mantıksal olarak, ölçülen özelliğe odak grup referans grupla aynı düzeyde sahip olsa bile, referans gruba göre doğru yanıt verme olasılığı daha düşük olabilmekte ve maddede DMF gözlenebilmektedir. Bu şekilde DMF gösteren maddeler içeren bir testte, bu maddelerin dikkate alınmaması gözlemlenen puanların hatalı olmasına neden olabilmektedir. Bu durum, bir grubun gerçek yapısının ya da özelliklerinin doğru olarak belirlenememesine neden olur. Bu nedenle DMF gösteren olarak belirlenen maddelere uygun şekilde müdahale edilmesi çok önemlidir.

Alanyazında bu durumda tavsiye edilen; uzman görüşü ile madde içeriğinin bir grup için avantajlı olabileceği kararı verilmesi durumunda, DMF gösteren maddelerin yenilenmesi ya da kaldırılmasıdır. Oysaki madde yazım süreci çok zordur. Bununla birlikte bazı durumlarda test madde havuzu nispeten küçük olabilmekte ve / veya testin büyük bölümünde DMF gösteren maddeler olabilmektedir. Bu gibi durumlarda DMF gösteren maddelerin kaldırılması pratikte ideal bir çözüm olmayabilir. Bu nedenle, özellikle nispeten yüksek miktarda DMF içeren testlerde, bu çalışmada önerildiği gibi DMF gösteren maddelerin atılmadan kullanılması ve test puanını DMF etkilerinden arındırarak düzeltilmiş puan elde etme fikri bu çalışmanın önemini ortaya koymaktadır. Nitekim testle ölçülmek istenen yetenek birinci boyutta yer almakta, diğer yetenekler ikinci boyuta konumlanarak DMF'ye neden olmaktadır (Angoff, 1973). Bu kapsamda ÇBMTK modellerinin kullanılabileceği; ÇBMTK ile birincil boyut kestirilirken DMF etkisinin kontrol edilebileceği düşünülmektedir. Diğer bir ifadeyle, ikincil yetenek parametresinin etkisi dikkate alındığında, madde ve birey parametrelerinde yanlılık düzeltmesi sağlaması beklenmektedir. Elbette çok boyutlu modellerde Tip I hatanın daha iyi kontrol altına alınması beklenir. Bununla birlikte hangi koşullarda en iyi ve ne düzeyde kontrol altına alındığı bilinmemektedir. Bu çalışma ile hangi koşulların hangi durumlarda

daha etkili olduğuna dair çok önemli bilgiler edinileceği düşünülmektedir. Öyle ki alanyazın incelemelerinde de DMF etkisinin kontrol edilmesinde ÇBMTK modellerinin kullanılması durumunun incelenmediği görülmektedir. Bu bağlamda, bu çalışmada olduğu gibi değişkenlik gösterebilecek farklı koşullar altında elde edilecek sonuçlar konusunda bilgi edinilmesi önem kazanmaktadır.

Genel olarak, ÇBMTK modelleri üzerinde deneysel araştırma yapan çok az çalışma bulunmaktadır. Deneysel olarak ÇBMTK'nın, TBMTK'da DMF'yi kontrol etme amacıyla kullanımını inceleyen bir çalışmaya ise rastlanmamıştır. Bu çalışmada, farklı koşullar altında DMF etkisinin kontrol edilmesinde, ÇBMTK modelinin kullanılması durumuna ilişkin önemli bilgiler sunulmuştur. Bununla birlikte elde edilen bilgiler gerçek veriler üzerinde denenerek örnek bir uygulama gösterilmiştir. Bu kapsamda bu çalışma, elde edilen bilgilerin gerçek durumlarda nasıl kullanılabileceğinin ortaya konması ve pratik çözümlerde bulunması bağlamında oldukça önemlidir. Tüm çalışmaların, gerçek durumlarda karşılaşılan problemlerin çözümüne katkı getirmesi beklenmektedir. Yapılan bu çalışma da gerçek durum incelemesi ile bu katkıyı sağladığından, bu anlamda oldukça önemlidir. Bu kapsamda bu çalışmanın bahsedilen teknikolojiler hakkında mevcut bilgilere katkı sağlayacağı düşünülmektedir.

Sınırlılıklar

Gerçek veri seti ile ilgili olarak, bireylere ait sadece isim ve cinsiyet bilgisine ulaşılabilmiştir. DMF incelemelerinde grup değişkeni olarak sosyoekonomik düzey, bölüm (sayısal, sözel) gibi değişkenlerin de dikkate alınması düşünülmüş, ancak gerekli bilgilere ulaşılamamıştır. Bu kapsamda DMF incelemelerinde grup değişkeni "cinsiyet" ile sınırlandırılmıştır.

BÖLÜM 2

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde, MTK, TBMTK, ÇBMTK, DMF ile ilgili kuramsal açıklamalara ve ilgili araştırmalara yer verilmiştir.

Ek Madde Tepki Kuramı

MTK, psikometri alanında kullanılan matematiksel modellerin bir ailesidir ve sınavları tasarlamak, analiz etmek ve puanlamak için kullanılır (Lord, 1953). Bireylerin test performansının bireylere ait yetenek tarafından açıklanabileceğini vurgular. Bununla birlikte bireyin madde üzerindeki performansı ile bu performansı etkileyen yetenek arasındaki ilişkinin monoton bir şekilde arttığını ifade eder (Hambleton ve Swaminathan, 1985). Bu ilişkiyi tanımlayan kavramlardan biri olan Madde Karakteristik Eğrisi (MKE) grafiksel ve matematiksel açıdan da güçlü bir temele dayanmaktadır. MTK'da MKE, bir maddeyi doğru yanıtlama olasılığının test tarafından ölçülen yetenek üzerindeki doğrusal olmayan regresyonunu ifade eder. Bireyler ve maddeler aynı ölçek üzerine ölçeklendirilir. Her bireyin, her bir madde için gizil bir yeteneğe sahip olduğu varsayılır (Baker, 2001;Reckase, 2000).

Klasik Test Kuramı, gözlenen puanlardan gerçek puanlara yönelik çıkarımlar yapmaya çalışırken, MTK, bireyin yeteneğinden, bir maddeyi doğru cevaplama olasılığı üzerine odaklanır. MTK bireylerin gerçek yeteneklerinin testlerden bağımsız olarak ölçüldüğü prensibine dayanmakta; aynı yeteneği ölçmeyi amaçlayan testlerde bireylerin yetenek ve gerçek puanlarının sabit kaldığını kabul etmektedir (Hambleton ve Jones, 1993). Klasik gerçek puan modelinin aksine, MTK modelleri yanlışlanabilir modellerdir ve bu kapsamda model-veri uyumu önemlidir. Bu kapsamda, MTK olasılıksal bir model olduğu için test / paralel formlar geliştirme, test eşitleme, BOBUT uygulamaları ve DMF analizleri gibi alanlarda avantaj sağlamaktadır (Embretson ve Reise, 2000).

Ölçülen özelliğin tek ya da çok boyutlu olmasına bağlı olarak tek ve çok boyutlu MTK modelleri bulunmaktadır. Eğitim ve psikoloji alanında TBMTK modelleri yaygın olarak kullanılmaktadır. Bununla birlikte testteki tüm maddelerin tek bir gizil özelliği

ölçmesi anlamına gelen tek boyutluluk varsayımı özellikle başarı ve yetenek testlerinde çoğu zaman tamamıyla karşılanamamaktadır (Reckase, 2009). Çünkü psikolojik değişkenler, tek boyutta ifade edilememektedir. Başarı testlerinde de maddelere cevap verirken, bireyler arası farklılık gösteren birden çok bilişsel süreç etkili olmaktadır.

Tek boyutluluk sağlanmadığında test alanlar tek bir özellik üzerinde sıralanamamaktadır. Çünkü çok boyutlu örtük uzay ile bireylerin tek bir yeteneği birebir eşleştirilememektedir. Bu gibi durumlarda TBMTK modelleri yerine bu modellerin çok boyutlu veri setleri için genişletilmiş hali olan ÇBMTK modellerinin kullanılması tavsiye edilir (Ackerman, 1992). Böylece, bireylerin maddelere verdikleri yanıtların olasılıkları ile çok boyutlu uzaydaki yeri arasındaki ilişki ifade edilebilmektedir. ÇBMTK'da da TBMTK'da olduğu gibi test maddesi esastır.

Çok Boyutlu Madde Tepki Kuramı

ÇBMTK, iki ya da daha çok gizil değişken ile cevaplayıcının maddeye doğru yanıt verme olasılığı arasındaki bağlantıyı matematiksel bir modelle açıklayan modellerdir (Ackerman vd., 2003). TBMTK'da bir özelliğin, bireylerin cevaplarını etkilediği varsayılırken, ÇBMTK'da bireylerin cevaplarını birçok özelliğin etkilediği varsaymaktadır. Reckase (2009) de, test maddesine verilecek cevabın testi alan bireylerin yeteneğine ve test maddesinin özelliğine bağlı olduğunu belirtmiştir. Bu durumda; testi alan bireyin soruyu anlaması, soruyu cevaplama ve kaydetme yolunu bilmesi, çeldiricilerde yer alan terimlerin anlamlarını bilmesi gibi birçok faktör etkili olmaktadır. Diğer bir ifadeyle, ÇBMTK'nın temelinde tek bir maddede çoklu bilişsel yeteneklerin olması yatmaktadır. TBMTK'dan farklı olarak her bir boyut için birey yeteneği farklı bir değer aldığından; bireye ait yetenekler vektör halinde tanımlanmaktadır.

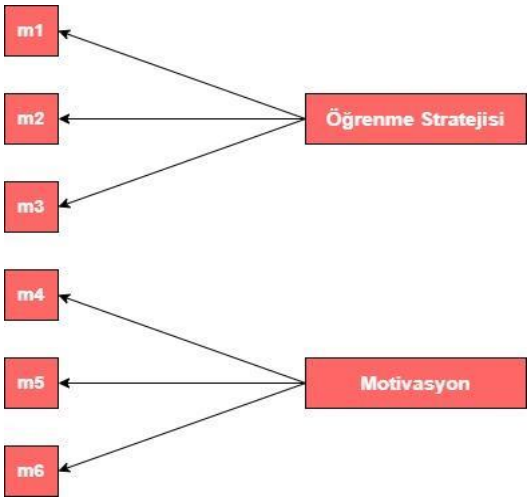
ÇBMTK'da boyut belirlemek önemli bir aşamadır. Ölçmede boyut, bireylerin sıralanabildiği ölçek olarak tanımlanmaktadır. Boyut kavramı varlıkların özelliklerinin, istenilen neden doğrultusunda karşılaştırılmak üzere sadeleştirilmesi olarak ifade edilmektedir. Jacoby'e (1991) göre maddeler arasındaki farklı ve dikkate değer değişkenlik kaynaklarının sayısı; madde kümesine ait boyutluluk olarak tanımlanabilmektedir. Diğer bir ifadeyle tek bir maddenin birden çok bilişsel yeteneği ölçmesidir.

Bir madde kümesinin boyut olarak ele alınabilmesi için, kümedeki bireylerin diğer bireylerden farklı ve ortak bir özelliğe sahip olması gerektiği söylenebilir. Boyut belirlemenin amacı; Fizikte 70, Kimyada 80 puan gibi bireyin ilgilenilen boyutlardaki yerinin tanımlanmasıdır. Matematik testindeki cebir ve geometri maddeleri gibi farklı alt boyutlardan oluşan bir test çok boyutlu olduğunda; öğrencinin performansını belirlemek için toplam puanın kullanabilme durumu tartışmalıdır (Meara vd., 2000).

ÇBMTK'da Boyutluluk

ÇBMTK'da, test maddelerinin boyutları ile bireylerin bu boyutlardaki performansının etkileşimine odaklanılmaktadır. Bireylerin farklı boyutlardaki performansı belirlenebilir. ÇBMTK'da birey yetenek düzeylerinden maddelere değer atanmaktadır. Diğer bir ifadeyle maddeler çok boyutlu uzayda modellenmektedir. Boyutluluk maddeler-arası boyutluluk (between-item dimensionality) ve madde-içi boyutluluk (within-item dimensionality) olmak üzere ikiye ayrılmaktadır (Wang vd., 2004). Buna ek olarak iki modelin bir arada kullanıldığı karmaşık yapılı model de bulunmaktadır. Bu kapsamda ilk olarak maddeler arası boyutluluk aşağıda kısaca açıklanmıştır.

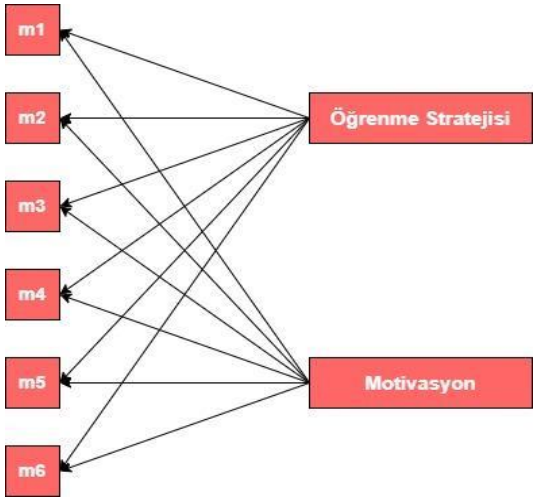
Basit Yapı (Çoklu Tek Boyutlu Model) - Maddeler Arası. Çoklu-tek boyutlu model olarak da adlandırılan bu modelde, birden fazla alt test bulunur ve her biri kendi içinde homojen bir yapıya sahiptir. Bu model, konu alanlarındaki kazanımlar için uygun bir seçenektir. Bu modelde her bir madde sadece bir boyutta yük vermekte; diğer bir ifadeyle madde ayırt ediciliği tek boyutta sıfırdan farklıdır. Uluslararası Matematik ve Fen Eğilimleri Araştırması (Trends in International Mathematics and Science Study-TIMSS) örneğinde olduğu gibi her biri tek bir yeteneği / özelliği ölçmektedir. Şekil 1'de basit yapılı çok boyutlu model örneği gösterilmiştir.



Şekil 1. Basit Yapılı Çok Boyutlu Model

Şekil 1 incelendiğinde, testin farklı tek boyutlu alt testlerden oluştuğu görülmektedir. Bu nedenle çok boyutlu modeller olarak adlandırılmaktadır. Her madde sadece bir yetenekle ilişkili olduğu bu modelde test verisinin boyutluluğu maddeler arası model olarak adlandırılmaktadır.

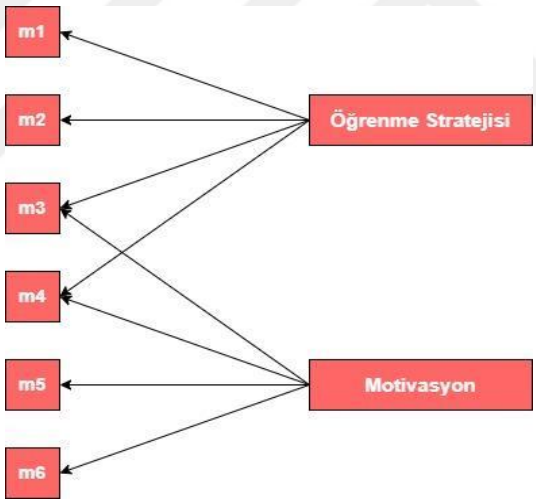
Karmaşık Yapı - Maddeler İçi. Her bir boyuttaki yükler, her maddenin ilgili boyuttaki etkisini gösterir ve ÇBMTK'da bu yükler a parametresi olarak ifade edilir. Diğer bir ifadeyle her maddenin her bir boyutta ayırt ediciliği vardır (Steinberg, Thissen ve Wainer, 2000). Maddeler, tek bir yetenekle ilişkili değildir; farklı yetenekler için kullanılırlar. Şekil 2'de karmaşık yapılı çok boyutlu model örneği gösterilmiştir.



Şekil 2. Karmaşık Yapılı Çok Boyutlu Model Örneği

Şekil 2 incelendiğinde, maddelerin birden çok yetenekle ilişkili olduğu görülmektedir. Bu modelde test verisinin boyutluluğu maddeler içi model olarak adlandırılır. Basit modelde a parametresi, bir boyut için geçerli iken; karmaşık modelde her boyut için farklı a parametreleri olduğu görülmektedir. Burada a parametresi faktör yükü gibi yorumlanır. İlk olarak, lineer olmayan faktör analiziyle boyutlar belirlenir; daha sonra madde parametreleri kestirilir. Karmaşık modelde her boyut için θ kestirilmekte ve çeşitli yollarla b,irleştirilerek tek bir θ değeri elde edilmektedir.

Karışık Model: Basit ve Karmaşık Yapı. Bu modelde de birden fazla alt test bulunur. Maddeler genel olarak tek bir yetenekle ilişkili olmakla birlikte bazı maddeler birden fazla yetenekle ilişkilidir. Bazı maddeler farklı yetenekler için kullanılmaktadır. Bu maddelerin her bir boyutta farklı ayırt ediciliği vardır. Şekil 3'de karışık model örneği gösterilmiştir.



Şekil 3. Karışık Model: Basit ve Karmaşık Yapı Birlikte Örneği

Şekil 3 incelendiğinde bazı maddelerin birden çok yetenekle ilişkili olduğu görülmektedir. Buna ek olarak bazı maddelerin ise tek bir yetenekle ilişkili olduğu gözlenmektedir. Bu kapsamda bu modelde test verisinin boyutluluğu karışık model olarak adlandırılır.

ÇBMTK genel olarak iki temel amaçla kullanılmaktadır. İlk amaç, bir test maddesini doğru cevaplayabilme yeteneğine ve becerisine sahip olan bireylerle test maddesinin özellikleri arasındaki etkileşimi ortaya koymaktır. Bu sayede, test maddelerinin karakteristikleri ile bireylerin yeteneklerinin uyumlu olup olmadığı

değerlendirilebilir. Bir diğer amaç ise test maddelerinin duyarlı olduğu boyutlara uygun şekilde karar verebilmektir. Bu, test yapısı ve içeriği hakkında daha sağlam ve geçerli kararlar almayı sağlar. Bununla birlikte ÇBMTK modeli yıllar içinde; farklı uygulama alanlarında yaygın olarak kullanılmıştır. Bu uygulamalar aşağıda maddeler halinde sıralanmıştır:

1. Eğitimde ölçme ve değerlendirme süreçlerinde
2. Psikolojik testlerin geliştirilmesinde ve analizinde
3. Sağlık alanında hastaların durumunu değerlendirmede
4. Dil ve çeviri çalışmalarında çeviri kalitesinin ölçülmesinde
5. Sosyal bilimlerde araştırmalarda
6. İş yerinde performans değerlendirmesinde
7. Yetenek ve yetenekli bireylerin belirlenmesinde
8. Eşleştirme ve sıralama süreçlerinde

ÇBMTK, ölçme ve değerlendirme süreçlerinde güçlük ve ayırıcılığın çok boyutlu kapsamda açıklanması, grup farklılıkları ve madde hassasiyetinin belirlenmesi, DMF'nin nedenlerinin belirlenmesi, teste madde seçiminin sağlanması, farklı madde gruplarının kalibrasyonu ve aynı çok boyutlu koordinat düzlemine yerleştirilmesi, ÇBMTK'da madde-birey etkileşiminin doğru şekilde modellenmesi ve veri indirgenmesi, bireysel farklılıkların belirlenmesi için çok boyutlu uzayın tanımlanması ile çok boyutlu uzayda bireyleri iyi tanımlayabilmek için maddelerin iyi bir şekilde yazılması alanlarında kullanılmaktadır. Bu uygulamalar, ÇBMTK'nın çok yönlü ve esnek kullanımının bir göstergesidir. ÇBMTK; testlerin geliştirilmesi, değerlendirme süreçlerinde kullanılması, madde ve grup analizlerindeki farklılıkların incelenmesi ve bireyler arasındaki farklılıkların anlaşılması gibi alanlarda önemli bir araç olarak kullanılmaktadır.

ÇBMTK modellerinin kullanım alanlarından biri de DMF'nin nedenlerinin belirlenmesidir (Reckase, 2009). İstenmeyen bir boyutun ortaya çıkması, DMF'nin nedenlerini açıklamada yardımcı olur ve bu, ÇBMTK'nın bir avantajıdır. Bunun yanı sıra; çoklu yetenek gerektiren bir maddeye doğru yanıt verebilmek için gereken yetenek düzeylerini ölçmede kullanılmaktadır. ÇBMTK'da doğru yanıtlama olasılığı tek bir puan yerine bir yetenek profili ortaya koymaktadır. Bu yaklaşım, bireylerin yeteneklerini daha

kapsamlı bir şekilde değerlendirmeye imkan sağlamakta ve daha detaylı bilgiler sunmaktadır. Bununla birlikte ÇBMTK öncelikle geniş ölçekli test uygulamaları başta olmak üzere başarı testleri için madde seçiminde kullanılabilmektedir. Çok boyutlu testlerde, testlerin ölçmek istediği özelliklere uygun maddelerin seçilmesi, testlerin güvenilir ve geçerli sonuçlar üretebilmesi için ÇBMTK kullanımı oldukça önemlidir. Veri seti çok boyutlu olduğunda ÇBMTK modellerinin TBMTK modellerinden daha üstün olduğu sonucuna ulaşan birçok araştırma mevcuttur. Buna ek olarak ÇBMTK, test eşitleme ve bilgisayar ortamında bireye uyarlanmış test uygulamalarında da kullanılmaktadır.

ÇBMTK'nın yukarıda belirtilen avantajlarının yanı sıra birtakım dezavantajları bulunmaktadır. Bunlar; boyutluluğun belirlenmesinde ve yorumlanmasında kullanılabilecek en uygun yöntem konusunda uzlaşma sağlanamaması, temel modellerin birçok formu ortaya konulmasına rağmen pratik yetersizliğinden dolayı bazılarının çok fazla uygulama alanı bulamaması, madde parametreleri tahminlerini elde etmek için en az 2000 kişilik cevaplayıcı gerekmesidir (Embretson ve Reise, 2000; Reckase, 2009). Bununla birlikte ÇBMTK'da az sayıda madde ile sağlıklı kestirim veremediğinden, çok sayıda parametre kestirimi yapılması gerekmektedir. Bu durum parametre kestiriminde standart hatanın da artmasına neden olabilmektedir.

Belirtilen avantaj ve dezavantajları dikkate alınarak kullanılmasına karar verilen ÇBMTK modellerinin kullanabilmesi için dikkate alınması gereken varsayımları bulunmaktadır. İlk olarak, bireyin ölçülen boyutta yerinin sabit olduğu, yani test alma sürecinde değişmediği kabul edilir. Benzer şekilde, aynı amaç için kullanılan uygulamalarda test maddesinin özelliğinin sabit olduğu varsayılır. ÇBMTK'da da TBMTK'nın varsayımlarından olan yerel bağımsızlık varsayımı geçerlidir. Bu varsayıma göre aynı yetenek vektöründe bulunan bireylerin cevapları sadece yetenek vektöründen ve madde parametre vektöründen etkilenir. Birinci varsayımla ilişkili olan bu varsayımda bireyin bir maddeye verdiği yanıtlar, diğer test maddelerinden bağımsızdır. Bu varsayımda; test maddelerinin birbirine ipucu vermemesi ve bir bireyin bir maddeye verdiği cevabın diğer bireyleri etkilememesi beklenir. Bununla birlikte, bireyin bir maddeye verdiği yanıtların diğer test maddelerinden bağımsız olduğu varsayımı her zaman tam olarak karşılanamaz. Örneğin kopya çekildiğinde birey cevapları birbirinden etkilendiğinden bu varsayım karşılanamaz. Bununla birlikte; bireyin test performansının dikkat, motivasyon gibi diğer bilişsel ve kişisel faktörlerden etkilenmesi nedeniyle yerel bağımsızlık varsayımı hiçbir zaman tam olarak karşılanamaz.

Her maddenin çok boyutlu uzaydaki yeri yani konumu için sadece bir tane doğru cevaplama olasılığı değeri vardır. ÇBMTK modellerinin çoğu, doğru-yanlış olarak puanlanan maddelerde, yetenek vektörünün bileşenlerinden herhangi biri arttığında doğru yanıtlama olasılığının da arttığını kabul eder. Diğer bir ifadeyle koordinat sisteminde bireyin yeri arttıkça, test maddesini doğru yanıtlama olasılığı da artar. Monotonicty (monoton artma) varsayımı denilen bu varsayım daha çok bilişsel yeteneklerin ölçülmesinde mantıklıdır. Bu varsayımların dikkate alınması, ÇBMTK modellerinin doğru şekilde uygulanabilmesi için önemlidir.

Yukarıda belirtilen varsayımlara uygun olan MTK modelleri çeşitli şekillerde sınıflandırılmaktadır. Testin boyutluluğu dikkate alındığında, tek boyutlu ve çok boyutlu olarak kategorize edilebilirler. Puanlama şekline göre ikili ve çoklu puanlanan olarak ayrılabilirler. Bununla birlikte, doğrusallık yapısına göre doğrusal ve doğrusal olmayan modeller olarak sınıflandırılabilirler (McDonald, 1982). TBMTK modelleri, içerdikleri parametre sayısıyla adlandırılmaktadır. Diğer taraftan, ÇBMTK modelleri ise parametre sayısıyla birlikte gizil özellikler arasındaki ilişkinin telafi edilebilir olup olmamasına göre telafisel (compensatory) veya telafisel olmayan (non-compensatory) olmasına göre adlandırılmaktadır (Ackerman, 1996; Reckase, 2009; Spencer, 2004).

Çok Boyutluluk İçin Genişletilmiş Lojistik Modeller

Telafisel olan ve olmayan modeller için genişletilmiş tek boyutlu modeller olmak üzere ikiye ayrılmaktadır. Telafisel olan modeller için genişletilmiş tek boyutlu modeller başlığı altında ikili ve çoklu puanlanan maddeler için ÇBMTK modelleri ayrı ayrı verilmiştir. İlgili başlıklar aşağıda verilmiştir.

Telafisel Olan Modeller için Genişletilmiş Tek Boyutlu Modeller

Bir yetenekteki zayıflığın, maddenin ilişkili olduğu diğer boyutta bulunan yüksek yetenek tarafından telafi edildiği varsayımına dayanmaktadır. Bir boyuttaki düşük yetenek, diğer boyuttaki yüksek yetenek tarafından telafi edilmektedir. Başka bir deyişle, toplam değerlerinin aynı olabilmesi için bir yetenek koordinatı düşükse, diğer yetenek değeri yüksek olmaktadır. Yetenek (θ), koordinatlarının doğrusal bileşimine bağlıdır. Maddenin doğru cevaplanma olasılığı, maddenin ilişkili olduğu yeteneklerin

olasılıklarının toplamını temsil etmektedir. Bu modele göre cevaplama olasılıklarını hesaplayabilmek için normal ogive veya lojistik form kullanılmaktadır (Reckase, 2009). Teleafisel olan modeller için genişletilmiş tek boyutlu modeller, ikili ve çoklu puanlanan maddeler için ÇBMTK modelleri olmak üzere ikiye ayrılmaktadır. Bu kapsamda ikili puanlanan maddeler için ÇBMTK modelleri; matematiksel fonksiyonlar; yetenek (θ), ayırıcılık (α), güçlük (D) ve şans parametresi üzerinden ele alınmaktadır. Boyut sayısı kadar α ve θ parametresi bulunur. Bununla birlikte yalnızca bir tane D parametresi vardır. Bu modeller aşağıda kısaca ayrı başlıklar halinde verilmiştir.

Çok Boyutluluk için Genişletilmiş Rasch Modeli. Çok Boyutluluk için Genişletilmiş 2 PLM'deki a değerinin 1'e eşitlenmiş halidir. a=1 olduğunda, tek boyutlu Rasch modeline benzer sonuçlar üretir. Tek farkı, yetenek parametresinin tek bir yapı yerine birden fazla yapıyı temsil eden koordinatların toplamı olmasıdır. q_{iu}, i maddesi için u gizil yeteneği ayrıştıran ağırlık matrisi olmak üzere Rost ve Carstensen (2002) tarafından çok boyutluluk için genişletilmiş Rasch modeli kullanımında Eşitlik 1'deki formül verilmiştir:

$$P\,(X = x \mid v,i) = \frac{\exp(\sum_{u=1}^{U} \mathrm{qiu}\theta\mathrm{vu} - \sigma\mathrm{i})}{1 + \exp(\sum_{u=1}^{U} \mathrm{qiu}\theta\mathrm{vu} - \sigma\mathrm{i})} \qquad (1)$$

Bu model, her bir yanıtın gizil yetenek parametrelerinin ayrıştırılarak birden fazla gizil boyut ile ilişkilendirilmesine olanak sağlamaktadır. Bu şekilde, yetenek parametresi birden fazla gizil değişken koordinatların toplamından elde edilen değeri yorumlayabilmektedir (Reckase, 2009).

Çok Boyutluluk için Genişletilmiş 2PLM. Reckase (1985) tarafından geliştirilmiş çok boyutlu ölçüm modellerinden biridir. Çok boyutluluk için genişletilmiş 2PLM, TBMTK'da yer alan 2PLM'den türetilmiştir. Burada yetenek parametresi tek bir yapının düzeyini değil, birden fazla yapıyı temsil eden gizil değişken koordinatlarının toplamını kullanmaktadır. Genişletilmiş 2PLM'e ait en genel formül Eşitlik 2'de verilmiştir.

$$P(U_{ij}= 1| \theta_j,a_i,d_i) = \frac{e^{ai\theta j+di}}{1+e^{ai\theta j+di}} \quad (2)$$

U_{ij} : j bireyinin i maddesine verdiği cevap

θ_j : j kişisinin yetenek düzeyi

a_i : i maddesinin eğim parametresi

di : i maddesinin kesişim parametresi

Formülde b parametresi yerine d parametresi kullanılmıştır. "d" parametresi, bu modeli tek boyutlu iki parametreli lojistik modelden ayıran en temel farklardan biridir. α ve θ parametreleri boyut sayısı kadardır. d parametresi ise farklı boyutlardaki yetenekler birbirini tamamladığından bir tanedir. Her bir koordinat boyutu için θ değerlerinin kuramsal olarak - ∞ ve + ∞ arasında olması beklenir. Bununla birlikte genellikle θ aralığı - 4 ve + 4 arasında değişmektedir. Sıfır noktası ve ölçme birimi her koordinat boyutu için keyfidir.

θ uzayında doğru yanıtlama olasılığı 0.5 olduğunda, a parametresi en yüksek değerini alır. Eğimle ilgili olan a parametresi, TBMTK'da olduğu gibi ayırıcılık parametresidir. Bununla birlikte, TBMTK'dan farklı olarak çok boyutlu genişletilmiş iki parametreli lojistik modelde her boyut için ayrı bir a parametresi bulunmaktadır. Öte yandan d parametresi kesişim parametresi olup TBMTK'daki güçlük ile aynı değildir. d parametresinin negatif değerinin, ayırt edicilik parametresinin değerine bölünmesi (-d / a_i) ile elde edilen değer maddeye ait göreceli güçlük değerini vermektedir. Bununla birlikte bu tür yapılan güçlük tahmininde bazı sorunlar olması nedeniyle, alternatif olarak, çizginin orijine olan uzaklığı hesaplanmaktadır. Bu hesaplama ile elde edilen b parametresi TBMTK ile aynı şekilde yorumlanmakla birlikte; TBMTK'daki b değeriyle karıştırılmaması için Çok Boyutlu Madde Güçlüğü (ÇBMG) olarak adlandırılmaktadır. ÇBMTK'da ÇBMG, diğer boyutlardaki vektör değerleri sıfır (0) olmak üzere, 0.5 olasılıkla doğru yanıt verme olasılığı anlamındadır. Bu vektör, sıfır vektörü olarak adlandırılmaktadır (Reckase, 2009). Bu model şans başarısının olmadığı ya da ihmal edilebildiği; bununla birlikte maddeler boyunca ayırt edicilik parametrelerinin farklı değerler aldığı durumlarda kullanılmaktadır.

Çok Boyutluluk için Genişletilmiş 3PLM. Çok Boyutluluk için Genişletilmiş 2PLM'ye c parametresinin yani şansla tahmin etme parametresinin eklenmiş halidir. Bu

bireylerin soruyu doğru yanıtlama durumu genelde yapı ile ilgili olmadığından, düşük yetenekli bireylerin bile belli bir olasılıkla soruyu yanıtlama ihtimali vardır. Model uyumu iyi olduğunda, elde edilen model, k boyutlu uzay için verilen $\theta = (\theta_1, \theta_2, \theta_3, \ldots \theta_k)$ koordinatlarında doğru yanıtların koşullu olasılığını kestirmektedir. Burada k, modelde yer alan boyut sayısıdır. θ düzeyindeki bir bireyin i maddesini doğru cevaplama olasılığını veren çok boyutluluk için genişletilmiş 3PLM'nin en genel haline ait formül Eşitlik 3'te verilmiştir:

$$P(u_{ij}=1|\theta_j, a_i, c_i, d_i) = c_i + (1-c_i)\frac{e^{ai\theta j+di}}{1+e^{ai\theta j+di}} \quad (3)$$

a_i : i maddesinin ayırıcılık gücü ile ilgili parametrelerin 1 x k vektörü

c_i : Şans parametresi

di : TBMTK modelindeki güçlük parametresi

Çok boyutluluk için genişletilmiş 3PLM'de cevaplara ait olasılıklar sıfıra doğru ilerlememektedir. Bu modelde şans parametresi de dikkate alındığından bu olasılıklar c parametresine doğru ilerlemektedir (Reckase, 2009). Bu model şans parametresinin ihmal edilemeyeceği durumlarda kullanılmaktadır.

Çok Boyutluluk İçin Genişletilmiş Normal - Ogive Modeller. ÇBMTK'nın temel işlemleri normal ogive kullanılarak yapılmaktadır. Bu modelde, çok boyutluluk uzayındaki yer ile bir test maddesine doğru cevap verme olasılığı arasındaki ilişki tanımlanmaktadır. Buna bağlı olarak mevcut istatistiksel kestirim programlarının birçoğu ogive modellere dayalı olarak da kestirim yapmaktadır (Reckase , 2009). Ogive modele ait formül Eşitlik 4'te verilmiştir:

$$P(U_{ij}=1|\theta_j, a_i, c_i, d_i) = c_i + (1-c_i)\frac{1}{\sqrt{2\pi}}\int_{-zi\,(\theta j)}^{\infty} e^{\frac{-t2}{2}}dt \quad (4)$$

U_{ij} : j bireyinin i maddesine verdiği cevap

θj : j bireyinin yetenek düzeyi

αi : i maddesinin ayırıcılığı

ci : i maddesinin en düşük asimptotu

di : i maddesinin güçlüğü (Schilling ve Bock, 2005)

Eşitlik 4'teki formülde z (θ_j), aθ_j ile d_i'nin toplamına eşittir. Bu modelde c parametresi sıfır olduğunda, çok boyutluluk için genişletilmiş iki parametreli lojistik model ile eşit sonuçlar vermektedir (Reckase, 2009). Bu modelde 1.70 sabiti kullanıldığında normal lojistik modelle arasında 0.01 gibi çok az bir fark olmaktadır. Bu durum bu iki modelin birbirleri yerine kullanılabileceği anlamına gelmektedir.

Çoklu puanlanan maddeler için ÇBMTK modelleri ise Çok Boyutlu Kısmi Modeli (ÇBKPM-Multidimensional Partial Credit Model), Çok Boyutlu Genelleştirilmiş Kısmi Modeli (ÇBGKPM- Multidimensional Generalized Partial Credit Model) ve Çok Boyutlu Aşamalı Tepki Modeli (ÇBATM- Multidimensional Graded Response Model)'dir. Bu modeller TBMTK'na ait modellerin çok boyutluluğa genelleştirilmiş halidir. Bu üç model aşağıda kısaca telafi edici model kapsamında açıklanmıştır.

Çok Boyutlu Kısmi Modeli (ÇBKPM) ve Çok Boyutlu Genelleştirilmiş Kısmi Modeli (ÇBGKPM). ÇBKPM'de maddeler, farklı boyutlarda farklı güçlüklere sahiptir. Aynı boyut içerisindeki cevap kategorilerinde ise maddeler aynı güçlüğe sahiptir. ÇBGKPM, tek boyutlu GKPM'den farklı olarak birbirinden ayrı güçlük ve eşik parametreleri içermemektedir.

Çok Boyutlu Aşamalı Tepki Modeli (ÇBATM). Muraki ve Carlson (1993) tarafından geliştirilmiş bu modelde tek boyutlu aşamalı tepki modelinde olduğu gibi bir test maddesi ile belirlenen görevi başarmış biri zaten daha alttakileri yapmış demektir. Model k tane kategoriye sahipse, k kategorinin başarıyla tamamlanabilmesi için öncelikle k-1 kategorinin başarıyla tamamlanması gerekmektedir (Reckase, 2009). ÇBAT modeline ilişkin normal ogive formülü Eşitlik 5'te verilmiştir:

$$P(u_{ij}=k|\theta_j) = \frac{1}{\sqrt{2\mu}} \int_{ai'\theta j+di,k+1}^{ai\theta j+dik'} e^{\frac{-t2}{2}} dt \qquad (5)$$

K : Madde puanı

a_i : ayırt edicilik değeri

d_{ik} : herhangi bir maddenin k. kategoriye ulaşma kolaylığı

İlgili kategoriden belirli bir puan almanın kolay olduğu durumlarda, d_{ik} parametresi yüksek ve pozitif değerler almaktadır. Tersi durumda diğer bir ifadeyle ilgili kategoriden belirli bir puan almanın zor olduğu durumlarda, d_{ik} parametresi büyük negatif değerler almaktadır.

Telafisel Olmayan Modeller için Genişletilmiş Tek Boyutlu Modeller

Ölçülen özellik açısından bir boyutta çok düşük olan birey her zaman diğer boyutta çok yüksek olmayabilir. Okuma becerisi iyi olmayan bir bireyin, matematik becerisi yüksek olsa bile soruyu anlamadığı için çözememesi örneğinde olduğu gibi bazı durumlarda gerçekçi de olmayabilir. Bu gibi durumlarda bireyin bir yetenekteki zayıflığının diğer boyuttaki yüksek yetenek tarafından telafi edilmediği modellerden telafisel olmayan modeller için genişletilmiş tek boyutlu modeller kullanılmaktadır. Yetenekler arası telafisellik özelliği bulunmayan bu modele kısmi telafisel model de denmektedir (Ackerman, Gierl ve Walker, 2003; Reckase, 2009). Bu modelde i maddesine doğru cevap verilme olasılığı Sypmson (1978) tarafından Eşitlik 6'daki formülle gösterilmiştir:

$$P_i(\theta) = c_i + (1 - c_i) \prod_{l=1}^{L} \frac{e^{(a_{il}\theta_l - d_{il})}}{1+e^{(a_{il}\theta_l - d_{il})}} \quad (6)$$

θ_l : l. Yetenek düzeyine karşılık gelen θ vektörünün elemanı
c_i : i maddesine ilişkin şans parametresi
a_{il} : l. yetenek düzeyinde i maddesine ilişkin ayırt edicilik parametresi
d_{il} : l. yetenek düzeyinde i maddesine ilişkin güçlük parametresi

Bu modelde bilişsel beceriler parçalara ayrılmakta ve her parçada tek boyutlu model kullanılmaktadır. Bir maddenin ilişkili her boyut için θ, α ve d parametresi kestirilmektedir. Telafisel modelde alt boyutlar birbirlerini etkilemektedir. Bu nedenle yetenekler birbirine bağımlı kabul edilmekte ve dolayısıyla doğru cevap olasılıkları toplanmaktadır. Bu modelde ise boyutlar birbirinden bağımsız kabul edildiğinden tüm

boyutlara ait yeteneklerin olasılıkları çarpımı maddenin doğru yanıtlanma olasılığını vermektedir. Örneğin, hem sembolleri bilmeyi hem de aritmetik bilgi gerektiren bir matematik sorusunda bu iki yapıyla ilgili çeşitli yetenek kombinasyonları vardır. Burada θ_1 ve θ_2 olmak üzere iki tane yetenek koordinatı tanımlanarak θ'lar için doğru yanıtlama olasılıklarının kombinasyonları oluşturulur. θ_1 ve θ_2 'nin düşük olması doğru yanıtlama olasılığının düşeceğini; θ_1 ve θ_2 kombinasyonunda tutarlı bir artış olması ise, test maddesinin her koordinat düzlemindeki farklara yanı θ'lara hassas olduğunu göstermektedir (Reckase, 2009).

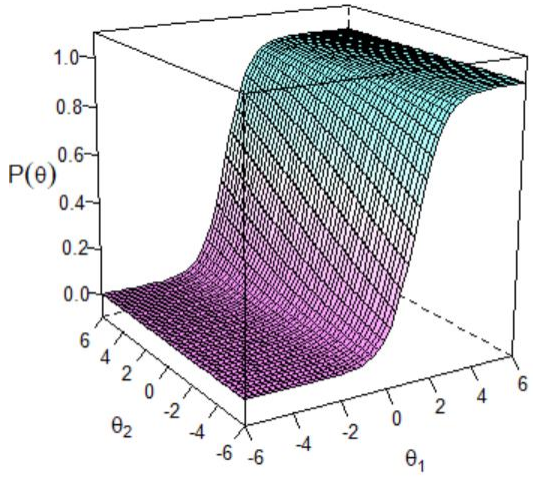
Farklı koordinat düzlemlerinde θ hızları da farklı olabilmektedir. Maddenin genel karakteristiği, maddelerin doğru yanıtlanma oranlarındaki artıştan izlenebilmektedir. Artışın yönü, test maddesinin ölçtüğü yapının düzeyi hakkında bilgi vermektedir. Modele ilişkin diğer bir özellik de boyut sayısının artmasıyla birlikte madde parametrelerinin yorumlamalarının da değişmesidir. Örneğin tek boyutlu modellerde θ=0 ve d=0 olduğunda p(θ)=0.5 olmakta; bununla birlikte kısmi telafi edici modellerde θ=0 ve d=0 olduğunda p(θ)=0.25 olmaktadır. Diğer bir ifadeyle θ=0 ve d=0 olduğunda, "m" boyut sayısı olmak üzere $p(\theta)= 0.5^m$ olmaktadır. Bu durumun nedeni telafisel modelde doğrusal fonksiyonla bir araya getirilen θ'ların; kısmi telafi edici modellerde çarpımla bir araya getirilmesidir. Bir diğer farkı ise; telafi edici modeller, birey-madde arasındaki ilişkiyi bütüncül bakış açısıyla değerlendirirken kısmi telafi edici modelin test maddesinin yetenek ve bilgi açısından farklı bölümleri olduğunda kullanılmasıdır (Reckase, 2009).

Çok Boyutlu Madde Tepki Kuramı Modelinin Parametreleri

ÇBMTK modeline ait parametreler ve parametre kestirim yöntemleri aşağıda ayrıntılı olarak verilmiştir.

Madde Karakteristik Yüzeyi (MKY)

TBMTK'deki madde karakteristik eğrisi yerini Madde Karakteristik Yüzeyine (MKY) bırakmaktadır. Çünkü çok boyutlu psikolojik yapılarda bireylerin maddeye doğru yanıtlama olasılıkları, bir tek yetenek düzeyine bağlı olarak belirlenememektedir. MKY, MKE gibi θ ve maddeyi doğru cevaplama olasılığını (P(θ)) içermektedir. Bu ilişki Şekil 4'te örnek olarak gösterildiği gibi çok boyutlu grafiksel düzlemde gösterilmektedir.



Şekil 4. Madde Karakteristik Yüzeyi Örneği

MKY'de maddelerin yorumlanmasında kolaylık sağlamak için Eş Olasılık Konturları (Çizgileri) kullanılmaktadır. Bunlar, MKY'ye ilişkin konturlardır. Bir bileşenin doğru yanıtlama olasılığı, o bileşendeki yetenek düzeyine dik çizilen doğrunun, asimptotu olduğu konturun değeridir. Eş olasılık çizgileri üzerinde bireylerin aynı noktada bulunmaları, maddeyi doğru cevaplama olasılıklarının eşit olduğunu göstermektedir. MKY'nin dikleşmesi ayırt ediciliğin arttığını göstermektedir.

Maddelerin birlikte gözlemlenmesinde her madde için ayrı kontur grafiği çizmek zorluk yaratmaktadır. Bu nedenle, birden fazla maddeyi aynı grafikte gösterebilmek için madde vektörleri kullanılmaktadır. Vektörün çizilebilmesi için ise uzunluk, orijine olan uzaklık ve yöne ihtiyaç vardır.

Uzunluk - Çok Boyutlu Madde Ayırıcılığı (ÇBMA). Madde ayırıcılığı madde karakteristik yüzeylerinin en dik olduğu noktanın eğimidir. ÇBMA, tek tek madde ayırt ediciliklerinin bir fonksiyonudur. TBMTK'dan farklı olarak vektörel bir büyüklük şeklinde tanımlanmaktadır. Burada vektörün uzunluğu, maddenin maksimum ayırıcılığı hakkında bilgi vermektedir. Uzunluk arttıkça ayırıcılık da artmaktadır. Madde iki boyutlu ise ÇBMA Eşitlik 7'deki formülle hesaplanmaktadır:

$$\text{ÇBMA} : \sqrt{ai1^2 + ai2^2} \qquad (7)$$

Vektörün ikinci veya dördüncü bölgede olması a parametrelerinden birinin negatif olduğunu göstermektedir. Bununla birlikte formülden dolayı her zaman pozitif (+) değer almaktadır. Bu nedenle vektör her zaman koordinat düzleminin birinci ve üçüncü bölgesindedir. Vektörün üçüncü bölgede olması maddenin çok kolay, birinci bölgede olması ise çok zor olduğunu göstermektedir. Buna ek olarak eş olasılık çizgileri (konturları) birbirine ne kadar yakınsa, madde o kadar ayırt edicidir. Her bir yetenekteki ayırt edicilik parametrelerinin bileşimini gösteren formül Eşitlik 8'de verilmiştir:

$$BMA_i = ÇBMA_i = \sqrt{\sum_{k=1}^{m} a_{ik}^2} \quad (8)$$

Maddenin hangi konuda ne kadar bilgi verdiğini gösteren a parametresi faktör yükü olarak da değerlendirilebilmektedir. Örneğin bir boyutta ayırıcılık yüksek, diğerinde düşük olması durumunda o maddenin bilgi verdiği ancak ilgili özelliğin daha az ölçüldüğü belirtilmektedir. Bununla birlikte iki boyutta da ayırıcılık düşükse o maddenin testten çıkarılması gerekmektedir (Reckase, 2009).

Orijine Olan Uzaklık - Çok Boyutlu Madde Güçlüğü (ÇBMG). Bir noktadaki eğim; o noktadaki doğru cevaplama olasılığına, a vektörünün elemanlarına ve α açısına bağlı olarak değişmektedir. Eğimin en dik olduğu noktanın orijinden olan uzaklığına madde güçlüğü denilmektedir. Diğer bir ifadeyle 0.5 olasılık konturunun, orijine olan uzaklığıdır. Buradan hareketle, madde vektörlerinin başlangıç noktasının, 0.5 olasılık konturundan başladığı belirtilebilir. Orijine olan uzaklık "$a^2 = b^2 + c^2$" şeklinde standart uzaklık formülü ile hesaplanmaktadır. m tane boyut varsa m-1 tane açı hesaplanmak üzere; aradaki açı komşu/hipotenüs yani kosinüs olarak hesaplanmaktadır. ÇBMG hesaplanmasına ilişkin formül Eşitlik 9'da verilmiştir:

$$D_i = \frac{-d_i}{\sqrt{\sum_{k=1}^{m} a_{ik}^2}} \quad (9)$$

d_i : Güçlük parametresi

a_i : i maddesi için ayırt edicilik parametresi vektörü

a_{ik} : i maddesinin k boyutu üzerindeki ayırt edicilik değeri

Yukarıdaki formülde hesaplanan D değerinin pozitif (+) olması maddenin zor olduğunu diğer bir ifadeyle doğru olarak yanıtlanabilmesi için yüksek düzeyde yeterlilik gerektirdiğini göstermektedir. Negatif (-) olması ise maddenin kolay olduğunu diğer bir ifadeyle doğru olarak yanıtlanabilmesi için düşük düzeyde yeterlilik gerektirdiği anlamına gelmektedir. Görüldüğü üzere ÇBMTK'da da madde güçlüğü TBMTK'daki gibi yorumlanmaktadır (Reckase ve McKinley, 1999).

Yön (α_i). Madde vektörünün, pozitif θ ekseniyle yaptığı açıdır. Açının ölçüsü hangi yetenek düzeyini ölçeceğini göstermektedir. Vektör 45^0'den küçükse diğer bir ifadeyle θ_1 değerine yakın ise maddeler öncelikli olarak θ_1'i; vektör 45^0'den büyükse diğer bir ifadeyle θ_2 değerine yakın ise maddeler öncelikli olarak θ_2'yi ölçmektedir. Vektör 45^0 ise diğer bir ifadeyle aralarındaki açı eşitse maddenin her iki yetenek düzeyini de eşit olarak ölçtüğü söylenebilmektedir. ÇBMA vektörünün α_i yönünü ifade eden matematiksel eşitlikse Eşitlik 10'da verilmiştir:

$$\alpha_i = \arccos. \frac{\alpha_i}{\text{MDİSCİ}} \quad (10)$$

Bu eşitlikte i maddesinin k ekseniyle yaptığı açı gösterilmektedir (Zhang, 2008).

Bilgi Fonksiyonu

ÇBMTK'daki madde bilgi fonksiyonun, TBMTK'daki madde bilgi fonksiyonuna genişletilmiş şeklidir (Reckase, 2009). Her yüzey için farklı madde bilgi fonksiyonu bulunmaktadır. Bu nedenle maddenin verdiği tüm bilgiyi belirlemek zor olmaktadır. Maddenin verdiği bilginin belirlenmesinde "istiridye kabuğu" şekli kullanılmaktadır. Bu şekil, her nokta için maddenin verdiği bilgiyi hesaplamaktadır. Bu bilgi maddenin ölçtüğü özelliği ne kadar iyi ölçtüğünü belirtmektedir. İkili puanlanan maddeler için TBMTK'daki bilgi fonksiyonuna bilginin yönünün de eklenmesiyle ÇBMTK'daki bilgi fonksiyonu elde edilmektedir (Ackerman, 2005). ÇBMTK'da bilgi fonksiyonunun matematiksel gösterimi Eşitlik 11'de verilmiştir:

Çok Boyutlu Bilgi = Ia (θ) = P_i (θ) θ_i (θ) [$\sum_{k=1}^{m} \alpha ik \cos \alpha ik$]2 (11)

$\alpha_i k$: Örtük özellik kompozisyonunda i maddesini temsil eden vektör (θ_1 ekseniyle yapılan açı ile gösterilmekte)

P_i (θ) : θ yetenek düzeyinde i maddesine doğru cevap verme olasılığı

Eşitlik 11'de verilen madde bilgi eşitliği incelendiğinde, maddenin verdiği bilgi düzeyinin maddeye ilişkin hata varyansı ile test orantılı olduğu görülmektedir. Bu durumda madde bilgisi her θ düzeyinde farklı değerler almaktadır (Baker ve Kim, 2004).

Test Fonksiyonu

TBMTK'daki test bilgi fonksiyonudur. ÇBMTK'da bu fonksiyon, madde puanları toplamının θ vektörü üzerine regresyonudur. Madde karakteristik yüzeylerinin toplamıdır. Eş olasılık konturları gösteriminde de madde bilgi fonksiyonları toplamıdır.

Çok Boyutlu Madde Tepki Kuramı Modelinin Parametre Kestirim Yöntemleri

MTK modelleri birey parametreleri ve madde parametreleri olmak üzere en az iki tür parametre kestirimi yapabilmektedir. Bu teknikler aşağıda kısaca açıklanmıştır.

Madde Parametre Kestirimi

MTK modelleri için madde parametreleri sıklıkla maksimum olabilirlik (Maximum Likelihood-ML) yaklaşımı kullanılarak kestirilmektedir. Eş zamanlı olabilirlik tekniği, birey ve madde parametrelerinin eş zamanlı kestirimini sağlayabilmekte olup kolay görünmekle birlikte uygulamada bazı sıkıntıları bulunmaktadır. Birincisi; doğrusal olmayan fonksiyonların daha zor çözülmesidir (Hambleton ve Swaminathan, 1985). İkincisi ise parametre kestirimi için örneklem büyüklüğünün 1000'den fazla olması gerekliliğidir (Lord ve Novick, 1968). Madde parametre kestirim teknikleri aşağıda kısaca açıklanmıştır.

Marjinal Maksimum Likelihood Tekniği (MML). Bock ve Aitkin (1981) tarafından geliştirilmiştir. G (θ) yetenek dağılımına sahip bir örneklemden rastgele seçilen bireylerin doğru cevap verme olasılığı ∫ P (θ) d G (θ)'dir. Bu örneklemden N sayıda birey seçilmiş ise, gözlenen veri için marjinal olabilirlik fonksiyonu Eşitlik 12'de verilmiştir:

$$LM\ (a,b,c \mid U) = \prod_{j=1}^{N} \int \prod_{j=1}^{N} Pi(\theta)\ uij Q(\theta) 1\text{-}uij d\ G\ (\theta) \qquad (12)$$

Bu teknikle maksimum Linear Model (LM) değerinin hesaplanmasıyla madde parametresi kestirimi gerçekleştirilmektedir. Bu teknik hem tüm modellerde hem de kısa ve uzun testlerde madde parametresi kestiriminde kullanılabilmektedir (Bock ve Aitkin, 1981). MML tekniğinde bazı özel durumlar hariç benzer yeteneğe sahip bireyler tarafından farklı maddelere verilen cevapların şartlı bağımsız olduğu varsayılmaktadır. Marjinal maksimum olabilirlik fonksiyonunu maksimize eden logaritma değerine ait formül Eşitlik 13'te verilmiştir:

$$\log LM = \sum_{l=1}^{s} r_l\, P\ (x_l) \qquad (13)$$

r_l : N bireyden oluşan gözlenen örneklemin içindeki x_l örüntüsünün frekansı

S : Gözlenen farklı desen sayısı

Madde parametreleri bilindiğinde; madde sayısı arttıkça birey parametreleri için, benzer olarak birey parametreleri biliniyorsa; örneklem büyüklüğü arttıkça madde parametreleri için daha doğru kestirimler yapabilmektedir. Bununla birlikte örneklem büyüklüğünün yeterli olması durumunda bile parametre kestiriminin doğruluğuna ilişkin garanti sağlanamayabilmektedir (Hulin vd., 1982).

Bock Aitkin Expectation - Maximization (BA-EM). EM algoritması, Bock ve Aitken (1981) tarafından MML tekniğine yeni bir yaklaşım olarak geliştirilmiştir. Bu yaklaşımın "expectation" (beklenti) ve "maximization" (maksimize etme) adımlarından oluşmaktadır. İlk adımda (expectation) istenilen log-olabilirlik değerleri için kestirilen parametreler kullanılarak bir fonksiyon oluşturulmakta; hesaplanan bu log-olabilirlik değerlerini maksimize eden parametrelerse ikinci adımda hesaplanmaktadır. Bu

algoritmada ilk adımda geçici kestirimler ve gözlenen veri üzerinde olasılıklı bir yapay veri seti oluşturulmaktadır. İkinci adımda ise parametreler kestirilmekte ve tekrar ilk adıma dönülerek kestirilen parametreler büyük ölçüde aynı olana kadar bu işlem tekrarlanmaktadır. BA-EM algoritmasısının veri yapısının üç boyutu geçmediği durumlarda kullanılması önerilmektedir. BA-EM tekniği tercih edilen bir teknik olmakla birlikte; kayıp veri olması ya da çok kategorili veri bulunması gibi durumlarda ise EM algoritmalarının uygulanması sıkıntılı olmakta; Bayesyan kestirim tekniklerinin kullanılması önerilmektedir.

Bayesyan Kestirim Teknikler. Bayesyan tekniklerin en önemli avantajı, bu teknikleri içeren programların ücretsiz olmasıdır. Buna karşın bu teknikle karmaşık modellere ait parametrelerin tutarlı kestirimleri için analizlerin saatler, hatta günler alması ise bu tekniğin en temel dezavantajıdır. Bayesyan model kestirimleri, küçük örneklemlerde farklı önsel dağılım değerlerinin farklı kestirimlere neden olması bakımından eleştirilmektedir. Bayes tekniğine ilişkin temel yaklaşımın formülü Eşitlik 14'te verilmiştir:

$$P\,(A \mid B) = \frac{P\,(B \mid A)\,P\,(A)}{P\,(B)} \qquad (14)$$

A olayının olasılığı P (A) iken, A olayı ile ilgili olasılık, B olayı gözlendikten sonra, B olayından sağlanan bilgiye dayalı olarak P (A | B) olmaktadır. Bayesyan yaklaşımlar da parametre kestirimlerinde olasılıklı bir dağılım kullanmakta olup bu olasılıklı dağılıma; önsel (prior) dağılım denmektedir. Dağılım veri seti gözlendikten sonra sonsal (posterior) dağılım olarak tekrar adlandırılmaktadır (Winkler, 1967).

Yetenek Parametresi Kestirimleri

Yetenek parametrelerinin kestirimi için genellikle parametrelerinin doğru kestirimini sağlayan ve yaygın kullanılan üç kestirim tekniği aşağıda verilmiştir.

Maksimum Olabilirlik (ML). Birey parametrelerini, bireyin madde cevaplarının olabilirliğini maksimize ederek kestirilmesine olanak tanıyan kestirim tekniğidir. Bu

teknik bütün maddelere doğru veya yanlış cevap verildiğinde parametre kestirimi yapamamaktadır (Embretson ve Reise, 2000). Hem EAP hem de MAP Bayesyan kestirim formundadır.

Maksimum Sonsal (MAP_Maximum a Posteriori). MAP kestirim tekniği, bir yeterlik düzeyini kestirmede, sonsal (posterior) dağılımı maksimize ederek birey yeterliği ile ilgili önsel bilgiyi ve olabilirlik fonksiyonunu birlikte kullanmaktadır. Tüm maddeler doğru veya yanlış olduğunda ya da madde yanıt örüntüleri çözümsüz veya birden fazla çözüme sahip olduğu durumlarda (odd pattern) ML kestirim yapamamaktadır. Bu tür sorunlar olduğunda, MAP tekniği, ML tekniğine göre daha avantajlıdır.

Beklenilen Sonsal (EAP_ Expected a posteriori) Yöntemler. Sonsal dağılımların ortalaması hesaplanarak yetenek parametresi kestirilmektedir. En küçük hata kareleri değerlerine sahip olması en avantajlı özelliğidir (Bock ve Mislevy, 1982). Bununla birlikte bazı durumlarda yanlı kestirimler yapabilmektedir.

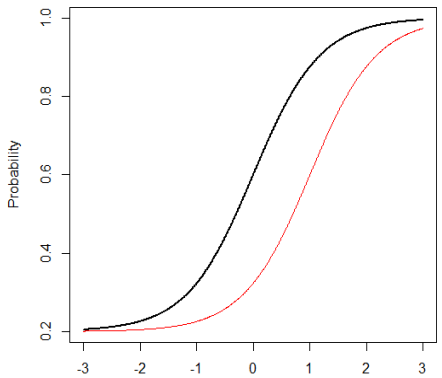
Değişen Madde Fonksiyonu

DMF, gruplardaki yetenek farklılıkları kontrol edildikten sonra maddenin farklı gruplarda farklı istatistiksel özellikler sergilemesi şeklinde tanımlanmaktadır (Angoff, 1973). Bir test maddesine verilen Y yanıtında sadece yetenek düzeyi (θ) ve tesadüfî hata etkili olmalıdır. Y yanıtının θ koşullu olasılığını gösteren $f(\theta)$'nin iki grup için karşılaştırıldığı varsayılsın. Y yanıtının koşullu olasılığının yalnızca θ'ya bağlı olması ve hata dağılımlarının her iki grup için aynı olması durumunda elde edilen formül Eşitlik 15'te verilmiştir:

$$f(\theta,\ G = R) = f(\theta,\ G = O) \tag{15}$$

Eşitlik 15'te, grup değişkeni G, çoğunluk veya avantajlı grup üyelerinden oluşan referans grup R ve azınlık veya dezavantajlı grup üyelerinden oluşan odak grup ise O ile gösterilmiştir. Y yanıtının koşullu olasılık dağılımının grup üyeliğinden bağımsız olması, DMF olmadığını göstermektedir. Aynı yeterlik düzeyinde yer alan bireylerin doğru yanıt

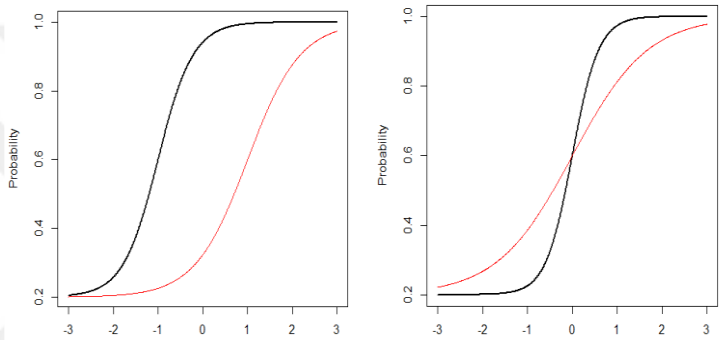
verme olasılıklarının eşit olması, grupların birbiri üzerinde avantajı olmadığını gösterir. Bununla birlikte aynı yeterlik düzeyinde ancak farklı gruplardaki bireylerin Y yanıtını seçme olasılıklarının farklı olması, grup üyeliği ile madde performansı arasında bağımlılık olduğunu gösterir. Bu durumda maddede dezavantajlı durumda olan grup, olasılığın düşük olduğu gruptur. Hataların eşit dağıldığı varsayıldığında ve yetenek düzeyi (θ) kontrol altına alındığında iki grup için madde fonksiyonunun değişmesi durumunda madde DMF gösterir. Bütün DMF belirleme tekniklerinin iki farklı amacı vardır. Bunlardan biri Eşitlik 15'te belirtilen H_0 hipotezini test etmektir. Bir diğer amacı ise Eşitlik 15'te belirtilen durumun ne ölçüde yanlışlandığını belirlemeye çalışmaktır (Penfield ve Camilli, 2007). DMF, grup üyeliği ile madde performansının koşullu bağımlılığı tarafından belirlenirken; θ ölçeği boyunca değişen olasılık değerleri tarafından da DMF türü belirlenmektedir. DMF türü gruplar arasında farklılık gösteren madde parametresine bağlı olarak, tek biçimli (uniform) ve tek biçimli olmayan (non-uniform) olmak üzere iki grupta değerlendirilir (Camilli ve Shepard, 1994). Tek biçimli (TB) DMF'de bir grupta bir maddeye doğru cevap verme olasılığı, tüm yetenek düzeylerinde diğer gruptan daha yüksektir. Gruplar arasında sadece b parametresi farklılık göstermektedir. Koşullu bağımlılık θ ölçeği boyunca göreli olarak aynı büyüklüğe sahiptir. Diğer bir ifadeyle madde θ ölçeği boyunca aynı grup için sürekli olarak göreli avantaj sağlamaktadır. Tek biçimli DMF örneği Şekil 5'te verilmiştir:



Şekil 5. Tek Biçimli DMF Örneği

Şekil 5'te görüldüğü üzere tek biçimli DMF'de madde yanıt eğrileri farklı olmakla birlikte kesişmez. Bir grubun diğer gruba kıyasla daha yüksek bir puan kazanma olasılığı θ aralığı boyunca az veya çok olacaktır. Koşullu bağımlılığın θ ölçeği boyunca

büyüklüğünün veya yönünün değişmesi ise Tek Biçimli Olmayan (TBO) DMF olarak bilinir. Koşullu bağımlılığın büyüklüğünün değişmesi, referans ve odak gruplar arasındaki olasılık farkının, farklı yetenek düzeylerinde nispeten farklılaşması olarak yorumlanabilir. Burada yön değişimi ise bir θ düzeyine kadar referans grup avantajlı iken o θ düzeyinden sonra odak grubun avantajlı olması durumunu ifade eder. Bu durum çapraz tek biçimli olmayan (crossing nonuniform) DMF olarak da bilinmekte olup zahmetlidir (Penfield ve Camilli, 2007). TBO DMF'de b parametresinin farklı olup olmadığına bakılmaksızın, a parametresi gruplar arasında farklılık gösterir. Tek biçimli olmayan ve çapraz tek biçimli olmayan DMF Şekil 6'da verilmiştir:



Şekil 6. Tek Biçimli Olmayan ve Çapraz Tek Biçimli Olmayan DMF Örneği

Şekil 6'da görüldüğü üzere TBO DMF ise bir grupta bir maddeye doğru yanıt verme olasılığı; bazı yetenek düzeylerinde diğer gruptan düşük, bazı yetenek düzeylerinde ise diğer gruptan yüksektir. Çapraz TBO DMF'de Madde Karakteristik Eğri'lerin (MKE) şekilleri iki grup arasında farklılık göstermekte ve genellikle MKE'leri kesişmektedir. Farkı türleri olan DMF'nin belirlenmesinde farklı yöntemler vardır. Bunlar Klasik Test Kuramı'na (KTK) ve MTK'ya dayalı yöntemler olarak iki grupta ele alınabilir (Atalay vd., 2012; Camilli ve Shepard, 1994; Ellis ve Raju, 2003; Millsap ve Everson, 1993). KTK'ya dayalı DMF Belirleme Yöntemleri; Dönüştürülmüş Madde Güçlüğü, Varyans Analizi, Mantel-Haenszel (Holland, 1985) ve Lojistik Regresyon Tekniği (LRT) (Swaminathan ve Rogers, 1990)'dir. MTK'ya dayalı DMF Belirleme Yöntemleri ise SIBTEST (Shealy ve Stout, 1993), Lord'un χ^2 Testi (Lord, 1980), Raju'nun Alan Ölçümleri (Raju,1988) ve MTK Olabilirlik Oranı (MTK-OO) (Thissen, 2001; Thissen vd., 1993) yöntemleridir. Bu yöntemlerden araştırma kapsamında kullanılanlar aşağıda ayrı başlıklar halinde kısaca açıklanmıştır.

Lojistik Regresyon (LR) Yöntemi

Lojistik Regresyon (LR) yöntemi Swaminathan ve Rogers (1990) tarafından geliştirilmiştir. LR yönteminde eşleşme değişkeni olarak toplam puan kullanılmakta; hem TB hem de TBO DMF belirlenebilmektedir (Gierl vd., 1999; Jodoin ve Gierl, 2001). DMF'yi araştırmak için; test puanını, grup değişkenini ve grup ile test puanları etkileşimini gösteren üç ayrı regresyon modeli ele alınmaktadır. Test puanı değişkeninin manidar olması modelin geçerliğini, grup değişkeninin manidar olması TB DMF'yi ve grup ile test puanları etkileşimi değişkeninin manidar olması ise TBO DMF'yi göstermektedir (Gierl vd., 1999). LR modelde etki büyüklükleri hesaplanırken yararlanılan standartlaştırılmış regresyon katsayıları (R^2) DMF'nin derecesini vermekte ve üç düzeyi bulunmaktadır (Hidalgo, Lopez-pina, 2004). Jodoin ve Gierl (2001) $\Delta R^{2'}$nin yordayıcısı olan SIB etki büyüklüğü ölçüsünü ölçüt olarak almıştır. ΔR^2 0.035'den küçükse A düzey DMF vardır ve DMF göz ardı edilebilir. ΔR^2 0.03 ile 0.070 arasında ise B düzey diğer bir deyişle orta düzey DMF varlığına işaret etmektedir. ΔR^2 0.070'e eşit veya büyükse ise C düzey DMF diğer bir ifadeyle yüksek düzeyde DMF olduğu belirtilmektedir.

SIBTEST

SIBTEST; Shealy ve Staut (1993) tarafından standardizasyon indeksinin bir modifikasyonu olarak geliştirilmiştir. SIBTEST ile hem madde hem de madde grubu (bundle) yanlılık incelemeleri yapılabilmektedir. Parametrik olmayan bir yöntem olan SIBTEST, tek ve çok boyutlu veriler için uygundur. SIBTEST'in en önemli özelliği; DMF incelemesi yapılan maddeler üzerindeki performansı karşılaştırmak için referans ve odak gruplarda regresyon düzeltmesi yapmasıdır. İlgili düzeltmeyle, referans ve odak grupların eşleştirilmesi MH gibi toplam test puanı veya ölçeklenmiş puanları kullanan yaklaşımlardan daha doğru olarak yapılabilmektedir. Gözlenen puanlar kullanıldığında ortaya konulamayacak olan fark, regresyon düzeltmesiyle ortaya çıkarılabilmektedir (Penfield ve Camili, 2006).

İki kategorili maddelerde TB olan ve TBO DMF belirlenmesinde kullanılabilmektedir. Yetenek düzeyi belirlenirken DMF sorgulaması yapılan maddeyi dikkate almamaktadır. Penfield ve Camili (2006)'ya göre tanımlanan düzey aralıkları H_0:

$B_{UNİ}$, | $B_{UNİ}$ |< .059 ise A düzeyi; H_1: $B_{UNİ} \neq 0$ ve .059 =<| $B_{UNİ}$ |< .088 ise B düzeyi; | $B_{UNİ}$ |≥.088 ise C düzeyi şeklindedir.

Lord'un χ^2

Lord'un χ^2 yöntemi; Lord (1980) tarafından geliştirilmiştir. MTK'ya dayalı bir yöntem olup her türlü madde yanıt modeline uygundur. Bununla birlikte madde parametreleri, istatistiksel testten önce ortak bir ölçü ile ölçeklendirilmelidir (Magis vd., 2010). Bu nedenle Lord'un χ^2 istatistiğinin hesaplanmasında öncelikle odak ve referans grupların madde parametreleri ve kovaryansları hesaplanmaktadır. Daha sonra ölçek denkliğini sağlamak için odak grubun parametre değerleri referans grubun parametre değerlerine dönüştürülmektedir. Dönüştürülmüş parametre ve kovaryans değerleri kullanılarak Lord'un χ^2 istatistiği hesaplanmaktadır. Gözlenen değerler, kritik değerle karşılaştırılmakta; DMF'nin olmadığını belirten yokluk hipotezi test edilmektedir. Hem TB hem de TBO DMF belirlemesinde kullanılabilmektedir (Cromwell, 2002). TBDMF analizlerinin yanı sıra ÇBDMF belirleme yöntemleri de bulunmaktadır. Bu yöntemler aşağıda ayrı başlıklar halinde verilmiştir.

Çok Boyutlu Değişen Madde Fonksiyonu

ÇBMTK'da ikinci bir faktör bulunur. Grupların bu ikinci faktör bakımından denk olmaması Y yanıtının koşullu beklenen değerinin gruplar arasında farklılaşmasına neden olabilir. Bu durum gruplardaki bireylerin bu ikinci boyuta sahip olma durumlarının farklılaştığını ve bu farklılığın grup üyelikleri ile ilişkili olduğunu gösterir. Diğer bir ifadeyle DMF söz konusudur. Söz konusu olan DMF'nin belirlenmesinde ÇBMTK'ya dayalı DMF belirleme yöntemleri geliştirilmiştir. Bunlar çok boyutlu SIBTEST (Stout vd., 1997), Lojistik Regresyon (Mazor vd., 1998), Madde Yanıt Teorisi Olasılık Oranı (MTK-LR) testi (Suh ve Cho, 2014) ve Çoklu Göstergeler Çoklu Nedenler (MIMIC) modeli (Lee vd., 2016) şeklinde sıralanabilir. Bu modeller tek boyutlu testler için tasarlanmış DMF algılama yaklaşımlarının çok boyutlu uzantıları olup bu yaklaşımlar arasında MIMIC ve MTK- OO yaygın olarak kullanılmaktadır. Bu durumda, MTK modelleriyle bağlantıları nedeniyle hem ikili hem de çoklu puanlanan maddelerde

DMF'yi tespit edebilmeleri ve kullanım kolaylıkları etkilidir. ÇBMTK'ya dayalı DMF belirleme yöntemlerinden kısaca aşağıda bahsedilmiştir.

Çok Boyutlu SIBTEST

Çok Boyutlu SIBTEST, Tek Boyutlu SIBTEST'in doğrudan bir uzantısıdır. Shealy ve Stout'ta (1993) tarafından sunulan bu istatistiksel test, çok boyutlu DMF modeline dayanmakta ve çok boyutlu verilerde DMF incelemektedir. Bazı iki boyutlu test verilerinde DMF'nin belirlenmesinde önerilmiştir. Bu durum iki boyutlu bir testte DMF'nin, amaçlanan iki boyuttan başka ikincil boyutsal etkiden kaynaklanacak şekilde modellendiği durumu ifade etmektedir.

Çok Boyutlu Veriler için MIMIC

Lee vd. (2016) tarafından MIMIC modelinin çok boyutlu bir uzantısını tanıtılmıştır. Bu model çok boyutlu yapılarda daha güvenilir DMF analizi yapabilmesi için önerilmiştir. Bu model, iki veya daha fazla gizil özelliği idare edebilmektedir. İki kategorili olarak puanlanan maddelerin η_1 ve η_2 olmak üzere iki gizil özelliği ölçtüğü varsayıldığında, MIMIC etkileşim modeline yönelik eşitlik Eşitlik 16'da verilmiştir:

$$y_i^* = \lambda_{1i}\eta_1 + \lambda_{2i}\eta_2 + \beta_i z + \omega_{1i}\eta_1 z + \omega_{2i}\eta_2 z + \varepsilon_i \quad (16)$$

y_i^* : i değişkeni gizil yanıtı
λ_{1i} : i maddesinin gizil özellik 1 (η_1) üzerindeki faktör yükleri
λ_{2i} : i maddesinin gizil özellik 2 (η_2) üzerindeki faktör yükleri
β_i : Eğim parametresi
z : Gruplama değişkeninin
ε_i : Tesadüfi hata

$\beta_i \neq 0$ olduğunda; β_i, gruplama değişkeninin i maddesi üzerindeki TB DMF etkisidir. ω_{1i} ve ω_2 ise $\omega_{1i} \neq 0$ ve / veya $\omega_{2i} \neq 0$ olduğunda; gruplama değişkeni ile iki gizil özellik arasındaki etkileşim nedeniyle i maddesi üzerindeki TBO DMF etkileridir.

Çok boyutlu MIMIC etkileşim modelinde, model tanımlamasını sağlamak için, iki gizil değişkenin iki değişkenli normal dağılıma sahip olduğu varsayılır.

Çok Boyutlu Veriler için MTK - LR

Suh ve Cho (2014) tarafından MTK-LR testinin ÇBMTK modelleri için genişletilmiş halidir. ÇBMTK modelleri için MTK-LR testi, TBMTK modelleri için MTK-LR testine benzer şekilde, iç içe (nested) modellerin model uyumunun karşılaştırılmasıyla değerlendirilmesine dayanmaktadır. ÇBMTK modelleri için MTK-LR testi bazı ek varsayımlar gerektirir. Pratikte iki gizil özelliğin her iki grupta sırasıyla ortalamaları 0 ve varyansları 1 olarak sabitlenir. Bu durumda birden çok gizil özellik arasında metrik belirsizlik sağlanır. Bununla birlikte; bu yöntemin uygulanabilmesi için boyutlar arası korelasyon sıfır olmalıdır.

Çok Boyutlu Veriler için Lojistik Regresyon

Mazor vd. (1998) tarafından eşleşen değişkenler olarak birden çok özelliği kapsamak için LR kullanılmıştır. Maddenin X_1 ve X_2 olmak üzere iki özellikle ilişkili olduğu varsayımında; Mazor vd. (1998) tarafından TB DMF belirlemede kullanılan Eşitlik 17'de verilen LR modeline ait eşitlik; iki özellik için Eşitlik 18'de gösterildiği şekilde genişletilmiştir:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 G \quad (17)$$

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 G. \quad (18)$$

Mazor vd. (1998) tarafından yapılan çalışmada sadece TB DMF için kullanımı düşünülmüştür. Bununla birlikte LR yaklaşımı, birden fazla özelliğin varlığında TBO DMF'yi tespit etmek için kullanımı Eşitlik 19'da verilmiştir:

$$z=\beta_0+\beta_1 X+\beta_2 G+\beta_3 XG \quad (19)$$

Eşitlik 19'da verilen TBO DMF'yi tespit etmek için kullanılan LR modelini takiben; iki özellikle (X_1 ve X_2) ilişkili bir madde için TBO DMF Eşitlik 20'de verilmiştir:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 G + \beta_4 X_1 G + \beta_5 X_2 G. \quad (20)$$

TB ve TBO DMF, LR'un tek boyutlu formundan aynı H_0 hipotezleri kullanılarak tanımlanabilmektedir. Eşitlik 18'de $\beta_3 \neq 0$ olduğunda diğer bir ifadeyle önemli bir grup etkisi varsa maddenin TB DMF gösterdiği ifade edilir. Eşitlik 20' de $\beta_4 \neq 0$ ve / veya $\beta_5 \neq 0$ olması, önemli bir grup ve toplam puan etkileşimi olduğunu göstermektedir. Bu durumda β_3'ün anlamlı olup olmadığına bakılmaksızın maddenin, TBO DMF gösterdiği belirtilir.

İlgili Araştırmalar

Bu çalışma kapsamında DMF belirleme teknikleri tek ve çok boyutlu kestirimler üzerinden ele alınmıştır. Bu bağlamda öncelikle genel olarak ÇBMTK ve DMF üzerine yapılan çalışmalara yer verilmektedir. Daha sonra çalışma ile yakından ilgili olan ÇBMTK ve DMF'nin birlikte yer aldığı çalışmalar özetlenmektedir. En son olarak, sırasıyla ÇBMTK, DMF ve ikisini birlikte inceleyen çalışmaların genel bir özeti ilgili araştırmalar özet başlığı altında verilmektedir.

Çok Boyutlu Madde Tepki Modeli ile İlgili Araştırmalar

Reckase vd. (1988) yaptıkları araştırmada gerçek ve benzetim yoluyla üretilmiş olmak üzere iki veri seti kullanmışlardır. Gerçek veri seti, 40 maddesi matematik kullanımı alt testinden ve 28 maddesi sosyal araştırmalar okuma testinden olmak üzere 2738 bireyden toplanan verileri içermektedir. Diğer veri seti ise 2 PL çok boyutlu lojistik modele göre, iki değişkenli standart normal dağılımdan seçkisiz seçilmiş, benzetim yoluyla üretilmiş 80 maddeden oluşan veridir. D değerleri ortalaması 0.00 ve standart sapması 0.60 olan normal dağılımdan seçkisiz olarak seçilmiştir. Maddeler 20'şer gruplara ayrılarak, her gruptaki maddelerin açısı sırasıyla 90^o, 45^o, 0^o olarak ve son grubun diğer bir ifadeyle son 20 maddenin açıları ise 0^o ile 90^o arasında eşit saçılıma sahip olacak şekilde ayarlanmıştır. Veri kümelerinin analizinde tek boyutlu analiz

programlarından olan LOGIST ve ÇBMTK analiz programlarından olan MIRTE kullanılmıştır. Programlardan elde edilen parametrelerle Q_3 istatistiği hesaplanmıştır. Daha sonra MIRTE sonuçlarına dayalı olarak maddeler kümelendikten sonra LOGIST ile madde kümeleri analiz edilmiş ve elde edilen parametrelerden tekrar Q_3 istatistiği hesaplanmıştır. Analizler sonucu kümelerin tek boyutluluk sayıltısını karşıladığı sonucuna ulaşılmıştır.

Zeng (1989) tarafından yapılan araştırmada TB modellerin ÇB'a karşı sağlamlığının belirlenmesi amaçlanmıştır. Veri seti olarak 50'şer maddelik ve 1000'er kişilik veri üretilmiştir. Veri üretiminde boyutların olasılığa vereceği katkı değişimlenmiştir. Karşılaştırmalarda MKE karşılaştırmaları ve parametre karşılaştırmaları kullanılmıştır. Araştırma sonucunda ana boyutun performansın %80'inden fazlasını açıkladığı durumlar için 1PL modelin, ÇB'a karşı oldukça dayanıklı olduğu; 2PL modelin ise 1PL modelden daha az dayanıklı olduğu bulunmuştur. Parametreler açısından incelendiğinde; a parametresinin boyutluluğa duyarlı olduğu bulunmuştur. Ayrıca araştırma sonucunda ana boyutun performansın en az %80'i açıkladığı durumlarda yetenek ve b parametresinin daha küçük hatalar ürettiği sonucuna ulaşılmıştır.

Spray vd. (1990) tarafından yapılan araştırmada telafisel olan ve telafisel olmayan ÇBMTK modellerinin madde ve test puanı üzerinden ölçme farklılıklarının karşılaştırılması amaçlanmıştır. Veri seti olarak benzetim koşuluyla 2000 kişilik 20 maddelik veri üretilmiştir. Araştırmanın sonucunda madde düzeyinde çok küçük ve anlamlı olmayan farklılaşmalar olmakla birlikte test puanı düzeyinde farklılaşmanın olmadığı görülmüştür. Buna ek olarak madde düzeyindeki farklılaşmaların gizil özellikler arasındaki korelasyon arttıkça azaldığı belirlenmiştir.

Luecht ve Miller (1992) tarafından yapılan araştırmada ÇB testlerin TB kalibrasyonu ve yorumlanması için yeni bir yöntem önerilmiştir. Bu amaç kapsamında 2000 kişi için 50 maddeden oluşan 2 boyutlu 3PLM ÇB normal ogive modele göre veri üretilmiştir. Referans açısı olarak 200 ve 700 belirlenmiştir. Bu açıların standart sapması 0, 5 ve 10 olacak şekilde değişimlenmiştir. ÇB maddeler ÇB uzaydaki yönlerine göre kümelenmiştir. Elde edilen her ayrı küme için TB kestirim yapılmıştır. Yetenek, güçlük ve ayırt edicilik için yapılan TB ve ÇB kestirim sonuçlarının incelenmesinde gerçek değerler dikkate alınmıştır. Araştırma sonucunda, önerilen bu yöntemle iki basamaklı kestirimin uyum indekslerinden taviz verilmeden kestirim dayanıklılığı ve yorumlanması açısından kazanç sağlanacağı sonucuna ulaşılmıştır.

Bolt ve Lall (2003) tarafından yapılan araştırmada Markov zinciri Monte Carlo (MCMC) tahmin algoritmasının ÇBMTK modellerinin parametrelerinin yeniden elde edilmesinde etkinliğininin incelenmesi; telafi edici olan ve olmayan ÇBMTK modellerinin nispi uyumunun karşılaştırılmasında bir Bayes faktör kriterinin faydasını göstermek amaçlanmıştır. Bu amaç kapsamında telafisel ÇB 2PLM ve telafisel olmayan ÇB gizil özellik modeli kullanılmıştır. Araştırmada hem gerçek hem de yapay veriler kullanılmış olup gerçek verileri Wisconsin Üniversitesi İngilizce Testi'ndeki 31 maddeye ait 6000 öğrencinin yanıtları oluşturmaktadır. Yapay veriler ise örneklem büyüklükleri (1000, 3000), madde sayısı (25, 50) ve gizil yetenekler arasındaki korelasyon (0, 0.3, 0.6) olacak şekilde üretilmiştir. Araştırmanın sonucunda gerçek veriler üzerinde yapılan incelemelerde MLTM'de parametre kestirimlerinin daha güç olduğu ve MLTM' ye göre M2PL'nin veriye daha iyi uyum sağladığı görülmüştür. Yapay veriler üzerinde yapılan incelemelerde ise gizil yetenekler arasındaki korelasyon arttıkça MLTM'de yapılan tahminlerin doğruluğunda düşüş olmakla birlikte her iki model için Metropolis-Hasting algotitması kullanılarak yapılan MCMC yönteminin parametreleri doğru tahmin ettiği belirtilmiştir. Araştırma sonucunda ayrıca; MCMC yöntemiyle M2PL modellerde parametrelerin doğruluğunun daha fazla olduğu sonucuna ulaşılmıştır.

Spencer (2004) tarafından benzetim yoluyla üretilmiş veri üzerinde yapılan araştırmada, TBMTK ve ÇBMTK modelleri altında model ve yetenek parametre kestirimlerinin karşılaştırması yapılmıştır. Araştırmada kapsamında ayrıca kullanılan yazılımların karşılaştırılmasına da yer verilmiştir. Araştırma sonucunda, yanıtlayıcıların ÇB gizil özelliklerinin, ÇB ölçme modeli ile TB ölçme modeline göre daha hassas ölçümlere ulaşıldığı belirlenmiştir.

De la Torre ve Patz (2005) tarafından yapılan araştırmada tek bir test içerisinde farklı boyutlardan gelen yeteneklerin eşzamanlı tahmininin hangi koşullar altında daha kesin yetenek tahminleri sağlayacağının araştırılması amaçlanmıştır. Bu amaç doğrultusunda veriler boyut sayısı (2,5), madde sayısı (10, 30, 50) ve yetenekler arasındaki korelasyon (0, 0.4, 0.7, 0.9) olacak şekilde değişimlenerek benzetim yoluyla elde edilmiştir. ÇB 3PLM ile parametre kestirimleri yapılmıştır. Araştırma sonucunda madde sayısı artışına bağlı olarak gizil yetenekler arasındaki ilişkinin arttığı bulunmuştur. Gizil yetenekler arasındaki korelasyon bire yaklaştığında TBMTK kullanımının; korelasyon sıfır olduğunda ise ÇBMTK kullanılarak yapılan eş zamanlı kestirimin daha iyi sonuçlar elde edileceği görülmüştür. Araştırmanın bir diğer sonucu eş zamanlı

kestirim yapmanın, parametre kestirimlerindeki yanlılığı ve standart hatayı azaltabileceğinin görülmesidir.

Levy (2006) tarafından yapılan araştırmada koşullu kovaryans teorisine göre Bayesian model kontrol prosedürlerini kullanarak MTK'de ÇB'nun incelenmesi amaçlanmıştır. Araştırma kapsamında boyutluluğu etkileyebilecek faktörler hipoteze edilerek bu faktörler incelemeye dâhil edilmiştir. Bu inceleme sonucunda; örneklem büyüklüğü, ÇB madde sayısı oranı, veri üretim modeli (telafisel ve telafisel olmayan modeller), boyutlar arası korelasyon ve boyut sayısı boyutluluğu etkileyen faktörler olarak belirlenmiştir.

Seungho Yang (2007) tarafından, TB ve ÇB modeller altında, tek boyutluluk varsayımının ihlalinin parametre tahmini ile model veri uyumuna olan etkisi incelenmiştir. Araştırma sonucunda, madde güçlük parametresi kestiriminde örneklem büyüklüğünün etkisi olduğu gözlenmiştir. Yetenek parametresi kestiriminde ise test uzunluğunun en büyük etkiye sahip olduğu görülmüştür. Diğer araştırmalardan farklı olarak ÇBMTK modellerinin, gizil özellikler arasındaki ilişki orta düzeyden yüksek düzeye kadar olan durumlarda, daha kesin kestirimlerde bulunduğu belirlenmiştir. Bununla birlikte gizil özellikler arası ilişki sıfıra yaklaştıkça TB Rasch modeli daha kesin kestirimlerde bulunduğu saptanmıştır.

Babcock (2009) tarafından yapılan araştırmada ÇBMTK modellerinin araştırılarak, alan yazında telafisel olmayan ÇBMTK modellerinin çok az uygulandığı görülmüştür. Bu bilgiden yola çıkılarak bu araştırmada yapay veri kullanılarak telafisel olmayan 2PL ÇBMTK modelinin uygulanabilirliğinin araştırılması amaçlanmıştır. Araştırmada tahmin yöntemi olarak Metropolis-Hastings algoritmasını kullanılmıştır. Araştırma sonucunda gizil özellikler arası ilişki düşük olmak koşuluyla her gizil özelliği temsil eden en az altı maddenin olması gerektiği belirlenmiştir. Buna ek olarak en az 4000 örneklem büyüklüğünün kabul edilebilir madde parametre tahminleri için gerekli olduğu görülmüştür. Sonuç olarak 2PL telafisel olmayan ÇBMTK modelinin kabul edilebilir tahminlerde bulunabilmesi için karşılanması güç varsayımlarının olduğunu ortaya konmuştur.

Çakıcı Eser (2015) tarafından yapılan araştırmada madde ve birey parametre değerlerinin TBMTK ve ÇBMTK modelleri kullanılarak daha kararlı kestirim yapabilmek için gerekli olan test uzunluğunun ve örneklem büyüklüğünün belirlenmesini amaçlamıştır. Kestirimlerin kararlılığının incelenmesinde yanlılık, RMSE ve kestirilen parametrelerle gerçek parametreler arasındaki korelasyondan yararlanılmıştır.

Araştırmada benzetim yoluyla üretilen iki kategorili maddeler ve yetenek değerlerinden elde edilen veriler kullanılmıştır. Araştırma sonucunda hem madde sayısı hem de örneklem büyüklüğündeki değişimden madde parametrelerine ait RMSE değerlerinin ve korelasyonların etkilendiği belirlenmiştir. Birey parametresi kestirimleri ise madde sayısındaki artıştan olumlu olarak etkilenip örneklem büyüklüğündeki değişkliklerden etkilenmemiştir.

Gül (2015) tarafından yapılan araştırmada iki kategorili (1-0) ÇB yapıların, farklı örneklem büyüklüğü, boyutlararası korelasyon ve boyut sayısı koşulları altında, TBMTK ve ÇBMTK modellerine dayalı olarak kestirilen birey ve madde parametre değerlerinin karşılaştırılması amaçlanmıştır. Birey ve madde parametre değerleri; korelasyon katsayıları, RMS ile değerlendirilmiştir. Araştırmada benzetim yoluyla iki, üç ve beş boyutlu basit yapıda iki kategorili ve 3PL telafisel modele uygun veriler üretilmiştir. Verilerin analizinde iki ve üç boyutlu veriler için BA-EM; beş boyutlu veri yapısı için ise MH-RM yöntemi kullanılmıştır. Araştırma sonucunda ÇBMTK'nın boyut sayısının ikiden fazla ve örneklem büyüklüğünün 1000'den fazla olması durumunda; TBMTK'nın ise boyut sayısının az ve örneklem büyüklüğünün yüksek olması durumunda daha iyi kestirimler yaptığı ve birey parametre değerlerini daha iyi kestirdiği görülmüştür. ÇBMTK'da daha az hatalı kestirimler yapılmak üzere TBMTK ve ÇBMTK ile yapılan analizlerde boyutların ilişki düzeyinin artmasının her iki modele dayalı olarak yapılan kestirimleri iyileştirdiği sonucuna ulaşılmıştır.

DeMars (2017) tarafından yapılan araştırmada iki kısmi telafi edici modeli, parametrelerin doğru kestirilmesi bakımından değerlendirmek ve kısmi telafi edici model için oluşturulan verilerin kısmi telafisel olan ya da telafisel modellerden hangisine daha uygun olduğunun belirlenmesi amaçlanmıştır. Araştırma verileri kısmi telafisel modellerden 'ürün' ve 'etkileşim' modellerinin 2PL ve 3PL türlerine uygun olacak şekilde üretilmiştir. Ayrıca veriler boyut sayısı (iki, dört) ve boyutlar arasındaki korelasyonlar değişimlenerek üretilmiştir. Bu araştırma sonucunda ürün modeli madde parametreleri hariç, kısmi telafi edici modellerin, makul bir doğrulukla tahmin edilebileceği görülmüştür.

Şahin (2012), Şahin ve Anıl (2017) yaptıkları çalışmalarda model veri uyumu, madde parametreleri ve standart hata kestirimleri üzerinde test uzunluğunun ve örneklem büyüklüğünün etkisini incelemişlerdir. Bu kapsamda 10, 20 ve 30 maddelik testler ve 150, 250, 350, 500, 750, 1000, 2000, 3000 ve 5000 örneklem büyüklükleri kullanmışlardır. Araştırmanın sonucunda, 1PL modelde 10, 20 ve 30 maddelik testler için

150 kişilik örneklemin yeterli olabileceği belirtilmiştir. Bununla birlikte 2PL modelde; 30 maddelik bir test için en az 250; 20 maddelik bir test için en az 500 ve 10 maddelik bir test için en az 750 kişilik örneklemlerin yeterli olabileceği sonucuna ulaşılmıştır. Buna ek olarak 3PL modelde ise 30 maddelik bir test için en az 350; 20 ve 30 maddelik bir test için ise en az 750 kişilik örneklemlerin yeterli olabileceği saptanmıştır.

Keçeoğlu (2018) tarafından yapılan araştırmada farklı boyutluluk ve örneklem özelliklerinde MTK uygulamalarına göre kestirilen parametrelerin değişmezliğinin incelenmesi amaçlanmıştır. Veri seti olarak Temel Eğitimden Ortaöğretime Geçiş (TEOG) uygulamasından elde edilen gerçek veriler kullanılmıştır. Araştırma kapsamında; TB parametrik modellerden 2PL, parametrik olmayan modellerden monoton homojenlik modeli; ÇBMTK için ise genişletilmiş 2PL model kullanılmıştır. Araştırmanın sonunda madde parametresi değişmezliğinin sağlanması için gereken örneklem büyüklüğünün, TB parametrik model için TB parametrik olmayan modele göre daha fazla olduğu görülmüştür. Test yapay olarak ÇB hale getirilerek ÇBMTK uygulaması yapılmıştır. ÇB testin parametrik ve parametrik olmayan ve ÇBMTK'ya göre analiz edilmesiyle elde edilen madde parametreleri için değişmezliğin sağlanamadığı belirlenmiştir. Örneklem büyüklüğü ve korelasyon türüne göre değişmeksizin, ÇB teste ilişkin, TB parametrik MTK ve ÇBMTK'den elde edilen yetenek parametreleri arasında manidar ve mükemmele yakın ilişki düzeylerinin olduğu görülmüştür. Buna ek olarak ÇBMTK ile TB ve ÇB testlere ait parametrik ve parametrik olmayan MTK'dan elde edilen güvenirlik katsayılarının evren değerine yakın büyüklükte olduğu belirtilmiştir.

Değişen Madde Fonksiyon ile İlgili Araştırmalar

Atar (2007) hem ikili puanlana (dichotomous) hem de çoklu puanlanan (polytomous) soruları içeren eğitim testlerindeki DMF'yi kapsamlı bir şekilde incelemiştir. DMF'nin çeşitli test koşullarında nasıl tezahür ettiğini anlamak için MTK-OO Testi, LR ve Genelleştirilmiş Lineer Gizil ve Karışık Modelleme (GLLAMM) gibi çeşitli istatistiksel yöntemler kullanmıştır. Bu çalışmada hem simülasyona dayalı veriler hem de gerçek veri (Florida Kapsamlı Değerlendirme Testi _FCAT) kullanılmıştır. Çalışmada örneklem büyüklüğü, referans ve odak grup örneklem büyüklüğü oranları, DMF büyüklüğü gibi faktörler değişimlenerek madde parametre kestirimi üzerindeki etkileri incelenmiştir. Çalışma sonucunda örneklem büyüklüğünün madde parametre

kestiriminde etkili olduğu bulunmuştur. DMF belirleme yöntemlerinin performanslarına yönelik olarak Tip I hata ve istatistiksel güç oranları incelenmiş, GLLAMM yönteminin farklı test koşullarında madde parametre kestiriminde diğer yöntemlerden daha etkili olduğu bulunmuştur.

Kim ve Jang (2009) Ontario Secondary School Literacy Testi'nde (OSSLT-Mart 2006) DMF incelenmiştir. Araştırmada İngilizce'nin alt öğrenme alanları ayrı ayrı boyutlar olarak ele alınarak dil ve kültürel özellikler bakımından öğrenciler arasında farklılaşma olup olmadığının belirlenmesi amaçlanmıştır. Araştırmada kelime bilgisine sahip bir gruptaki öğrencilerinin avantaja sahip oldukları fakat gramer, okuma ve yazma becerilerinde ise diğer gruptaki öğrencilerin daha iyi olduğu sonucuna ulaşılmıştır.

Furlow vd. (2009) DMF'ye neden olacak şekilde referans grubu odak gruptan çok daha üstün yetenek düzeyi parametresine sahip olarak tasarlayarak Tip I hata ve istatistiksel güç oranları incelemiştir. Araştırma kapsamında DMF incelenmesinde SIBTEST kullanılmış; analizler sonucunda DMF etki büyüklüğü, madde öbek oranı yüzdesi, örneklem büyüklüğü arttıkça istatistiksel güç oranının arttığı gözlenmiştir. Bununla birlikte tüm koşullarda Tip I hatanın beklenen değere yakın olduğu belirlenmiştir.

Huang (2010) PISA, PIRLS ve TIMSS uluslararası büyük ölçekli sınavlarında maddelerde gözlenen DMF'nin dil, kültür ve eğitim müfredatı değişkenlerinin hangisinden kaynaklandığını, TB ve ÇB DMF belirleme yöntemleriyle incelenmesini amaçlamıştır. Araştırma sonucunda ÇB testlerde klasik yöntemlerin çok fazla DMF gösteren madde tespit ettiği belirlenmiştir. Bununla birlikte ÇB Rasch tabanlı çok düzeyli lojistik modeli kullanarak daha az sayıda ve daha az hata ile DMF gösteren madde tespiti yapıldığı sonucuna ulaşılmıştır.

ÇBMTK ve Değişen Madde Fonksiyonun Birlikte Ele Alındığı İlgili Araştırmalar

Bolt ve Stout (1996) DMF belirlemede SIBTEST yönteminin ÇB testlerde nasıl kullanıldığını göstermeyi amaçlamışlardır. Bu yöntemle DMF'ye neden olan durumların da saptanabileceğini öne sürmüşlerdir. Çok kategorili gerçek ve üretilmiş veri setleri kullanılan bu araştırma sonucunda, bir veri setinin içerdiği DMF miktarını artan ve kontrol edilen Tip I hata ile belirleyebileceği belirtilmiştir.

Stout vd. (1997) iki boyutlu olarak hazırlanan bir testte DMF belirlemede kullanılan MULTISIB (Shealy ve Stout, 1993) yöntemini incelemeyi amaçlamışlardır. Çalışma sonucunda DMF'ye sebep olan ikincil boyutun MULTISIB tekniği ile en aza indirgendiği, Tip I hata ve istatistiksel güç oranlarının iyi sonuçlar verdiği belirtilmiştir. Bununla birlikte ÇB testlerde toplam test puan yerine boyutlardan elde edilen puanların alt boyut puanı olarak ele alınmasının daha uygun olacağı sonucuna ulaşılmıştır.

Liaw (2015) başlangıç yeteneğin ölçülmesi sürecinde, TB ve ÇB modellerin DMF belirleme durumlarının karşılaştırmasını amaçlamıştır. Boyutlar arası korelasyon, referans ve odak grup oranları, başlangıç yetenek etki düzeyleri, ikincil yetenek etki düzeyleri ve DMF içeren madde oranları değişimlenmiştir. Araştırma sonucunda yanlılık bakımından en kötü sonuçları TB model vermiştir. ÇB modellerle ise DMF gösteren maddeler kaldırılmadan da uygun sonuçlar elde edilebileceği sonucuna ulaşılmıştır.

Bulut vd. (2017) ÇB testlerde; DMF etki büyüklüğü ve biçimi, test uzunluğu, boyutlar arası korelasyon, örneklem büyüklüğü, odak referans grupları arasındaki ortalama farklılıkları koşullarında MIMIC kullanarak Tip I hata ve istatistiksel güç oranlarının nasıl değiştiğinin belirlenmesini amaçlamışlardır. Araştırma sonucunda örneklem büyüklüğü ve ortak madde sayısı arttıkça istatistiksel güç oranının ve TBDMF'yi belirleme gücünün arttığı belirtilmiştir. Bununla birlikte odak ve referans grup ortalamalarının birbirine yakın olmasının da gücü arttırdığı sonucuna ulaşılmıştır. Sonuç olarak Multi-MIMIC yöntemi TB DMF olan maddelerde güvenilir sonuçlar verebiliyor iken TBO maddeler için kullanılmasının yanıltıcı sonuçlar verebileceği belirtilmiştir.

Bulut ve Suh (2017) araştırmalarında farklı koşullar altında iki ya da daha fazla örtük özelliğe sahip ÇB testlerde DMF belirlenmesinde kullanılan MIMIC, MTK-OO ve LR yöntemleri ve bu yöntemlerden elde edilen sonuçlar hakkında bilgi vermeyi amaçlamıştır. Araştırmada TBO DMF için LR'nun kullanılabildiği sonucuna ulaşılmıştır. Bununla birlikte TB DMF için de kısa testlerde LR ve MIMIC'in, MTK- OO'dan daha iyi sonuçlar verdiği; uzun testlerde ise sonuçların birbirine yakın olduğu görülmüştür. Tip I hata için LR, istatistiksel güç için ise MTK-OO yöntemi ile daha yüksek ve dengeli sonuçlar elde edilmiştir.

İlgili Çalışmaların Genel Değerlendirmesi

ÇBMTK kapsamında yapılan çalışmalar incelendiğinde, alan yazında ağırlıklı olarak KTK ve MTK uygulamalarından elde edilen parametre kestirimlerinin karşılaştırıldığı, MTK'nin özellikle parametrik uygulamalarından elde edilen parametre kestirimlerinin ve model uyumlarının karşılaştırıldığı araştırmaların olduğu belirlenmiştir (Kart Arslan, 2019; Luecht, 2003; Spencer, 2004). Birçok araştırmada da (Anderson, 1999; MacDonald ve Paunonen, 2002; Rogers ve Ndalichako, 2000) ikili puanlanan maddelerden oluşan testlerde; test geliştirmede ve değerlendirmede KTK ile TBMTK karşılaştırılmıştır. Bununla birlikte ÇBMTK'yı konu edinen; MTK ve ÇBMTK'nın birlikte ele alındığı; kümeleme analizi ile ÇBMTK'nın sonuçlarının, parametrik ve parametrik olmayan yöntemlerinin, TBMTK ve ÇBMTK modellerinin karşılaştırıldığı çalışmalar da yer almaktadır. Yapılan parametre kestirim araştırmalarında birey ve madde parametrelerinin kestirim kararlığı örneklem büyüklüğü, test uzunluğu, madde oranı, boyut sayısı, boyutlar arası korelasyon, yetenek dağılımı gibi özelliklerin manipüle edilmesi ile belirlenmeye çalışılmıştır. Çalışmalar incelendiğinde genel olarak boyutlar arası korelasyonun 0.00-0.95 aralığında, boyut sayısı olarak bir ve iki boyut çok az da olsa üç ve üstü boyutun alındığı, 2PL modelin kullanıldığı, örneklem büyüklüğünün 500-6000 aralığında ve test uzunluğunun 10 ile 80 madde aralığında alındığı görülmüştür. Parametre değişmezliğinin göstergesi olarak ise standart hata ortalamaları, yanlılık, RMSE değerleri, hata varyansları, kestirilen parametrelerle gerçek parametreler arasındaki korelasyondan yararlanılmış; temel boyutluluk incelemesinde DIMTEST kullanılmıştır. Yapılan araştırmalarda hem gerçek veri hem de simülatif veri üzerinde araştırmalar olmakla birlikte yöntem karşılaştırmalarının simülatif veri ile gerçekleştirildiği görülmektedir.

DMF ile ilgili yapılan çalışmalar incelendiğinde, DMF incelemelerinin hem gerçek veriler hem de simülasyon verileri üzerinde gerçekleştirildiği görülmüştür. Gerçek veriler için genellikle geniş ölçekli testlerin kullanıldığı çalışmalar (Ayan, 2011; Erdem, 2015; Kim ve Jang, 2009) olduğu görülmektedir. Simülasyon çalışmalarında ise farklı koşullarda DMF belirleme tekniklerinin incelendiği ve farklı tekniklerin birbirleri ile karşılaştırıldığı çalışmaların (Ayva Yörü, 2019; Bolt ve Stout, 1996; Chen ve Jin, 2018; Erdem Keklik, 2012; Furlow vd., 2009; Walker ve Göçer Şahin, 2016; Yıldıztekin, 2020) yer aldığı görülmektedir. Simülasyon çalışmaları kapsamında, gerçek test uygulamalarında karşılaşılabilecek durumlar (örneklem büyüklüğü, örneklem oranı,

gruplara ait yetenek farklılıkları, boyutlar arası korelasyon, DMF büyüklüğü, DMF türü, test uzunluğu vb.) dikkate alınarak farklı koşullar incelenmiş; koşullara göre DMF'nin nasıl değiştiği istatistiksel güç oranı, Tip I hata, yanlılık, RMSEA gibi bazı ölçütler dikkate alınarak yorumlanmıştır. Bunlara ek olarak DMF'nin test eşitlemeye etkisinin incelendiği ve test eşitlenmesinde kullanılan yöntemlerin karşılaştırıldığı (Atalay Kabasakal, 2014); DMF belirleme tekniklerinin farklı puanlama durumlarında incelendiği (Selvi, 2013) ve DMÖ Fonksiyonun incelendiği çalışmalar (Douglas vd., 1996; Vandenberghe ve Gierl, 2001) olduğu görülmektedir.

ÇBMTK ile DMF yöntemlerinin birlikte kullanıldığı çalışmalar incelendiğinde, ÇB testlerde, farklı koşulların DMF belirleme süreci üzerindeki etkilerinin DMF teknikleri ile belirlenmesi (Bulut vd., 2016; Walker ve Göçer Şahin, 2016), ÇB testlerde DMF belirleme yöntemlerinin kullanılması (Bolt ve Stout, 1996; Bulut ve Suh, 2017; Oshima vd., 1997; Stout vd., 1997; Walker vd., 2008) ve TB-ÇB modellerin DMF belirleme durumlarının karşılaştırılması (Liaw, 2015) çalışmalarının yer aldığı görülmektedir. DMF ile ilgili çok sayıda çalışma yer almasına rağmen, DMF'nin bu çalışma kapsamındaki şekilde boyutluluk etkisi üzerinden ele alındığı bir çalışmaya rastlanmamış; sadece bir çalışmada (Liaw, 2015), DMF ile boyutluluk arasındaki ilişki incelenmiş; buna karşın bu çalışmada iddia edildiği ve ele alındığı koşullarda yapılan bir çalışmaya rastlanmamıştır. DMF incelemelerinde yöntem olarak genellikle SIBTEST, MULTISIB, DFIT, MULTI-DFIT, MH, LR; MTK-OO, MIMIC ve Multi-MIMIC modellerinin tercih edildiği görülmektedir. Yapay veri setlerine ait sonuç incelemelerinde ise sıklıkla Tip I hata, istatistiksel güç, yanlılık ve RMSEA değerlerinin kıyaslandığı görülmüştür. Çalışmalarda veri setleri genellikle iki boyutlu olarak tasarlanmış olup üç boyutlu veri setleri ile oldukça sınırlı sayıda çalışma yer almaktadır. Boyutluluk ile DMF biçimi ve DMF büyüklüğü arasında ilişkiyi çalışmada ele alındığı şekilde inceleyen ve ele alınan iddianın gerçek veri seti ile desteklendiği bir çalışmaya rastlanmamıştır. Bu çalışmada farklı koşullarda DMF nedeniyle TB'u sağlanmayan testlerde ÇB yetenek kestiriminden elde edilen veri setlerine yönelik yapılan DMF analizlerinin Tip I hata ve istatistiksel güç oranları üzerindeki etkisi incelenmiştir. Yapılan simülatif çalışma gerçek veri seti ile de incelenerek desteklenmiştir.

BÖLÜM 3

YÖNTEM

Bu bölümde çalışmanın modeli, simülasyon deseni, tez kapsamında istenilen koşullarda veri setlerinin üretilmesi, simülasyona ait geçerlik çalışmaları, verilerin analizi ve gerçek verilere ilişkin bilgiler yer almaktadır.

Araştırmanın Modeli

Bu çalışmada, simülasyon deseninde dikkate alınan koşullar dahilinde veri üretimi gerçekleştirilmiştir. Elde edilen veri setlerine TB ve ÇB MTK kestirimleri yapılmıştır. Bu çalışmada, ÇBMTK modellerinin birincil yeteneği tahmin ederken, TBMTK'ya ait DMF'yi kontrol etmede etkili olup olmayacağı farklı simülasyon koşullarında incelenmiştir. Bu yönüyle çalışma simülatif deneysel araştırma niteliğindedir (Morris vd., 2017). Buna ek olarak simülasyon çalışmaları alanyazındaki kuramsal bilgi birikimine katkı sağlayarak kuramların gelişimine önemli katkılarda bulunabilmesi özelliği nedeniyle teknikolojik yaklaşım olarak da kabul edilmektedir (Happach ve Tilebein, 2015; Harwell vd., 2017). Hoaglin ve Andrews (1975) tarafından da istatistiksel örnekleme deneyleri olarak ele alınması gerektiği savunulmaktadır. Bu kapsamda bu çalışmada DMF gösteren maddeler nedeniyle TB'nun sağlanamadığı testlerde, TBMTK ve ÇBMTK kestirtimlerinin DMF'yi kontrol altına almaya yönelik etkilerinin farklı simülasyon koşulları altında incelenmesi amaçlanmıştır. Yapılan bu çalışmanın alanyazındaki kuramsal bilgi birikimine katkı sağladığı; dolayısıyla temel araştırma niteliğindeki bir simülasyon çalışması olduğu söylenebilmektedir.

Simülasyon Deseni

Birincil yetenek tahmininin doğruluğu ve kesinliği ile ilişkili değişimlenen çeşitli koşulların etkilerini araştırmak için, bu çalışmada Monte Carlo yaklaşımı kullanılmıştır. Monte Carlo yaklaşımında parametre tahminlerinin dağılımını elde etmek, tek bir veri setinden mantıklı olmayan sonuçlar elde etme şansını azaltmak ve gerçek parametreleri

yeniden örnekleyebilmek için birden fazla tekrar (replikasyon) yapılmaktadır. Bu çalışma kapsamında belirtilen simülasyon koşullarına göre oluşturulan her veri seti için 100 tekrar yapılmıştır. Simülasyon desenine ait sabit tutulan ve değişimlenen koşullar belirlenirken literatürdeki çalışmalardan yararlanılmıştır. Çalışmada kullanılan tüm deneysel koşullar Tablo 1'de sunulmuştur.

Tablo 1

Simülasyon Deseni 1

	Koşullar	Düzeyler	Düzey Sayısı
Sabit Tutulan Koşullar	MTK Modeli	TB 2PL	1
	Yetenek (θ)	-3≤ θ≤3	1
	Madde Güçlüğü (b)	0≤b≤1	1
	Madde Ayırt Ediciliği (a)	1≤a≤2	1
	Madde Sayısı (k)	60	1
	Örneklem Büyüklüğü (n)	1500	1
Değişimlenen Koşullar	Boyut Sayısı (k)	2, 3	2
	DMF Yüzdesi	%10, %20, %30	3
	DMF Büyüklüğü	.50, .75	2
	DMF Türü	TB, TBO	2
Toplam Koşul Sayısı			2x3x2x2=24
Tekrar Sayısı			100
Toplam Veri Seti			24x100=2400

Tablo 1 incelendiğinde çalışma kapsamında veriler üretilirken farklı koşulların dikkate alındığı görülmektedir. Bu koşullar aşağıda açıklanmaktadır. Simülasyona dayalı üretilen verilerin doğası gereği bazı varsayımlar yapılmaktadır. Çalışma sonuçlarının kalitesinde, bu varsayımların geçerliği etkili olmaktadır.

Sabit Tutulan Koşullar

Bu çalışma kapsamında çok boyutlu analizler için gerekli koşullar da dikkate alınarak örneklem büyüklüğü, tepki kategori sayısı, toplam madde sayısı, madde ayırıcılığı, madde güçlüğü, boyutlar arası korelâsyonlar, veri üretiminde kullanılan teknik, veri üretiminde kullanılan model ve yetenek parametreleri sabit tutulmuştur. Bu koşulların sabitlenmesinde etkili olan koşullar aşağıda başlıklar halinde verilmiştir.

Örneklem Büyüklüğü. Karıştırıcı etkilerden kaçınma ilkesine dayalı olarak örneklem büyüklüğü sabit tutulmuştur. Örneklem büyüklüğünün DMF tespitinde Tip I

hata ve istatistiksel güç oranları üzerinde bir etkisi bulunmaktadır. Yeterli DMF saptama gücü için minimum örneklem büyüklüğü referans ve odak grupta en az 200 (Johnson Frotman, 2007; Li vd., 2012; Wyse ve Mapuranga, 2009); orta büyüklükteki örneklem büyüklüğü için 400, 500 ile 600 ve büyük örneklemlerde ise 750, 800, 1000, 1200, 2000 ile 3000 kişi olarak belirlendiği görülmüştür (Atar, 2007; Atar ve Kamata, 2011; Bolt, 2002; Carter, 2011; Duncan, 2006; Finch ve French, 2008; French ve Maller, 2007; Jodoin ve Gierl, 2001; Johnson Frotman, 2007; Kristjansson vd., 2005; Li vd., 2012; Wang vd., 2013; Wyse ve Mapuranga, 2009). Büyük örneklemlerdeki durumun incelenmesi amacıyla odak ve referans grup, önceki çalışmalarda (Finch ve French, 2007; Köse, 2010; Lee ve Zhang, 2017; Walker vd., 2006) olduğu gibi gruplar dengeli olacak şekilde 750 birey referans ve 750 birey odak grup olmak üzere toplam 1500 birey olarak sabitlenmiştir.

Tepki Kategori Sayısı. Başarı testlerinde olduğu gibi eğitim alanında kullanılan test maddelerinin çoğunlukla çoktan seçmeli formatta yapılandırıldığı görülmektedir. Bu tür testler de genel olarak iki kategorili olarak puanlanmaktadır (Potenza ve Dorans, 1995). Bu kapsamda bu araştırmada tepki kategori sayısı iki (1-0) olarak sabitlenmiştir.

Toplam Madde Sayısı. Toplam madde sayısının sabitlenmesinde; DeMars ve Lau (2011) tarafından yapılan çalışmada elde edilen, test uzunluğunun kişi ve madde parametrelerinin kurtarılmasını ve DMF'nin boyutunu ciddi şekilde etkilemediği sonucu etkili olmuştur. Araştırmada sabitlenecek toplam madde sayısının belirlenmesinde alanyazın incelenmiş; test uzunluğunun 10 ile 80 madde aralığında (De la Torre ve Patz, 2005; Kelecioğlu, 2001; Köse, 2010; Özer-Özkan, 2012; Spray vd., 1990; Zeng, 1989) alındığı tespit edilmiştir. Bu çalışmada çok boyutlu yapılara ait incelemelerin yapıldığı çalışmalar (Ansley ve Forsyth, 1985; Duncan, 2016; Finch ve French, 2007) da olduğu gibi toplam madde sayısı 60 olarak belirlenmiştir.

Madde Ayırıcılığı. MTK modellerinde madde ayırt edicilik indeksinin (a) normal aralığı 0.00-2.00 olarak (Hambleton vd., 1991); yüksek ayırt edicilik aralığı ise 1.35 ile 1.69 olarak (Hambleton, 1989) tanımlanmıştır. Bu çalışmada madde ayırıcılığı TB ve TBO verilerin üretiminde; tek biçimli (uniform) dağılımda minimum 1.00, maksimum 2.00) olacak şekilde üretilmiştir.

Madde Güçlüğü. Araştırma kapsamında simülasyona dayalı olarak üretilen verilerde madde güçlüklerinin gerçek alan uygulamalarını yansıtması beklenmektedir. Bu bakımdan madde güçlüğünün genellikle orta düzeyde olması, zor ve kolay olan maddelerin ise az olması istenmektedir. Bu nedenle; madde güçlükleri (b), tek biçimli (uniform) dağılımda minimum 0.00, maksimum 1.00 olacak şekilde üretilmiştir.

Birey Yetenek Parametreleri. Yetenek parametresi, teorik olarak -∞ ve +∞ sonsuz arasında değerler alsa da, bireylerin çoğunun θ değeri -3 ve +3 arasında değiştmekte; sıklıkla θ, ortalaması 0, standart sapması 1 olacak şekilde kestirilmektedir (DeMars, 2016). Bu çalışmada testi alanların θ_1 ve θ_2 için (gerçek) yetenek puanları iki değişkenli normal dağılımdan alınmıştır. θ_1 ve θ_2 dağılımı, odak ve referans grup için, ortalama 0 ve standart sapma 1 olacak şekilde sabitlenmiştir.

Boyutlar Arası Korelâsyon. Alan yazında boyutlar arası korelâsyon için düşük korelâsyondan (0.30) yüksek korelâsyona (0.95) kadar çok çeşitli değerler kullanılmıştır (Ansley ve Forsyth, 1985; Reckase vd., 1988; Batley ve Boss, 1993; Tate, 2003; Flora ve Curran, 2004; Finch, 2010; Liaw, 2015; Soysal, 2017; Yıldıztekin, 2020; Zhang, 2012). Bu çalışmada çok boyutlu çalışmalarda (Kahraman Capar, 2001; Zhang, 2012) olduğu gibi boyutlar arası korelâsyon 0.5 olarak sabitlenmiştir.

Veri Üretiminde Kullanılan Teknik. Veri üretiminde, önceki simülasyon çalışmalarında (Cohen vd., 1996; Li vd., 2012) kullanılan Monte Carlo simülasyonu kullanılmıştır. Monte Carlo simülasyon çalışmaları, eğitim alanında özellikle ölçme ve psikometri alanlarında yapılan araştırmalarda önemli bir rol oynamakta; bu çalışmalarda genellikle MTK modeli incelenmektedir (Harwell, 1997; Harwell vd., 1996). Simülasyona dayalı MTK çalışmalarında ise ücretsiz ve açık kaynaklı programlama dili olan R programlama dilinin kullanımının son on beş yıldır arttığı gözlenmektedir. Bu durumda, madde parametrelerini kestirme, madde yanlılık analizleri gibi birçok MTK temelli analizin R programlama dilinde yer alan paketlerle yapılabilmesinin etkili olduğu düşünülmektedir. Monte Carlo simülasyon çalışmalarında, gerçek veri kümelerinin aksine gerçek yetenek ve gerçek boyutluluk yapısı önceden bilinmektedir. Bu nedenle tahminler bilinen parametrelerle karşılaştırılabilmektedir. Bununla birlikte bu teknikte deneysel durumlar için gerçekçi veri koşulları modellenebilmektedir.

Veri Üretiminde Kullanılan Model. DMF gösteren maddeler, ayırt edicilik (a) ve güçlük (b) parametrelerinde farklılık göstermektedir (Finch ve French, 2007). Hambleton vd. (1991) tarafından yapılan çalışmada, DMF belirlemede 3PL modelin kullanılmasıyla güçlü bir test istatistiği elde edilemediği belirtilmiştir. Bu çalışmalar ve ilgili benzer çalışmalar (Cohen vd., 1996; Finch ve French, 2007; Köse, 2010; 2012; Lee ve Zhang, 2017; Liaw, 2015; Li vd., Oshima vd., 1997; Özdemir, 2015; Şahin, 2012) dikkate alınarak bu çalışmada, odak ve referans gruplara ilişkin madde yanıtları ile madde ve yetenek parametreleri 2PL modele uygun olarak üretilmiştir.

Değişimleme Yapılan Koşullar

Bu çalışma kapsamında modelde dikkate alınan boyut sayısı, DMF gösteren madde oranı, DMF türü ve DMF büyüklüğü değişimlenmiştir. Bu koşulların değişimlenmesinde etkili olan bazı gerekçeler bulunmaktadır. Bunlar aşağıda başlıklar halinde verilmiştir.

Boyut Sayısı. Boyut sayısını doğru belirlemek, gerçek veri analizinde her zaman basit ve kolay değildir (Reckase, 2009). Performansla ilgili boyut sayısını araştırmanın önemli olacağı düşünülmektedir. Alanyazın incelendiğinde boyut sayısı olarak bir ve iki boyut çok az da olsa üç ve üstü boyutun alındığı (Ackerman, 1989; Ansley ve Forsyth, 1985; Doğan, 2002; de la Torre ve Patz, 2005; Luecht ve Miller, 1992; Soysal, 2017) görülm. Alanyazın dikkate alınarak modelde ele alınan boyut sayısı iki ve üç olarak değişimlenmiştir (de la Torre ve Patz, 2005; Yıldıztekin, 2020).

DMF Gösteren Madde Oranı. Alanyazın incelendiğinde DMF gösteren madde oranının yetenek parametre kestirimi üzerinde etkili olduğu (Walker vd., 2012; Wells vd., 2002; Zumbo, 2003) simülasyona dayalı olarak üretilen veri setlerinde 0 ile % 40 oranları arasında DMF gösteren madde oranlarının kullanıldığı görülmektedir (Fidalgo vd., 2000; Holms Finch ve French, 2007; Meade ve Lautenschlager, 2004). Bu çalışmanın veri üretiminde, testin geçerliğini olumsuz yönde etkilediği bilinen DMF gösteren maddelerin oranı önceki çalışmalarda (Atar, 2007; Erdem Kara, 2019; Gierl vd., 2000; Jodoin ve Gierl, 2001; Kabasakal vd., 2014; Liaw, 2015;) kullanılan %10, %20 ve %30 olarak

değişimlenmiştir. 60 maddelik veri setinde sırasıyla 6, 12 ve 18 madde DMF içerecek şekilde değişimlenmiştir.

DMF Türü. Alanyazında simülasyona dayalı yapılan çalışmalarda, DMF türünün çalışma kapsamında etkili olabilecek bir değişken olarak değişimlendiği görülmektedir. Bu çalışmalarda maddelerin yalnızca ayırıcılık ve yalnızca güçlük parametrelerine odak grup için eklemelerin yapıldığı Tek Biçimli (TB) olan ve Tek Biçimli Olmayan (TBO) DMF türleri ele alınmıştır (Cappaert, 2014; Dainis, 2008; Flowers vd., 2002; Golia, 2014; González-Romá vd., 2006; İnal, 2018; Oshima vd. 1997; Stark vd., 2006). ÇBMTK kullanımıyla madde ve yetenek parametrelerinde yanlılık düzeltmesi sağlanmasında DMF türünün (TB ve TBO) etkili olabileceği düşünülmektedir. Bu kapsamda TB ya da TBO-DMF'de nasıl düzeltme sağlayacağı incelenmiştir.

DMF Büyüklüğü. Alanyazındaki çalışmalar DMF büyüklüğünün yetenek kestirimi üzerinde etkili olduğunu göstermiştir (González-Romá vd., 2006; Stark vd., 2006; Zumbo, 2003). Bununla birlikte simülasyona dayalı çalışmalarda 0.10 ile 0.75 arasında değerlerin kullanıldığı görülmektedir (Bilican Demir, 2014; Cao vd., 2017; Erdem Keklik, 2014; Fidalgo vd., 2000; González-Romá vd., 2006; Kabasak vd., 2014). Bu kapsamda bu çalışmada veri üretiminde DMF miktarı, alanyazın (Atar ve Kamata, 2011; Stark vd., 2006; Tunç, 2016) dikkate alınarak 0.50 (orta) ve 0.75 (yüksek) olarak değişimlenmiştir. ÇBMTK kullanımıyla madde ve yetenek parametrelerinde yanlılık düzeltmesinde orta ya da yüksek DMF'den hangisinde daha iyi düzeltme sağlayacağının incelenmesi önemli görülmektedir. Çalışma kapsamında veriler üretilirken belirlenen koşullara uygun veri üretmek için; R Programı (R Core Team, 2019) kullanılmıştır.

Verilerin Üretimi

Veri üretiminde R içerisinde yer alan "mirt" (Chalmers, 2012) paketinden yararlanılmıştır. Belirtilen paket ve fonksiyonlar kullanılarak beş temel aşamada yürütülen veri üretim algoritması aşağıda verilmiştir. Öncelikle iki boyutlu veri üretimi için; Gerçek puanlar (Theta) üretiminde R'da bulunan "rmvnorm" fonksiyonu kullanılmıştır. Birey sayısı ve yetenek dağılım özellikleri tanımlanarak çalışmada belirlenen koşullar dahilinde referans ve odak gruplar için yetenek parametreleri

üretilmiştir. Bireylerin gerçek puanları, odak ve referans grupta aynı olacak; rastgele normal dağılıma uygun olacak şekilde üretilmiş; daha sonra Q matrisi tanımlanmıştır. Madde sayısı, yanıt kategori sayısı ve ilgili MTK modeli tanımlanarak madde parametreleri üretilmiştir. Bu amaçla alanyazın dikkate alınarak 2PLM'e ilişkin madde parametrelerinden a (ayırt edicilik) için a_1 ve a_2 parametreleri uniform dağılım 1.00 ve 2.00 arasında; b (madde güçlüğü) için ise b_1 ve b_2 parametreleri tek biçimli dağılımda minimum 0; maksimum 1 olacak şekilde belirlenmiştir. Bu modelde şans parametresinin (c) sıfır olduğu varsayılmaktadır (Lee, 2007).

Madde parametreleri üretiminde öncelikle referans grup için parametreler üretilmiş; odak grup için üstüne eklemeler yapılarak DMF gösteren maddeler üretilmiştir. Daha sonra referans grup ile odak grup bir araya getirilerek 1-0 bütünleşik veri seti üretilmiştir. Kişi ve madde parametrelerine dayalı olarak, R'deki plogis fonksiyonu kullanılarak her bir madde için doğru yanıtlama olasılıkları hesaplanmıştır. Hesaplanan olasılıklar, R'deki "runif" fonksiyonu kullanılarak her bir olasılık, tek biçimli bir dağılımdan rastgele bir sayı ile karşılaştırarak madde yanıtlarına dönüştürülmüştür. $P(X_{ij} = 1 \mid \theta_{j1}) = 1$ rastgele sayıdan büyükse, sınava giren kişinin maddeyi doğru yanıtladığı (1 puan); büyük değilse, sınava giren kişinin maddeyi yanlış yanıtladığı (0 puan) kabul edilmiştir. Üç boyutlu veri üretimi için, a_3 ve b_3 parametreleri de tanımlanmış, iki boyutlu veri üretiminde kullanılan fonksiyonlar kullanılarak referans ve odak gruplar için yetenek parametreleri üretilmiş, Q matrisi tanımlanmıştır. Kişi ve madde madde için iki boyutlu verilerde olduğu gibi doğru yanıt olasılığı hesaplanmıştır. Veri üretimine yönelik R kodlarının bir kısmı EK 1'de verilmiştir.

DMF Gösteren Verilerin Oluşturulması

Verilere DMF Yüklenmesi. İki boyutlu verilere DMF yüklenmiştir. Tüm DMF maddeleri, referans grup lehine simüle edilmiştir. Bunun için öncelikle; %10 için 6; %20 için 12 ve %30 için 18 "sample" komutu ile rastgele seçilmiştir. TB DMF üretiminde seçilen maddeler için referans grubun b parametrelerine sırasıyla .50 ve .75 büyüklüğünde DMF eklenmiştir. TBO DMF üretiminde seçilen maddeler için ise referans grubun a parametrelerine sırasıyla .50 ve .75 büyüklüğünde DMF eklenmiştir. Daha sonra belirtilen işlemler üç boyutlu veriler için tekrarlanmıştır. DMF gösteren madde oranı (%10, %20, %30), modelde dikkate alınan boyut sayısı (2, 3), DMF büyüklüğü (.50, .75)

ve DMF türü (TB, TBO) şeklinde değişimlenmiştir. Yukarıda belirtilen adımlar bu koşulların her biri için tekrarlanmıştır. Tekrar sayısı (100) olmak üzere her bir koşula ilişkin ilgili sayıda tepki örüntüleri elde edilmiştir.

100 Tekrarlı Veri Setlerinin Oluşturulması. Simülasyon deseninde yer alan koşullar dikkate alınarak öncelikle DMF gösteren madde oranı %10 ve boyut sayısı iki için işlemler yürütülmüştür. Bu kapsamda DMF büyüklüğü .50, DMF türü TB ve TBO 100 tekrarlı yanıt örüntüsü elde edilmiştir. Daha sonra DMF büyüklüğü .75, DMF türü TB ve TBO için yanıt örüntüsü elde edilmiştir. Aynı işlemler DMF gösteren madde oranı %10 olan üç boyut için de elde edilmiştir. Daha sonra %20 ve %30 DMF gösteren maddelere ait yanıt örüntülerinin elde edilmesinde %10 için yapılan işlemler tekrarlanmıştır.

Simülasyona Dayalı Verilerin Geçerlik İncelemeleri

Araştırma kapsamında simülasyon desenine bağlı olarak üretilen veriler kullanılmadan önce, üretilen verilerde elde edilmek istenilen koşulların ne ölçüde gerçekleştirilebildiğinin belirlenmesine yönelik geçerlik incelemeleri yapılmıştır. Simülasyona dayalı üretilen verilerin geçerliği kapsamında sırasıyla üretilen verilerin model veri uyumları, boyutluluğu, DMF gösteren maddeleri ile bu maddelerin DMF türleri ve DMF büyüklükleri incelenmiştir. Bu kapsamda yapılan incelemeler aşağıda ayrı başlıklar altında verilmiştir.

Veri Setlerine ait Model-Veri Uyumu İncelemeleri

Simülasyon desenine bağlı olarak 2 PL üretilen veri setlerinin, 2PL model ile uyum sağlayıp sağlamadığının kontrol edilebilmesi için ‘mirt’ (Chalmers, 2012) paketinden yararlanılmıştır. Bu kapsamda “mirt” fonksiyonunda yer alan “itemtype” bölümü sırasıyla “Rasch”, “2PL”, “3PL” olarak değiştirilerek üç farklı model kurulmuştur. ANOVA ile model karşılaştırmaları yapılarak AIC, BIC, -2logLik model uyum istatistikleri; iterasyon sayıları ve p değerleri incelenmiştir. Veri setinin her biri için belirtilen model karşılaştırmaları yapılarak elde edilen sonuçlar 24 koşul x 100 tekrar için ayrı ayrı matrisler halinde kaydedilerek değerlendirilmiştir. Yapılan değerlendirme

sonucunda tüm koşullarda üretilen veri setlerinin 2PL model ile uyum sağladığı; model veri uyumlarının %100 olduğu belirlenmiştir. Bu kapsamda model veri uyumuna yönelik bir ihlal olmadığı için bu bölüm tablo olarak verilmemiştir.

Boyutluluk Kapsamında Yapılan İncelemeler

Üretilen verilerin istenilen boyut yapısında olup olmadığının kontrolü için kaç faktörlü olduğu faktör Analizi ve paralel analizle incelenmiştir. Daha sonra üretilen veri setlerinin bir, iki ve üç boyutlu yapıdan hangisine daha iyi uyum sağladığı "mirt" analiziyle test edilmiştir. Analizlerde R dilinde yer alan *'psych'* (Revelle, 2022) ve '*sirt*' (Robitzsch, 2022) ve 'mirt' (Chalmers, 2012) paketlerinden yararlanılmıştır.

İlk olarak veri setleri 1-0 olarak puanlandığı için; ikili puanlanan madde yanıtlarından oluşan veri setlerine yönelik tetrakorik korelasyon matrisleri oluşturulmuştur. Faktör analizi kapsamında AFA yapmadan önce AFA varsayımları kontrol edilmiştir. Bu kapsamda KMO ve Bartlett testleri sonuçları incelenmiş; tüm veri setlerinde KMO test sonuçlarının 0,90'dan büyük ve Bartlett testlerinin ise manidar olduğu tespit edilmiştir ($p<.05$). Bu doğrultuda veri setlerinin AFA için uygun olduğuna karar verilmiştir.

Karar verildikten sonra '*psych*' paketi içerisinde yer alan '*fa.parallel*' fonksiyonu kullanılarak tüm veri setlerinde analiz sonuçları elde edilmiştir. Yamaç birikinti grafikleri incelendiğinde, özdeğerler arasındaki farkın iki faktörlü yapılarda ikinci faktörden sonra önemli oranda azaldığı görülmektedir. Diğer bir ifadeyle, ikinci faktörden sonraki faktörlerin varyansa olan katkıları, ilk iki faktörün katkılarıyla kıyaslandığında önemli oranda azalmakta ve birbirine yaklaşmaktadır. Paralel analiz sonucu incelendiğinde iki faktörlü verilerin, iki faktörlü ve iki bileşenli yapıya uyum sağladığı görülmektedir. Faktör analizi sonucu korelasyon matrisine dayalı standartlaştırılmış yükler (pattern matrix - desen matrisi) de incelenerek veri yapısının iki faktörlü yapıya uygunluğu teyit edilmiştir.

Üç faktörlü yapılar için yamaç birikinti grafikleri incelendiğinde, özdeğerler arasındaki farkın üçüncü faktörden sonra önemli oranda azaldığı görülmektedir. Diğer bir ifadeyle, üçüncü faktörden sonraki faktörlerin varyansa olan katkıları, ilk üç faktörün katkılarıyla kıyaslandığında önemli oranda azalmakta ve birbirine yaklaşmaktadır. Paralel analiz sonucu incelendiğinde üç faktörlü verilerin, üç faktörlü yapıya uyum

sağladığı görülmektedir. Veri setinin her biri için belirtilen analizler yapılarak elde edilen sonuçlar 24 koşul x 100 tekrar için ayrı ayrı matrisler halinde kaydedilerek değerlendirilmiştir. Boyut sayısına uyumun incelenmesinde Model 1 tek; Model 2 iki ve Model 3 üç boyutlu yapıya uyumu göstermek üzere üç farklı model kurulmuştur. Yapılan mirt analizleri sonucunda ANOVA ile model karşılaştırmaları yapılarak AIC, BIC, -2logLik model uyum istatistikleri; iterasyon sayıları ve p değerleri incelenmiştir. Veri setinin her biri için belirtilen model karşılaştırmaları yapılarak elde edilen sonuçlar 24 koşul x 100 tekrar için ayrı ayrı matrisler halinde kaydedilerek değerlendirilmiştir. Yapılan tüm incelemelere ait sonuçlar Tablo 2'de gösterilmiştir.

Tablo 2

Simülasyon Deseninde Belirlenen Boyut Sayısına Uyum Sonuçları

		100 Tekrarda Yanlış Belirlen Boyut Sayısı		
Koşullar	Beklenen Boyutluluk	Faktör Analizi	Paralel Analiz	Mirt
1.	2 Boyutlu	0 (% 100)*	0 (% 100)	2 (% 98)
2.	2 Boyutlu	1 (% 99)	0 (% 100)	2 (% 98)
3.	3 Boyutlu	0 (% 100)	0 (% 100)	0 (% 100)
4.	3 Boyutlu	0 (% 100)	0 (% 100)	0 (% 100)
5.	2 Boyutlu	0 (% 100)	0 (% 99)	3 (% 97)
6.	3 Boyutlu	0 (% 100)	0 (% 100)	0 (% 100)
7.	3 Boyutlu	0 (% 100)	0 (% 100)	0 (% 100)
8.	2 Boyutlu	0 (% 100)	1 (% 99)	3 (% 97)
9.	2 Boyutlu	2 (% 98)	0 (% 100)	2 (% 98)
10.	3 Boyutlu	0 (% 100)	0 (% 100)	4 (% 96)
11.	3 Boyutlu	0 (% 100)	0 (% 100)	2 (% 98)
12.	2 Boyutlu	0 (% 100)	1 (% 99)	3 (% 97)
13.	2 Boyutlu	0 (% 100)	0 (% 100)	0 (% 100)
14.	2 Boyutlu	1 (% 99)	0 (% 100)	1 (% 99)
15.	3 Boyutlu	0 (% 100)	0 (% 100)	(% 100)
16.	3 Boyutlu	0 (% 100)	0 (% 100)	(% 100)
17.	2 Boyutlu	0 (% 100)	0 (% 100)	(% 100)
18.	3 Boyutlu	0 (% 100)	0 (% 100)	(% 100)
19.	3 Boyutlu	0 (% 100)	0 (% 100)	(% 100)
20.	2 Boyutlu	1 (% 99)	0 (% 100)	(% 98)
21.	2 Boyutlu	0 (% 100)	0 (% 100)	(% 98)
22.	3 Boyutlu	1 (% 99)	0 (% 100)	(% 100)
23.	3 Boyutlu	1 (% 99)	1 (% 99)	(% 100)
24.	2 Boyutlu	0 (% 100)	0 (% 100)	0 (% 100)

*Değişimlenen ve analiz sonucu elde edilen boyut arasındaki uyum yüzdesi

Tablo 2 incelendiğinde her bir koşul için faktör analizi ve paralel analizle 100 tekrarda yanlış olarak belirlenen boyut sayısının olmadığı ya da bir tane olduğu; mirt analiziyle 100 tekrarda yanlış boyuta uyum sağlama sayısının ise sıfır ile dört arasında

değiştiği görülmektedir. Faktör analizi, paralel analiz ve mirt analiziyle 100 tekrarda değişimlenen ve analiz sonucu elde edilen boyut sayısı arasındaki uyum yüzdelerinin ise %96 ile %100 arasında değiştiği görülmektedir. Bu kapsamda boyut sayısının belirlenememe oranı incelendiğinde nominal alfa değeri olan %5'in altında olduğu diğer bir ifadeyle kabul edilebilir sınırlarda olduğu görülmektedir. Tüm bu sonuçlar birlikte değerlendirildiğinde üretilen verilerin istenilen boyut sayısında olduğu; boyutluluk açısından geçerli olduğu söylenebilir.

DMF Yüklenme Durumlarının Kontrol Edilmesi

Seçilen maddelere DMF'lerin yüklenip yüklenmediği DMF tekniklerinden LR, SIBTEST ve Lord'un χ^2 teknikleri ile kontrol edilmiştir. Her bir koşul için doğru ve yanlış olarak belirlenen DMF gösteren madde sayıları ve oranları hesaplanmıştır. Tablodaki değerlerin elde edilmesinde yapılan hesaplamalar aşağıda birer örnek koşul üzerinde gösterilmiştir.

Tabloda verilen, "Doğru Olarak DMF'li Belirlenen Madde Sayısı ve Oranları" değerlerin elde edilmesi kapsamında birinci koşul hesaplaması örnek olarak açıklanmıştır. Bu koşulda toplam madde sayısının %20'si diğer bir ifadeyle 60 maddeden 12'si DMF'li olarak belirlenmiştir. Bu doğrultuda 1. Koşul kapsamında 100 tekrarlı veri seti için 12x100=1200 maddenin DMF göstermesi beklenmektedir. Yapılan hesaplamalar sonucunda 1. koşul için LR ile 1144 maddenin diğer bir anlatımla DMF göstermesi planlanan maddelerin %95'inin; SIBTEST ile 1156 maddenin (%96), Lord'un χ^2 ile ise 1160 maddenin (%96) doğru bir şekilde DMF'li olarak belirlenebildiği görülmüştür.

Tabloda verilen, "DMF Yüklenmediği Halde DMF Bulunan Madde Sayısı ve Oranları" değerlerin elde edilmesi kapsamında birinci koşul hesaplaması örnek olarak açıklanmıştır. Tüm koşullarda madde sayısı 60 ve tekrar sayısı 100 olmak üzere 6x100=6000 madde bulunmaktadır. Yapılan hesaplamalar sonucunda 1. koşul için LR ile 6000 maddeden 292'sinin diğer bir anlatımla toplam madde sayısının %5'inin; SIBTEST ile 226 maddenin (%4), Lord'un χ^2 ile ise 168 maddenin (%3), DMF yüklenmediği halde DMF'li olarak belirlendiği görülmüştür. Tüm koşullar için belirtilen hesaplamalar tekrar edilmiş ve elde edilen sonuçlar Tablo 3'te verilmiştir.

Tablo 3

Doğru ve Yanlış Olarak DMF'li BelirlenenMadde Sayıları ve Oranları

Koşul	Doğru Olarak DMF'li Belirlenen Madde Sayısı ve Oranları			DMF Yüklenmediği Halde DMF Bulunan Madde Sayısı ve Oranları		
	LR	SIBTEST	Lord'un χ^2	LR	SIBTEST	Lord'un χ^2
1.	1144 (%95)	1156 (%96)	1160 (%96)	292 (%5)	226 (%4)	168 (%3)
2.	1140 (%95)	1154 (%96)	1150 (%96)	224 (%4)	195 (%3)	146 (%2)
3.	1156 (%96)	1166 (%97)	1160 (%96)	275 (%5)	209 (%4)	170 (%3)
4.	1136 (%95)	1156 (%96)	1144 (%95)	228 (%4)	212 (%4)	158 (%3)
5.	1747 (%97)	1762 (%97)	1758 (%97)	287 (%5)	230 (%4)	184 (%3)
6.	1750 (%97)	1759 (%97)	1740 (%96)	235 (%4)	231 (%4)	166 (%3)
7.	1731 (%96)	1734 (%96)	1751 (%97)	259 (%4)	234 (%4)	157 (%3)
8 .	1737 (%96)	1757 (%97)	1738 (%96)	259 (%4)	219 (%4)	172 (%3)
9.	582 (%97)	576 (%96)	575 (%95)	211 (%4)	197 (%3)	145 (%2)
10.	572 (%95)	562 (%94)	571 (%95)	211 (%4)	208 (%3)	127 (%2)
11.	570 (%95)	579 (%96)	570 (%95)	231 (%2)	295 (%5)	152 (%3)
12.	575 (%95)	551 (%92)	556 (%92)	245 (%4)	250 (%4)	169 (%3)
13.	1150 (%96)	1166 (%97)	1161 (%96)	240 (%4)	177 (%3)	134 (%2)
14.	1130 (%95)	1160 (%97)	1159 (%96)	188 (%3)	190 (%3)	147 (%2)
15.	1146 (%96)	1168 (%97)	1164 (%97)	238 (%4)	201 (%3)	126 (%2)
16.	1134 (%95)	1150 (%95)	1149 (%96)	212 (%4)	231 (%4)	152 (%3)
17.	1754 (%97)	1768 (%98)	1764 (%98)	248 (%4)	170 (%3)	115 (%2)
18.	1740 (%96)	1760 (%97)	1718 (%95)	300 (%5)	223 (%4)	172 (%3)
19.	1728 (%96)	1736 (%96)	1726 (%95)	319 (%5)	171 (%3)	198 (%3)
20.	1747 (%97)	1763 (%97)	1730 (%96)	279 (%5)	155 (%3)	121 (%2)
21.	558 (%92)	578 (%96)	558 (%93)	185 (%3)	189 (%3)	133 (%2)
22.	560 (%93)	574 (%95)	554 (%92)	180 (%3)	177 (%3)	110 (%2)
23.	550 (%91)	556 (%92)	576 (%96)	205 (%3)	205 (%3)	147 (%2)
24..	546 (%91)	559 (%93)	558 (%93)	199 (%3)	194 (%3)	113 (%2)

Tablo 3 incelendiğinde; tüm koşularda 100 tekrarlı veri setlerinde doğru olarak belirlenen DMF gösteren madde sayılarının 551 ile 1768 arasında; DMF gösteren maddeleri doğru belirleme oranının ise %92 ile %98 arasında değiştiği görülmektedir. Bununla birlikte tüm koşullarda tekniklerden en az biri %95 oranında DMF gösteren maddeleri belirleyebilmektedir. Tüm koşullarda DMF gösteren maddelerin belirlenememe oranı %8 ile %10 arasında değişmektedir. Bununla birlikte tüm koşullarda DMF gösteren maddelerin belirlenememe oranı incelendiğinde tekniklerden en az birinde nominal alfa değeri olan %5'in altında olduğu diğer bir ifadeyle kabul edilebilir sınırlarda olduğu görülmektedir.

Tablo 3 incelendiğinde tüm koşularda 100 tekrarlı veri setlerinde DMF yüklenmediği halde DMF'li olarak belirlenen madde sayılarının 110 ile 319 arasında; maddeleri yanlış olarak DMF'li belirleme oranının ise %2 ile %5 arasında değiştiği görülmektedir. Bu kapsamda tüm koşullarda DMF gösteren maddelerin yanlış belirlenme oranları incelendiğinde nominal alfa değeri olan %5'in altında olduğu diğer bir ifadeyle

kabul edilebilir sınırlarda olduğu görülmektedir. Elde edilen tüm sonuçlar birlikte değerlendirildiğinde, verilerin istenilen DMF yapısında üretilebildiği; veri üretiminin, DMF gösteren maddelerin elde edilmesi kapsamında geçerli olduğu söylenebilmektedir.

DMF Türlerinin Kontrol Edilmesi

DMF gösteren maddelerin simülasyon desenine bağlı olarak üretilmek istenilen DMF türüne uygunluğu DMF tekniklerinden LR, SIBTEST ve Lord'un χ^2 teknikleri ile kontrol edilmiştir. Her bir koşul ve tekrar için analiz sonucunda elde edilen sonuçlar ve grafikler incelenmiştir. Yapılan incelemeler sonucunda tüm koşullarda verilerin, simülasyon deseninde üretilmek istenilen DMF türüne uygun olarak üretildiği görülmüştür. Tüm koşul ve tekrarlarda, DMF türüne yönelik veri üretiminde uyum %100 olduğu için bu bölüm tablo olarak verilmemiştir.

DMF Büyüklüklerinin Kontrol Edilmesi

Seçilen maddelere istenilen büyüklüklerde DMF'lerin yüklenip yüklenmediği DMF tekniklerinden LR, SIBTEST ve Lord'un χ^2 teknikleri ile kontrol edilmiştir. Her bir koşul için doğru ve yanlış olarak belirlenen DMF büyüklükleri ve oranları hesaplanmıştır. Simülasyonda belirlenen DMF büyüklüklerinin incelenmesi kapsamında birinci koşul hesaplaması örnek olarak açıklanmıştır. Tüm koşullarda madde sayısı 60 ve tekrar sayısı 100 olmak üzere 6x100=6000 madde bulunmaktadır. Yapılan hesaplamalar sonucunda 1. Koşul için LR ile 6000 maddeden 292'sinin diğer bir anlatımla toplam madde sayısının %5'inin; SIBTEST ile 226 maddenin (%4), Lord'un χ^2 ile ise 168 maddenin (%3), belirlenen DMF büyüklüğü koşulunu sağlamadığı görülmektedir. Tüm koşullar için ilgili belirtilen hesaplamalar tekrar edilmiş ve elde edilen sonuçlar Tablo 4'te verilmiştir.

Tablo 4

Simülasyon Deseninde Belirlenen DMF Büyüklüklerinin İncelenmesi

	Yanlış Belirlenen DMF Büyüklükleri ve Oranları		
Koşullar	LR	SIBTEST	Lord'un χ^2
1.	274 (%5)	245 (%4)	230 (%4)
2.	298 (%5)	236 (%4)	155 (%3)
3.	279 (%5)	229 (%4)	173 (%3)

(Devam ediyor)

Tablo 4 (Devam)

Simülasyon Deseninde Belirlenen DMF Büyüklüklerinin İncelenmesi

	Yanlış Belirlenen DMF Büyüklükleri ve Oranları		
Koşullar	LR	SIBTEST	Lord'un χ^2
4.	238 (%4)	208 (%3)	203 (%3)
5.	291 (%5)	233 (%4)	198 (%3)
6.	238 (%4)	239 (%4)	185 (%3)
7.	250 (%4)	234 (%4)	202 (%3)
8 .	239 (%4)	206 (%3)	178 (%3)
9.	205 (%3)	201 (%3)	152 (%3)
10.	210 (%4)	199 (%3)	127 (%2)
11.	255 (%4)	285 (%5)	152 (%3)
12.	239 (%4)	254 (%4)	190 (%3)
13.	205 (%3)	181 (%3)	121 (%2)
14.	173 (%3)	189 (%3)	139 (%2)
15.	138 (%2)	208 (%3)	196 (%3)
16.	230 (%4)	219 (%4)	156 (%3)
17.	299 (%5)	198 (%3)	121 (%2)
18.	286 (%5)	219 (%4)	206 (%3)
19.	219 (%4)	156 (%3)	231 (%4)
20.	275 (%5)	151 (%3)	239 (%4)
21.	202 (%3)	184 (%3)	138 (%2)
22.	190 (%3)	197 (%3)	156 (%3)
23.	200 (%3)	199 (%3)	148 (%2)
24.	183 (%3)	190 (%3)	198 (%3)

Tablo 4 incelendiğinde; tüm koşulara ait 100 tekrarlı veri setlerinde DMF büyüklükleri yanlış olarak belirlenen madde sayılarının 121 ile 299 arasında; DMF büyüklüklerinin yanlış belirlenme oranının ise %2 ile %5 arasında değiştiği görülmektedir. Bu kapsamda tüm koşullarda DMF büyüklüklerinin yanlış belirlenme oranları incelendiğinde nominal alfa değeri olan %5'in altında olduğu diğer bir ifadeyle kabul edilebilir sınırlarda olduğu görülmektedir. Bu kapsamda elde edilen sonuçlar değerlendirildiğinde, verilerin istenilen DMF büyüklüğüne uygun olarak üretilebildiği; veri üretiminin, DMF büyüklüğü kapsamında geçerli olduğu söylenebilmektedir.

Verilerin Analizi

Bu çalışmada simülasyon deseni dikkate alınarak üretilen veri setleri için ilgili varsayımlar kontrol edilmiştir. Simülasyona dayalı veri setleri üzerinde DMF belirleme analizleri yapılmıştır. Araştırmanın örneklemi büyük olduğu için yöntemlerin benzer sonuçlar vermesi beklenmektedir. Bununla birlikte her bir yöntemin kendine ait sınırlılığı ve farklı koşullardan etkilenme durumu olduğundan (Wiberg, 2007) alanyazında bu

yöntemlerden birkaçının birlikte kullanılması, yöntemlerde uzlaşı olanların DMF gösterdiğine karar verilmesi tavsiye edilmektedir. Ek olarak yöntemlerin birtakım sınırlılıkları olabilmektedir. Örneğin Mantel- Haenszel tekniği, özellikle büyük örneklemlerde güçlü bir DMF belirleme tekniği olarak bilinmekle birlikte TBO DMF'yi belirlemede etkili değildir. Bu sınırlılık, model veri uyumu sağlandığında ve örneklem yeterince büyük olduğunda MTK yaklaşımlarında ve logistik regresyona dayalı tekniklerde görülmemektedir (Sireci vd., 2005). Bu kapsamda bu çalışmada MTK yaklaşımlarından Lord'un χ^2, SIBTEST ve LR analizleri yapılmıştır. Farklı yöntemler kullanılması, elde edilen sonuçların yöntemden kaynaklı olmadığının ortaya konulmasında etkili olacaktır. Üç farklı DMF yöntemi kullanılmasında etkili olan bir diğer durum, DMF türünün ve DMF büyüklüğünün devreye girmesiyle yöntemler arası farkın da ortaya çıktığının düşünülmesidir.

TB kestirimlerden elde edilen yetenek parametreleri üzerinde Lord'un χ^2, SIBTEST ve LR analizleri yapılmıştır. TB kestirimler için DMF sonuçlarına ait Tip I hata ve istatistiksel güç oranları elde edilmiştir. Daha sonra aynı veri seti üzerinde ÇBMTK ile kestirim yapılarak elde edilen yetenek parametleri üzerinde tekrar Lord'un χ^2, SIBTEST ve LR analizleri yapılmış, DMF sonuçlarına ait Tip I hata ve istatistiksel güç oranları elde edilmiştir.

Bu çalışma kapsamında boyutluluğu ÇB veriler üzerinde TBMTK kestirim sonucu elde edilen veri setlerine yönelik DMF analizlerine ait Tip I hata ve istatistiksel güç oranlarının; ÇBMTK kestirimi sonucu elde edilen veri setlerine yönelik DMF analizlerine ait Tip I hata ve istatistiksel güç oranlarından manidar bir şekilde farklılaşıp farklılaşmadığının tespit edilebilmesi için Faktöriyel ANOVA gerçekleştirilmiştir. Faktöriyel ANOVA, R dilinde yer alan *'car'* paketinden (Fox vd., 2012); anlamlılık düzeyi .005'ten düşük çıkan değişkenlere yönelik Post-Hoc testleri için ise '*emmeans'* paketinden (Lenth vd., 2019) yararlanılmıştır.

Gerçek verilerin analizinde; düzenlenen gerçek veri seti üzerinden sırasıyla tek ve çok boyutlu yetenek kestirimleri yapılmıştır. Çok boyutlu yetenek kestirimine ait veriler kullanılırken birinci yetenek düzeyi dikkate alınmıştır. Öncelikle tek boyutlu kestirim üzerinden elde edilen veri seti üzerinde LR, SIBTEST ve Lord'un χ^2 teknikleri kullanılarak DMF analizleri gerçekleştirilmiş, Tip I hata ve istatistiksel güç oranları hesaplanmıştır. Daha sonra aynı şekilde çok boyutlu kestirim üzerinden birinci yetenek düzeyi dikkate alınarak LR, SIBTEST ve Lord'un χ^2 teknikleriyle DMF analizleri gerçekleştirilmiş, Tip I hata ve istatistiksel güç oranları hesaplanmıştır.

Araştırma kapsamında tek ve çok boyutlu kestirimler sonucu LR, SIBTEST ve Lord'un χ^2 tekniklerinden elde edilen Tip I hata ve istatistiksel güç oranları arasında manidar farklılıklar olup olmadığının belirlenebilmesi için MANOVA analizi gerçekleştirilmiştir. MANOVA analizinin yapılabilmesi için gözlemlerin bağımsız, örneklem büyüklüğünün yeterli, varyans-kovaryans matrislerinin homojen, bağımsız değişkenin her grubu için her bir bağımlı değişken çifti arasında doğrusal bir ilişki olması; verilerin çoklu normal dağılım göstermesi; tek veya çok değişkenli aykırı değerlerin ve çoklu doğrusallığın olmaması varsayımlarının karşılanması gerekmektedir (Field, 2005; Howell, 2014). Bu kapsamda analizler yapılmadan ilgili varsayımlar kontrol edilmiştir. Varyansların homojenliği Levene testi kullanılarak incelenmiş, varyansların homojenliği varsayımının sağlandığı görülmüştür. Diğer varsayımlar *'Gerçek Veri Çalışması'* başlığı altında kontrol edilmiştir. Bu doğrultuda varsayımların karşılandığı belirlendikten sonra analize geçilmiştir.

Gerçek Veri Çalışması

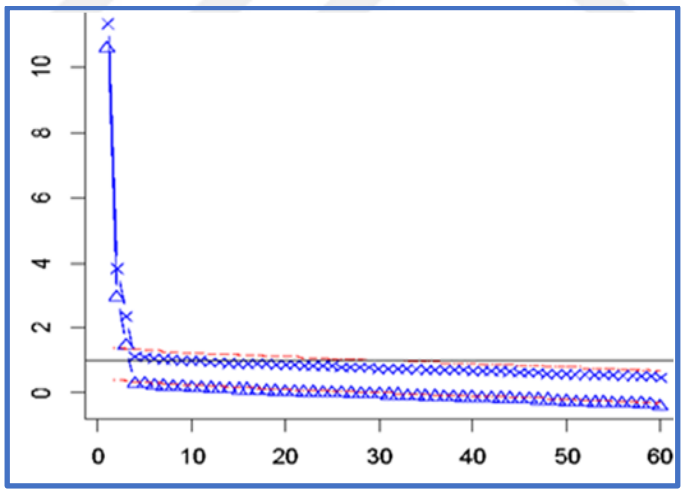
Araştırmanın ikinci kısmında simülasyon çalışması ile elde edilen bulgular, gerçek veri kullanılarak kontrol edilmiştir. Simülasyon çalışmasından elde edilen sonuçların gerçek veri sonuçlarıyla ne ölçüde uyum gösterdiği incelenmiştir. Bu kapsamda 954 kişi, 75 maddeden oluşan çoktan seçmeli bir Edebiyat sınavının verileri kullanılmıştır. İlgili veri seti R ortamına aktarıldıktan sonra bireylerin maddelere verdikleri doğru ve yanlış yanıtlar, 1-0 şeklinde ikili yanıt tepkilerine dönüştürülmüştür. Veri setinde maddelerin hiçbirini yanıtlamamış 13 birey veri setinden çıkarılmış, sonuç olarak geriye 941 birey 75 maddeden oluşan bir veri seti elde edilmiştir. Daha sonra simülasyon deseninde yer alan koşullardan birine olabildiğince uyum sağlayacak şekilde, veri setinden Bootstrap (yeniden örnekleme) yöntemi ile rastgele 60 madde seçilmiş, 941 bireyin bu maddelere yönelik yanıtlarını içerecek şekilde yeni bir veri seti oluşturulmuştur. Kayıp verilerin tamamlanmasında çoklu atama yöntemi kullanılmıştır. Çoklu atama yönteminde en az üç veri seti olması tavsiye edilmektedir. Bu kapsamda çoklu atama yöntemi ile kayıp verilerek tamamlanarak beş ayrı veri seti elde edilmiştir. Her bir veri seti üzerinden ayrı ayrı TB ve ÇB yetenek kestirimleri yapılarak ortalamaları alınmış; yetenek kestirimlerine yönelik veri setleri elde edilmiştir.

Veri setinin boyutluluk yapısının incelenmesinde kaç faktörlü olduğu faktör analizi ve paralel analizle; daha sonra üretilen veri setlerinin bir, iki ve üç boyutlu yapıdan hangisine daha iyi uyum sağladığı "*mirt*" analiziyle test edilmiştir. Analizlerde R dilinde yer alan *'psych'* (Revelle, 2022), '*sirt*' (Robitzsch, 2022) ve '*mirt*' (Chalmers, 2012) paketlerinden yararlanılmıştır. İlk olarak veri setleri 1-0 olarak puanlandığı için; ikili puanlanan madde yanıtlarından oluşan veri setlerine yönelik tetrakorik korelasyon matrisleri oluşturulmuştur. Faktör analizi kapsamında analiz yapılmadan önce varsayımları kontrol edilmiştir. Faktör analiz yapılabilmesi için verinin tek değişkenli ve çok değişkenli normallik göstermesi, doğrusal olması, örneklem büyüklüğünün faktörleştirmeye uygun olması; çoklu bağlantı sorununun, kayıp veri ve uç değerlerin olmaması gerekmektedir.

Normallik varsayımının kontrolünde çarpıklık ve basıklık katsayıları ile Kolmogrov-Smirnov ve Shapiro-Wilk istatistiksel yöntemleri kullanılmaktadır (Howell, 2014). Büyük örneklemlerde iyi çalışmadığı için Shapiro Wilk ve Kolmogorov-Smirnov testleri kullanılmamıştır. Grafiksel olarak her bir maddeye ait histogramlara ve saçılma diyagramlarına bakılmış ve maddelerin normal dağılım gösterdikleri görülmüştür. Çarpıklık ve basıklık katsayıları ±1 aralığında ise dağılımın normal, ±2 aralığında ise kabul edilebilir düzeyde normal olarak kabul edilmektedir (George ve Mallery, 2020). Bu kapsamda çalışmada elde edilen çarpıklık ve basıklık katsayılarının ±2 aralığında olduğu; normallik varsayımını karşıladığı görülmüştür. Bu durum tek değişkenli normallik varsayımının sağlandığı şeklinde yorumlanmıştır. Çok değişkenli normallik varsayımı kontrolü için Bartlett Küresellik Testi sonuçlarına bakılmış ve manidar olduğu görülmüştür ($\chi 2 = 14808{,}713_{(1225)}$; p<.05). Bu durum çok değişkenli normallik varsayımının karşılandığını göstermektedir. Çok değişkenli normallik varsayımının sağlanmış olması da doğrusallık varsayımın sağlandığı şeklinde yorumlanabilmektedir. Çoklu bağlantılılık sorununu test etmek üzere Tolerans ve VIF değerleri incelenmiştir. Maddelerin tümünde Tolerans değerlerinin 0.1'den büyük; VIF değerlerinin de 10'dan küçük olduğu görülmüştür. Bu durum çoklu bağlantı sorunu olmadığı şeklinde yorumlanabilmektedir. Tek değişkenli uç değer hesaplamasında öncelikle z değerleri hesaplanmıştır. Z değerleri (-3,+3) aralığında olmayan üç birey silinmiş veri setinde 938 birey kalmıştır. Çok değişkenli uç değer hesaplamasında Mahalanobis uzaklıkları hesaplanmış; hesaplanan bu değerin olması gerektiği gibi χ^2 /sd=86.661'den büyük olmadığı görülmüştür.

Yanıt matrisinin faktörleştirmeye uygunluğunun belirlenebilmesinde Kaiser-Mayer-Olkin (KMO) ve Bartlett testleri sonuçları incelenmiştir. Test sonucunda KMO değeri 0.971 olarak bulunmuştur. 0.94'ün üstündeki KMO değerleri "mükemmel" olarak yorumlandığından; yanıt matrisinin faktör analiz için mükemmel olduğu sonucuna ulaşılmıştır. Bartlett testi sonucunun ise manidar olduğu ($p<.05$) tespit edilmiştir. Tüm sonuçlar birlikte değerlendirildiğinde veri setinin faktör analizi için uygun olduğuna karar verilmiştir.

Karar verildikten sonra '*psych*' paketi içerisinde yer alan '*fa.parallel'* fonksiyonu kullanılarak tüm veri setlerinde analiz sonuçları elde edilmiştir. Yamaç birikinti grafikleri incelendiğinde, özdeğerler arasındaki farkın üçüncü faktörden sonra önemli oranda azaldığı görülmektedir. Faktör analizi sonucu korelasyon matrisine dayalı standartlaştırılmış yükler (pattern matrix-desen matrisi) incelenmiştir. İlgili matris incelendiğinde 8., 12., 21. ve 56. maddelerin ikinci; 1., 16. ve 37. maddelerin ise üçüncü boyutta yer aldığı görülmektedir. Paralel analizle verinin kaç faktörlü ve kaç bileşenli yapıda olduğu incelenmiştir. Gözlenen ve paralel analiz sonucunda elde edilen faktörlerin özdeğerlerine göre yamaç-birikinti grafiği Şekil 7'de verilmiştir.



Şekil 7. Paralel Analiz Yamaç-Birikinti Grafiği

Şekil 7 incelendiğinde özdeğerler arasındaki farkın üçüncü faktörden sonra önemli oranda azaldığı görülmektedir. Şekilde 7 incelendiğinde keskin platodan sonraki bileşenlerin varyansa yaptıkları katkı görece küçüktür. Faktör sayısına karar vermede bu nokta önemli olup bu noktaya kadar olan noktaların arası, olası faktör sayısı hakkında araştırmacıya bilgi vermektedir (Gorsuch, 1974). Diğer bir ifadeyle, üçüncü faktörden

sonraki faktörlerin varyansa olan katkıları, ilk üç faktörün katkılarıyla kıyaslandığında önemli oranda azalmakta ve birbirine yaklaşmaktadır. Paralel analiz sonucu incelendiğinde verinin üç faktörlü yapıya uyum sağladığı görülmektedir. Bu incelemelere ek olarak faktör sayısına karar vermek için 60 değişken, 938 kişi ve 100 tekrar ile Horn'un paralel analizi yapılmıştır. Yapılan paralel analiz sonucunda birinci, ikinci ve üçüncü faktörlerin rastgele hesaplanan özdeğerleri daha küçük değerler olarak bulunmuştur. Bu sonuç üç faktörü tutmamız yönündeki fikri desteklemektedir.

Paralel analiz ve faktör analizi sonuçları birlikte değerlendirildiğinde verilerin üç boyutlu yapıya uyumlu olduğu görülmektedir. Veri setinin bir, iki ve üç boyutlu yapıdan hangisine daha iyi uyum sağladığının belirlenmesinde "*mirt*" analizi yapılmıştır. Model 1, tek; Model 2, iki ve Model 3, üç boyutlu yapıya uyum için kurulan modelleri göstermektedir. Model karşılaştırmalarına ait değerler Tablo 5' de verilmiştir.

Tablo 5

Modellerin Karşılaştırma Sonuçları

	AIC	BIC	log.Lik	p
Model 1	59444.03	60025.28	-29602.01	.001*
Model 2	59198.17	60065.20	-29420.08	
Model 2	59198.17	60065.20	-29420.08	.001
Model 3	59157.69	60305.66	-29341.85	
Model 1	59444.03	60025.28	-29602.01	.001
Model 3	59157.69	60305.66	-29341.85	

*p<.001

Tablo 5 incelendiğinde Model 1 ve Model 2 arasında yapılan model karşılaştırması sonucunda LRT (likelihood ratio test) istatistiğinin p değeri manidar çıkmıştır. Bu bulgu iki boyutlu modelin, tek boyutlu modele göre veriye daha iyi uyum sağladığı şeklinde yorumlanabilir. Buna ek olarak, modellere ait AIC, BIC ve -2 log-Lik değerleri incelendiğinde bu değerlerin Model 2'de daha düşük olması da bu durumun doğruluğunu desteklemektedir. Model 2 ile Model 3 arasında yapılan model karşılaştırması sonucunda LRT istatistiğinin p değeri manidar çıkmıştır. Bu bulgu üç boyutlu modelin, iki boyutlu modele göre veriye daha iyi uyum sağladığı şeklinde yorumlanabilir. Bununla birlikte modellere ait AIC, BIC ve -2log-Lik değerleri incelendiğinde bu değerlerin Model 3'te daha düşük olması da bu durumun doğruluğunu desteklemektedir.

Model 1 ile Model 3 arasında yapılan model karşılaştırması sonucunda LRT istatistiğinin p değeri manidar çıkmıştır. Bu bulgu üç boyutlu modelin, tek boyutlu modele göre veriye daha iyi uyum sağladığı şeklinde yorumlanabilir. Buna ek olarak, modellere ait AIC, BIC ve -2log-Lik değerleri incelendiğinde bu değerlerin Model 3'te daha düşük olması da bu yorumun doğruluğunu desteklemektedir.

Tablo 5'te verilen üç model; AIC, BIC ve -2log-Lik değerlerine göre karşılaştırıldığında Model 3 için ilgili değerlerin daha düşük olduğu görülmüştür. İlgili üç model arasında en iyi uyum sağlayan modelin, Model 3 olması verinin üç boyutlu yapıya daha iyi uyum sağladığını göstermektedir. Veri seti incelendiğinde bu boyutların Edebiyat, Yazım Kuralları ve Dil Bilgisi şeklinde adlandırılabileceği düşünülmektedir. Veri setinin simülasyon koşullarına uygunluğunun belirlenmesinde LR, SIBTEST ve Lord'un χ^2 teknikleri kullanılarak DMF analizi yapılmıştır. Gerçek veri seti ile igili sadece öğrencilerin cinsiyet bilgilerine ulaşıldığı için DMF analizlerinde grup değişkeni olarak cinsiyet alınmıştır.

DMF analizlerinden elde edilen sonuçlar EK 3'de verilmiştir. EK 3'de verilen tablo incelendiğinde LR analizi sonucunda 1., 5., 8., 12., 16., 21., 27., 37., 40. ve 56. maddelerin, SIBTEST analizi sonucunda 1., 8., 16., 21., 37., 40. ve 56. maddelerin ve Lord'un χ^2 analizi sonucunda da 1., 8., 16., 21., 37., 40. ve 56. maddelerin DMF gösterdiği görülmektedir. Üç yöntemde de ortak olan 1., 8., 16., 21., 37., 40. ve 56. maddeler DMF gösteren maddeler olarak belirlenmiştir. Bu kapsamda 60 maddeden 7'si, diğer bir ifadeyle veri setinin %12'sinin DMF gösterdiği belirlenmiştir. Bu durum simülasyon koşullarından düşük DMF yüzdesi (%10 DMF) koşuluna karşılık gelmektedir. Buna ek olarak elde edilen sonuçlar incelendiğinde DMF gösteren maddelerin yüksek düzeyde DMF içerdiği görülmektedir. Veri setlerine ait sonuçlar ve madde yanıt eğrileri incelendiğinde; maddelerin simülasyon koşullarından TBO DMF koşuluna karşılık geldiği görülmektedir. Tüm sonuçlar birlikte değerlendirildiğinde gerçek veri seti simülasyon koşullarından DMF yüzdesi %10, DMF büyüklüğü yüksek, DMF türü TBO ve boyut sayısı 3 koşuluna karşılık gelmektedir.

Değerlendirme Ölçütleri

Elde edilen verilerde Tip I hata hesaplanmasında; her bir koşul için DMF yüklenmediği halde DMF bulunan maddeler belirlenmiştir. Tip I hatanın hesaplanma formülü Eşitlik 22'de verilmiştir:

$$Tip\ I\ Hata\ = \frac{1}{Replikasyon\ Sayısı} * \left(\frac{Yanlış\ olarak\ DMF'li\ belirlenen\ madde\ sayısı}{DMF'siz\ belirlenen\ madde\ sayısı}\right) \quad (22)$$

Eşitlik 22 kullanılarak örnek bir Tip I hata hesaplaması aşağıda verilmiştir. Bu kapsamda örnek olarak seçilen koşullardan birinde LR analizi sonucunda DMF yüklenmediği halde DMF bulunan madde sayısı 1. tekrarda üç; 2. tekrarda iki; 3. tekrarda üç; ……. 100. tekrarda iki olarak belirlenmiştir. Bu koşulda, LR için Tip I hatanın hesaplanması Eşitlik 23'te verilmiştir:

$$1.Tip\ Hata\ = \frac{1}{100} * \left(\frac{3+2+3+\cdots+2}{48}\right) = 0.048 \quad (23)$$

Eşitlik 23 incelendiğinde Tip I hatanın 0.048 olarak hesaplandığı görülmektedir. Elde edilen Tip I hatanın düzeylerini belirlemede Bradley'in (1978) sınıflandırması kullanılmıştır. Tutucu (conservative), korunan (maintained) ve şişirilmiş (inflated) şeklinde yapılan sınıflandırmada sınıflandırılmış nominal α değerleri Tablo 6'da verilmiştir (Akt: Price, 2014).

Tablo 6

Bradley'in Tip I Hata Oranlarına İlişkin Sınıflandırması

Düzey	Değer Aralığı
Tutucu (Conservative)	$\alpha < 0.045$
Korunan (Maintained)	$0.045 < \alpha < 0.055$
Şişirilmiş (Inflated)	$\alpha > 0.055$

Tablo 6 incelendiğinde, 0.045 ile 0.055 aralığı olarak belirlenen korunan (maintained) düzeydeki α değeri, Tip I hatanın istatistiksel olarak doğru tanımlanması için gerekli olan düzeydir. α değeri 0.045'ten küçük olduğu tutucu Tip I hata gösteren bir tekniğin gücü azalmaktadır. 0.05'ten yüksek olduğu durumlar ise şişirilmiş Tip I hata

olarak kabul edilmekte (akt., Price, 2014), Şişirilmiş Tip I hata oranı ise artması doğru olan H_0 hipotezinin reddedilme oranını artırmaktadır. Tip I hata ve istatistiksel güç oranlarının değerlendirilmesi kapsamında, Bradley'in referansı (Bradley's criterion) dikkate alınmıştır. Bradley (1978)'in referansına göre (akt., Bilican, 2014); her bir DMF yaklaşımı için Tip I hata oranı 0.025 - 0.075 aralığında olduğunda, hatanın kontrol altına alınabildiği kabul edilmiştir. Bu kapsamda; bu koşulda yukarıda verilen örnekte Tip I hatanın kontrol altına alınabildiği görülmektedir. Elde edilen verilerde istatistiksel güç oranlarını hesaplama formülü Eşitlik 24'te verilmiştir:

$$Güç\ Oranı = \frac{1}{Replikasyon\ Sayısı} * \left(\frac{Doğru\ olarak\ DMF'li\ belirlenen\ madde\ sayısı}{DMF'li\ belirlenen\ madde\ sayısı}\right) \quad (24)$$

Eşitlik 24 kullanılarak örnek bir istatistiksel güç oranı hesaplaması aşağıda verilmiştir. Bu kapsamda her bir koşul için DMF yüklendiği halde DMF bulunmayan maddeler belirlenmiştir. Bu kapsamda örnek olarak seçilen koşullardan birinde SIBTEST analizi sonucunda DMF yüklendiği halde DMF bulunmayan madde sayısı 1. tekrarda sıfır; 2. tekrarda sıfır; …. 36. tekrarda bir; ….; 52. tekrarda bir; …. 100. tekrarda sıfır olarak belirlenmiştir. DMF yüklendiği halde DMF bulunmayan madde sayısının 1. tekrarda sıfır olması DMF gösteren maddelerin 12'de 12 belirlenebildiğini; 36. tekrarda 1 olması 12'de 11 DMF belirlenebildiğini göstermektedir. Bu koşulda, SIBTEST için istatistiksel güç oranının hesaplanması Eşitlik 25'de verilmiştir:

$$Güç\ Oranı = \frac{1}{100} * \left(\frac{12+12+\cdots+11+\cdots+11+\cdots+12}{12}\right) = 0.998 \quad (25)$$

Eşitlik 25 incelendiğinde istatistiksel güç oranının 0.998 olarak hesaplandığı görülmektedir. Elde edilen istatistiksel güç oranının düzeylerini belirlemede Bradley'in (1978) sınıflandırması kullanılmıştır. Bradley (1978)'in referansına göre (akt., Bilican, 2014); istatistiksel güç oranı için 0.80 ölçüt olarak ele alınmış ve 0.80'den büyük olan oranlar için yeterli istatistiksel güç yorumu yapılmıştır. Bu kapsamda; bu koşulda SIBTEST tekniğinin yeterli istatistiksel güce sahip olduğu görülmektedir.

Ana Etkiler ile Etkileşimler için Faktöriyel ANOVA. Simülasyon çalışması kapsamında değişimlenen faktörlerin ana etkileri ile bu faktörler arasındaki etkileşim etkilerinin manidarlığının incelenmesinde alanyazındaki çalışmalara benzer olarak (Carroll, 2015; Harwell vd., 2018; Rockoff, 2018; Wood, 2011) Tip I hata ve istatistiksel güç oranları Faktöriyel ANOVA kullanılarak analiz edilmiş; sonuçlar, tablo ve grafikler yardımıyla desteklenmiştir. Faktöriyel ANOVA yapılabilmesi için gözlemlerin bağımsız, varyansların homojen ve dağılımların normal olması varsayımları karşılanmalıdır (Field, 2005; Howell, 2014). Bu kapsamda analizler yapılmadan ilgili varsayımlar kontrol edilmiştir. Varyansların homojenliği Levene testi kullanılarak incelenmiş, varyansların homojenliği varsayımının sağlandığı görülmüştür. Normallik varsayımının kontrolünde çarpıklık ve basıklık katsayıları ile Kolmogrov-Smirnov ve Shapiro-Wilk istatistiksel yöntemleri kullanılmaktadır (Howell, 2014). Çarpıklık ve basıklık katsayıları ±1 aralığında ise dağılımın normal, ±2 aralığında ise kabul edilebilir düzeyde normal olarak kabul edilmektedir (George ve Mallery, 2020). Bu kapsamda çalışmada elde edilen çarpıklık ve basıklık katsayılarının genel olarak ±2 aralığında olduğu; normallik varsayımını karşıladığı görülmüştür. Bu sınırların dışında kalan değerler için veri yapısına uygun olan yöntemler kullanılarak veri dönüştürme yapılmıştır (Hair vd., 2019; Tabachnick ve Fidell, 2013). Faktöriyel ANOVA analizi sonucunda elde edilen p değeri haricinde, elde edilen sonuçların pratik açıdan da anlamlı olduğunun ortaya konabilmesi için örneklem büyüklüğünden bağımsız olan etki büyüklüğü katsayısı da raporlanmalıdır. Farklı etki büyüklüğü katsayıları olmakla birlikte Faktöriyel ANOVA analizi sonuçlarında genellikle eta kare ve kısmi eta kare katsayılarının raporlandığı görülmektedir (Clark ve LaHuis, 2012; Norouzian ve Plonsky, 2018). Eta kare etki büyüklüğü katsayısını hesaplamada kullanılan formül Eşitlik 26'da verilmiştir:

$$\eta\, 2 = \frac{Kareler\ Toplamı_{faktör\ etkisi}}{Kareler\ Toplamı_{toplam}} \tag{26}$$

Eşitlik 26'da verilen eta kare katsayısı, bütün faktörler için toplanan kareler toplamının bir yüzdesine karşılık gelmektedir. Bu katsayı değeri 0 ile 1 arasında değişiklik göstermekte (Lakens, 2013) bağımsız değişken bağımlı değişkendeki varyansın tamamını açıkladığında katsayı 1'e eşit olmaktadır. Faktöriyel ANOVA'da

sonucu raporlanan bir diğer etki büyüklüğü katsayısı olan kısmi eta kare katsayısını hesaplamada kullanılan formül Eşitlik27'de verilmiştir:

$$\eta p2 = \frac{Kareler\ Toplamı_{faktör\ etkisi}}{Kareler\ Toplamı_{faktöretkisi} + Kareler\ Toplamı_{hata}} \quad (27)$$

Eşitlik 27'de verilen kısmi eta kare katsayısının ANOVA çalışmasında ilgilenilen faktöre ilişkin kareler toplamının, ilişkili hata varyansı ile faktörün kendi kareler toplamına bölünmesi sonucunda elde edildiği görülmektedir (Brown, 2008). Kısmi eta kare katsayısı, eta kare katsayısından daha yüksek sonuçlar vermekte olup alanyazında genellikle tek yönlü ANOVA sonucunda eta kare; çok yönlü ANOVA sonucunda ise kısmi eta kare katsayısının raporlandığı görülmektedir (Fritz vd., 2012). Bu çalışmada çalışmanın amacı ve koşulları dikkate alınarak ana etkiler için eta kare, etkileşim etkileri içinse kısmi eta kare katsayıları yorumlanmıştır. Eta kare katsayılarının yorumlanmasında .01 (düşük), .06 (orta) ve .14 (yüksek) ve kısmi eta karenin yorumlanmasında ise; $.01 \leq \eta p^2 < .06$ (düşük), $.06 \leq \eta p^2 < .14$ (orta) ve $\eta p^2 \geq .14$ (yüksek) ölçütleri (Cohen, 1988; Gray ve Kinnear, 2012) dikkate alınmıştır.

BÖLÜM 4

BULGULAR VE YORUMLAR

Bu bölümde araştırma problemleri kapsamında elde edilen bulgular ve bunlara ilişkin yorumlar yer almaktadır. Elde edilen bulgular ayrı başlıklar halinde verilmiştir. Gerçek veri uygulamasından elde edilen bulgular ise ilgili bulguları destekleyecek şekilde yorumlanmıştır.

Tek ve Çok Boyutlu Kestirimlere Yönelik DMF'li Madde Yüzdeleri, Boyut Sayısı, DMF Büyüklüğü ve DMF Türüne Göre Tip I Hata ve İstatistiksel Güç Oranları Ortalamalarının İncelenmesi

Bu çalışma kapsamında DMF kaynaklı çok boyutlu verilerde tek biçimli ve tek biçimli olmayan DMF türlerinin TBMTK ve ÇBMTK kestirimlerinin farklı koşullar altındaki performanslarının incelenmesi amaçlanmıştır. Bu amaçla DMF gösteren maddelerin belirlenmesinde LR, SIBTEST ve Lord'un χ^2 teknikleri kullanılmıştır. DMF belirleme tekniklerinin performansları, Tip I hata ve istatistiksel güç oranları kullanılarak karşılaştırılmıştır.

Belirtilen karşılaştırmanın yapılabilmesi için öncelikle simülasyona dayalı olarak elde edilen veri setleri üzerinde TB ve ÇB yetenek kestirimleri yapılmıştır. Kestirimlere yönelik elde edilen veri setleri üzerinde ayrı ayrı LR, SIBTEST ve Lord'un χ^2 DMF belirleme teknikleri kullanılarak Tip I hata ve istatistiksel güç oranları hesaplanmıştır. Bu kapsamda öncelikle tek boyutlu kestirimler üzerinde her bir koşul ve 100 tekrar için belirtilen teknikler kullanılarak Tip I hata ve istatistiksel güç oranları hesaplanmıştır. 100 tekrara ait sonuçların ortalamaları alınmıştır. 100 tekrar için elde edilen ortalamalar Tablo 7'de gösterilmiştir.

Tablo 7

Tek Boyutlu Yetenek Kestirimi için DMF'li Madde Yüzdeleri, Boyut Sayısı, DMF Büyüklüğü ve DMF Türüne Göre Tip I Hata ve İstatistiksel Güç Oranları Ortalamaları

				Tip I hata Ortalaması			İstatistiksel Güç Oranları Ortalaması		
DMF Yüzdesi	Boyut Sayısı	DMF Büyüklüğü	DMF Türü	LR	SIB-TEST	Lord'un χ^2	LR	SIB-TEST	Lord'un χ^2
%10	2	Düşük	TB	0.063	0.042	0.040	0.61	0.55	0.42
			TBO	0.061	0.045	0.041	0.58	0.58	0.40
		Yüksek	TB	0.071	0.046	0.063	0.85	0.67	0.60
			TBO	0.069	0.049	0.061	0.80	0.69	0.58
	3	Düşük	TB	0.070	0.044	0.049	0.55	0.52	0.40
			TBO	0.066	0.049	0.048	0.52	0.55	0.39
		Yüksek	TB	0.090	0.050	0.059	0.72	0.57	0.53
			TBO	0.088	0.055	0.064	0.68	0.60	0.51
%20	2	Düşük	TB	0.079	0.060	0.055	0.58	0.58	0.40
			TBO	0.073	0.062	0.050	0.57	0.65	0.38
		Yüksek	TB	0.150	0.092	0.060	0.86	0.70	0.62
			TBO	0.144	0.098	0.063	0.84	0.76	0.57
	3	Düşük	TB	0.087	0.064	0.060	0.45	0.54	0.38
			TBO	0.081	0.069	0.058	0.44	0.58	0.35
		Yüksek	TB	0.148	0.069	0.078	0.60	0.65	0.45
			TBO	0.114	0.099	0.080	0.50	0.73	0.43
%30	2	Düşük	TB	0.098	0.090	0.057	0.50	0.59	0.38
			TBO	0.087	0.092	0.059	0.45	0.64	0.36
		Yüksek	TB	0.198	0.126	0.070	0.88	0.69	0.70
			TBO	0.178	0.137	0.078	0.80	0.75	0.55
	3	Düşük	TB	0.108	0.090	0.073	0.47	0.52	0.35
			TBO	0.098	0.096	0.060	0.42	0.56	0.33
		Yüksek	TB	0.188	0.123	0.092	0.64	0.63	0.51
			TBO	0.138	0.157	0.100	0.48	0.65	0.40

Tablo 7 incelendiğinde Tip I hata ortalamalarının 0.038 ile 0.198 arasında olduğu görülmektedir. LR tekniği sonuçlarının TB için 0.063 ile 0.198; TBO için 0.061 ile 0.178; SIBTEST sonuçlarının TB için 0.042 ile 0.126; TBO için 0.045 ile 0.157; Lord'un χ^2 tekniği sonuçlarının TB için 0.040 ile 0.092; TBO için 0.041 ile 0.100 arasında olduğu görülmektedir. Tablo 7'de istatistiksel güç oranları ortalamaları ise 0.33 ile 0.88 arasındadır. LR tekniği için istatistiksel güç oranları ortalamalarının TB için 0.45 ile 0.88; TBO için 0.42 ile 0.84, SIBTEST sonuçlarının TB için 0.52 ile 0.70; TBO için 0.55 ile 0.76, Lord'un χ^2 testi sonuçlarının ise TB ve TBO için 0.35 ile 0.70 arasında olduğu görülmektedir.

Tablo 7 incelendiğinde DMF yüzdesi arttıkça tüm tekniklerde Tip I hata oranın arttığı, istatistiksel güç oranının ise düştüğü görülmektedir. Magis ve De Boeck (2011) de yaptıkları simülasyon çalışmasında DMF gösteren madde sayısı arttıkça Tip I hata oranlarının önemli ölçüde büyüdüğünü, özellikle büyük örneklemlerde bu büyümenin çok fazla olduğunu belirtmişlerdir. Buna ek olarak Wood (2011) tarafından yapılan çalışmada

da bu araştırma sonuçlarıyla paralel olarak; testte bulunan DMF gösteren madde oranının artmasıyla Tip I hata oranının arttığını belirtmiştir. Çalışmada güç oranları için elde edilen sonuç alanyazın (Atar ve Kamata, 2011; Hidalgo ve Lopez-Pina, 2004; Wyse ve Mapuranga, 2009) sonuçlarıyla paralellik göstermektedir. Buna ek olarak yapılan bu araştırmaya benzer olarak DMF oranı azaldıkça SIBTEST'in gücünün arttığı tespit edilmiştir (Kilmen, 2016).

Tablo 7 boyut sayısı bakımından incelendiğinde yine her bir teknik için boyut sayısı arttıkça Tip I hata oranının arttığı, istatistiksel güç oranın düştüğü görülmektedir. SIBTEST için de genel olarak boyut sayısı arttıkça istatistiksel güç oranı düşmekle birlikte elde edilen değerler birbirine yakındır. Boyut sayısı arttıkça istatistiksel güç oranın düşmesi, DMF gösteren maddelerin daha zor belirlenebildiğini göstermektedir. Bununla birlikte bu artışın çok büyük olmadığı gözlenmektedir. Boyut sayısına göre Tip I hata oranlarındaki en fazla artışın ve istatistiksel güç oranlarındaki en fazla düşüş DMF oranı %30 olduğunda gerçekleştiği görülmektedir.

Tablo 7 DMF büyüklüğü açısından incelendiğinde, DMF büyüklüğü arttıkça tüm teknikler için elde edilen Tip I hata ve istatistiksel güç oranlarının arttığı görülmektedir. Diğer bir anlatımla DMF büyüklüğünün yüksek olması durumunda, teknikler DMF gösteren maddeleri daha kolay belirleyebilmektedir. Çalışmada elde edilen bu bulgu alanyazındaki (Garrett, 2009; Magis ve De Boeck, 2012; Stark vd., 2006; Stephens-Bonty, 2008; Su ve Wang, 2005; Tunç, 201; Wang ve Su, 2004) DMF büyüklüğü arttıkça DMF belirlemede tekniklerinin I. Tip hata ve güç oranlarının arttığı sonucuna ulaşan çalışmalarla desteklenmektedir. Bu çalışma sonuçlarına paralel olarak Erdem Keklik (2012) b parametresine; Chang vd. (1996) ise a parametresine eklenen DMF büyüklüğündeki artışının, tekniğin gücünü arttırdığını ifade etmiştir.

Tablo 7 DMF türü bakımından incelendiğinde, ise diğer koşullardan farklı olarak tüm tekniklerin benzer şekilde çalışmadığı görülmektedir. LR tekniğinin TB DMF'nin belirlenmesinde daha yüksek Tip I hata ve istatistiksel güç oranları verdiği; TB ve TBO DMF'nin belirlenmesindeki bu farklığın DMF yüzdesi arttıkça arttığı görülmektedir. SIBTEST'in ise LR tekniğinin aksine; TBO DMF'nin belirlenmesinde daha yüksek Tip I hata ve istatistiksel güç oranları verdiği görülmektedir. SIBTEST için de LR tekniğinde olduğu gibi TB ve TBO DMF'nin belirlenmesindeki bu farklık DMF yüzdesi arttıkça artmaktadır. Lord'un χ^2 tekniğinde ise TB ve TBO DMF'nin belirlenmesinde benzer Tip I hata oranlarının elde edildiği görülmektedir.

Araştırmada kullanılacak verileri elde etmek için R dilinde yer alan *"ltm"* paketinden yararlanılmıştır. Elde edilen tekrarlı veri setleri üzerinde çok boyutlu yetenek parametre kestirimleri yapılmıştır. Kestirimlerin tüm tekrarlarda tek seferde yapılabilmesi için; çok boyutlu yetenek parametre kestirimleri için fonksiyon yazılmıştır. Koşullardan örnek olarak seçilen iki koşul için elde edilen birinci yetenek parametrelerine ait ilk dokuz, ortada yer alan dokuz ve son dokuza ait değerler EK 2' de verilmiştir. Çok boyutlu yetenek kestirimleri üzerinden 24 koşulx100 tekrar için elde edilen veri setlerine yönelik LR, SIBTEST, Lord'un χ^2 teknikleriyle DMF analizleri gerçekleştirilmiştir. DMF analizleri sonucu Tip I hata ve istatistiksel güç oranları hesaplanmıştır. Elde edilen sonuçların 100 tekrar için ortalamaları Tablo 8'de gösterilmiştir.

Tablo 8

Çok Boyutlu Yetenek Kestirimi için DMF'li Madde Yüzdeleri, Boyut Sayısı, DMF Büyüklüğü ve DMF Türüne Göre Tip I Hata ve Güç Oranları Ortalamaları

				Tip I Hata Oranları Ortalaması			İstatistiksel Güç Oranları Ortalaması		
DMF Yüzdesi	Boyut Sayısı	DMF Büyüklüğü	DMF Türü	LR	SIB-TEST	Lord'un χ^2	LR	SIB-TEST	Lord'un χ^2
%10	2	Düşük	TB	0.0287	0.0254	0.0265	0.990	0.995	0.996
			TBO	0.0255	0.0256	0.0254	0.985	0.990	0.991
		Yüksek	TB	0.0293	0.0279	0.0271	0.975	0.988	0.982
			TBO	0.0254	0.0282	0.0269	0.973	0.985	0.985
	3	Düşük	TB	0.0299	0.0260	0.0283	0.978	0.985	0.980
			TBO	0.0321	0.0259	0.0272	0.975	0.980	0.982
		Yüksek	TB	0.0312	0.0279	0.0297	0.987	0.990	0.988
			TBO	0.0260	0.0286	0.0284	0.985	0.988	0.987
%20	2	Düşük	TB	0.0298	0.0268	0.0276	0.985	0.990	0.990
			TBO	0.0252	0.0275	0.0263	0.980	0.986	0.984
		Yüksek	TB	0.0312	0.0274	0.0282	0.970	0.975	0.971
			TBO	0.0290	0.0278	0.0268	0.968	0.972	0.968
	3	Düşük	TB	0.0324	0.0292	0.0299	0.986	0.988	0.983
			TBO	0.0296	0.0290	0.0263	0.983	0.971	0.979
		Yüksek	TB	0.0344	0.0305	0.0312	0.978	0.982	0.982
			TBO	0.0319	0.0299	0.0303	0.974	0.980	0.975
%30	2	Düşük	TB	0.0330	0.0274	0.0284	0.977	0.982	0.982
			TBO	0.0268	0.0270	0.0280	0.975	0.984	0.978
		Yüksek	TB	0.0340	0.0283	0.0296	0.960	0.970	0.975
			TBO	0.0337	0.0288	0.0289	0.960	0.970	0.972
	3	Düşük	TB	0.0363	0.0299	0.0314	0.953	0.964	0.972
			TBO	0.0319	0.0312	0.0298	0.950	0.958	0.964
		Yüksek	TB	0.0335	0.0313	0.0323	0.968	0.956	0.958
			TBO	0.0317	0.0321	0.0315	0.964	0.950	0.960

Tablo 8 incelendiğinde Tip I hata ortalamaları 0.0252 ile 0.0363 arasında değişmekte olduğu gözlenmektedir. LR tekniği sonuçlarının TB için 0.0287 ile 0.0363; TBO için 0.0254 ile 0.0337; SIBTEST sonuçlarının TB için 0.0254 ile 0.0313; TBO için

0.0256 ile 0.0321; Lord'un χ^2 testi sonuçlarının TB için 0.0265 ile 0.0323; TBO için 0.0254 ile 0.0315 arasında olduğu görülmektedir. Tablo 8 incelendiğinde istatistiksel güç oranları ortalamalarının ise 0.950 ile 0.996 arasında olduğu görülmektedir. LR tekniği için istatistiksel güç oranları ortalamalarının TB için 0.953 ile 0.990; TBO için 0.950 ile 0.985, SIBTEST sonuçlarının TB için 0.956 ile 0.990; TBO için 0.950 ile 0.985, Lord'un χ^2 testi sonuçlarının ise TB için 0.958 ile 0.996; TBO için 0.960 ile 0.991 arasında olduğu görülmektedir.

Tablo 7 ve Tablo 8, Tip I hata ortalamaları bakımından birlikte incelendiğinde; tek boyutlu kestirimlere ait Tip I hata ortalamalarının 0.038 ile 0.198 arasında; çok boyutlu kestirimlere ait Tip I hata ortalamalarının ise 0.0252 ile 0.0363 arasında olduğu görülmektedir. İstatistiksel güç ortalamaları bakımından incelendiğinde ise tek boyutlu kestirimlere ait istatistiksel güç ortalamalarının 0.33 ile 0.88 arasında; çok boyutlu kestirimlere ait istatistiksel güç ortalamalarının ise 0.950 ile 0.996 arasında olduğu görülmektedir. Sonuçlar genel olarak değerlendirildiğinde araştırma dâhilindeki tüm koşullar ve alt düzeylerinde çok boyutlu kestirimlerden elde edilen Tip I hata oranlarının düşük; güç oranlarının ise daha yüksek olduğu görülmektedir.

DMF yüzdesi arttıkça LR, SIBTEST ve Lord'un χ^2 teknikleri sonucu elde edilen istatistiksel güç oranlarının düştüğü görülmektedir. Yapılan bu araştırmaya benzer olarak DMF oranı azaldıkça SIBTEST'in gücünün arttığı tespit edilmiştir (Kilmen, 2016). Atalay Kabasakal vd. (2014) çalışmasında ise bu çalışmadan farklı olarak DMF oranının SIBTEST'in gücünü etkilemediği belirtilmiştir. Bu çalışmada DMF oranı azaldıkça SIBTEST sonucu elde edilen istatistiksel güç oranları düşmekle birlikte sonuçlar birbirine oldukça yakın olduğu için istatistiksel güç üzerinde fazla bir etkisi olmadığı söylenebilir. Araştırma sonucunda tüm tekniklerde DMF büyüklüğü arttıkça Tip I hata ve istatistiksel güç oranları artmaktadır. Alanyazında bu araştırmanın bulgularıyla paralel olacak şekilde, hem tek hem de çok boyutlu kestirimlerde DMF büyüklüğü arttıkça Tip I hata ve istatistiksel güç oranlarının arttığı belirlenmiştir (Awour, 2008; Garrett, 2009; González-Romá vd., 2006; Hidalgo ve Lopez-Pina, 2004; Magis ve De Boeck, 2012; Kilmen, 2016; Kim, 2010; Su ve Wang, 2005; Sünbül ve Sünbül, 2016; Tunç ve Kutlu, 2018; Wyse ve Mapuranga, 2009).

TB DMF Türü için; Tek Boyutlu ve Çok Boyutlu Kestirimlerinin, DMF Belirleme Tekniği Boyut Sayısı, DMF Yüzdesi, DMF Büyüklüğü Koşullarındaki Tip I Hata Oranlarının İncelenmesi

Birinci araştırma sorusu kapsamında belirlenen koşullarda kestirim yöntemlerinin performanslarını belirlemek amacıyla yapılan Faktöriyel ANOVA; TB ve TBO DMF türleri için ayrı ayrı gerçekleştirilmiştir. Yapılan analizde kestirim teknikleri, DMF belirleme teknikleri, DMF yüzdeleri, DMF büyüklükleri faktörleri için ana etkiler ve bu faktörlerin bütün ikili etkileşimleri hesaplanmıştır. Kestirim faktörü ile diğer faktörlerin ikili etkileşimlerine ait sonuçlar tablolar halinde verilerek raporlaştırılmıştır. Ana etki ile etkileşim etkilerinden manidar olanlar yorumlanmıştır. Birinci araştırma sorusuna ait alt araştırma soruları kapsamında TB ve TBO DMF türü için yapılan Faktöriyel ANOVA sonuçları aşağıdaki başlıklar altında tablolar halinde verilmiştir. Bu inceleme kapsamında TB DMF türü için tek boyutlu ve çok boyutlu kestirimlere ait performanslar Tip I hata oranları bakımından incelenmiş, elde edilen sonuçlar Tablo 9'da verilmiştir.

Tablo 9

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

	Kareler Toplamı	*Sd*	*F*	*p*	η^2	ηp^2
Kestirim (K)	0.051740	1	721.3598	.000	.498	.957
Teknik (T)	0.009164	2	63.8804	.000	.088	.799
DMF yüzdesi (DY)	0.009172	2	63.9375	.000	.088	.780
Boyut sayısı (BS)	0.000202	1	2.8210	.103	.002	.081
DMF büyüklüğü (DB)	0.004388	1	61.1814	.000	.042	.656
K:T	0.007000	2	48.8003	.000	.067	.753
K:DY	0.007215	2	50.2975	.000	.070	.759
K:BS	0.000026	1	0.3631	.551	.000	.011
K:DB	0.003841	1	53.5548	.000	.037	.625
Hata	0.011066	32				

Tablo 9'da TB DMF türü için Tip I hata sonuçları yer almaktadır. Buna göre '*Kestirim*' ana etkisinin ($F(1,32) = 721.3598$, $p < .001$, $\eta^2 = .498$) manidar ve yüksek düzey bir etkiye sahip olduğu; '*Teknik*' ($F(2, 58) = 63.8804$, $p < .001$, $\eta^2 = .088$) ve '*DMF yüzdesi*' ($F(2,32) = 63.9375$, $p < .001$, $\eta^2 = .088$) ana etkilerinin ise manidar ve orta düzey bir etkiye sahip oldukları görülmektedir. Buna karşın '*DMF büyüklüğü'* ($F(1,32) = 61.1814$, $p < .001$, $\eta^2 = .042$) ana etkisinin ise manidar ve düşük düzey bir etkiye sahip

olduğu görülmektedir. Bunlara ek olarak '*Boyut sayısı*' ana etkisinin ($F(1,32) = 2.8210$, $p > .05$, $\eta^2 = .002$) ise manidar olmadığı görülmektedir.

'*Kestirim*' ile ikili etkileşimler incelendiğinde; '*Kestirim*' x '*Teknik*' ($F(2,32)=48.8003$, $p < .001$, $\eta p^2 = .753$), '*Kestirim*' x '*DMF yüzdesi*' ($F(2,32) =50.2975$, $p <.001$, $\eta p^2 = .759$) ve '*Kestirim*' x *'DMF büyüklüğü'* ($F(1,32) = 53.5548$, $p < .001$, $\eta p^2=.625$) ikili etkileşimlerinin manidar ve yüksek düzey bir etkiye sahip oldukları görülmektedir. Buna ek olarak '*Kestirim*' x '*Boyut sayısı*' ($F(1,32) = 0.3631$, $p >.05$, $\eta p^2=.011$) ikili etkileşiminin ise manidar olmadığı görülmektedir. Bu kapsamda TB DMF türü dikkate alındığında, Tip I hata üzerindeki en önemli ana etkinin '*Kestirim*'e ait olduğu belirtilebilir. Kestirim ana etkisinin Tip I hata oranları üzerinde yüksek düzey etkisi olduğu; '*Teknik*' ve '*DMF yüzdesi*' ana etkilerinin ise orta düzey etkileri olduğu görülmektedir. Buna karşın; '*Boyut sayısı*' faktörünün hem ana etkileri hem de '*Kestirim*' ile etkileşimleri manidar değildir. Bu nedenle, TB DMF gösteren maddelerin Tip I hata oranları üzerinde '*Boyut sayısı*' faktörün etkili olmadığı ifade edilebilir.

'Teknik' ve *'DMF yüzdesi'* ana etkileri istatistiksel orta düzey bir etkiye sahipken; '*Kestirim*' ile etkileşimlerini veren '*Kestirim*' ile *'Teknik' ve 'Kestirim'* ile *DMF yüzdesi'* ikili etkileşimlerinin ise yüksek düzey etkiye sahip olduğu görülmektedir. Buna ek olarak TB DMF türü dikkate alındığında, Tip I hata üzerinde en önemli etkileşim etkilerinin, '*Kestirim*' ile '*Teknik*' ve '*Kestirim*' ile '*DMF yüzdesi'*ne ilişkin ikili etkileşimler olduğu belirtilebilir. Bu doğrultuda TB DMF gösteren maddelerin Tip I hata oranları üzerinde hem '*Kestirim*' ile '*Teknik*' hem de '*Kestirim*' ile '*DMF yüzdesi'* faktörlerinin birlikte daha etkili olduğu ifade edilebilir.

'DMF Büyüklüğü' ana etkisi düşük düzey bir etkiye sahiptir. '*Kestirim*' ile etkileşimlerini veren '*Kestirim*' ile '*DMF büyüklüğü'* ikili etkileşimlerinin ise yüksek düzey etkiye sahip olduğu görülmektedir. Bu doğrultuda TB DMF gösteren maddelerin Tip I hata oranları üzerinde '*Kestirim*' ile *'DMF büyüklüğü'* faktörlerinin birlikte daha etkili olduğu ifade edilebilir. Faktöriyel ANOVA sonucu manidar olan ve orta-yüksek düzey etkiye sahip olan ana faktörlere yönelik ileri incelemeler için; '*stats*' (R Core Team, 2019) paketinde yer alan; eşit örnek hacimlerinde kullanılan ve "*family-wise*" %95 güven aralığı değerlerini daha az hata ile veren (Howell, 2014) '*TukeyHSD*' fonksiyonundan yararlanılarak Tukey testi gerçekleştirilmiştir. TB DMF türü için Tip I hata oranları üzerinde manidar etkiye sahip olan '*Kestirim'* yönteminin ikili karşılaştırmaları Tablo 10'da verilmiştir.

Tablo 10

TB DMF Türü İçin Tip I Hata Oranlarının Kestirim Yöntemine Göre İkili Karşılaştırmaları

Kestirim yöntemi	Fark	Standart hata	t (Tukey)	p
ÇB – TB	-0.0536	0.002	-26.858	.001**

*p<.05 **p<.01; ÇB: Çok Boyutlu; TB: Tek Boyutlu

Tablo 10 incelendiğinde TB DMF Türü için Tip I hata oranları üzerinde Çok Boyutlu (ÇB) kestirim yöntemi lehine ve manidardır (t=-26.858; p<.01). ÇB kestirim yöntemi, Tek Boyutlu (TB) kestirim yöntemine göre daha düşük Tip 1 hata değerleri üretmiştir. TB DMF türü için kestirim yöntemlerine ait Tip I hata değişimleri Şekil 8'de verilmiştir.



Şekil 8. Kestirim Ana Etkisi için Tip I Hata Oranları

Şekil 8'de görüldüğü gibi çok boyutlu kestirimlerden elde edilen Tip I hata ortalamalarının tek boyutlu kestirimlere göre düştüğü görülmektedir. TB DMF türü için Tip I hata oranları üzerinde manidar etkiye sahip olan '*Teknik'* faktörünün ikili karşılaştırmaları Tablo 11'de verilmiştir.

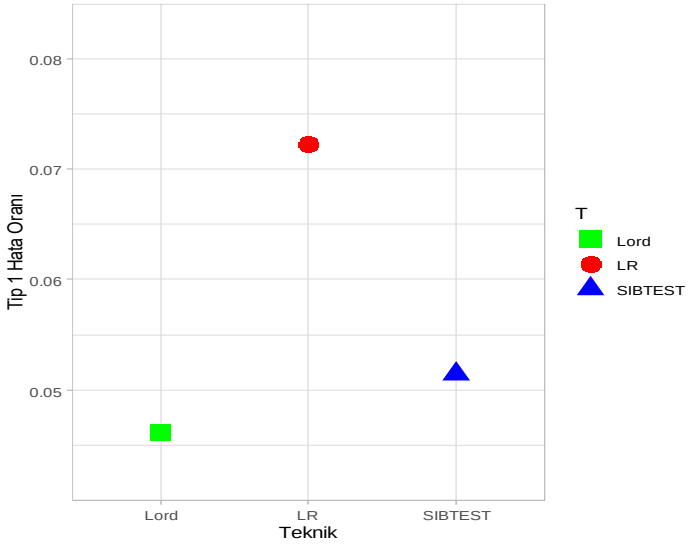
Tablo 11

TB DMF Türü İçin Tip I Hata Oranlarının Metotlara Göre İkili Karşılaştırmaları

Teknikler	Fark	Standart hata	t (Tukey)	p
Lord – LR	-0.02615		-10.918	.001**
Lord – SIBTEST	-0.00532	0.00239	-2.224	.082
LR – SIBTEST	0.02082		8.695	.001**

*p<.05 **p<.01

Tablo 11 incelendiğinde, TB DMF türü için Tip I hata oranları üzerinde Lord'un χ^2-LR (*t*=-10.918; *p*<*.01*) ve LR-SIBTEST (*t*=8.695; *p*<*.01*) teknikleri arasında manidar farklılıklar bulunmuştur. Lord'un χ^2 tekniği, LR tekniğine göre; SIBTEST tekniği ise LR tekniğine göre daha düşük Tip 1 hata değerleri üretmiştir. TB DMF türü için tekniklere ait Tip I hata değişimleri Şekil 9'da verilmiştir.



Şekil 9. Teknik Ana Etkisi için Tip I Hata Oranları

Şekil 9'de tekniklere göre Tip I hata değerleri yer almaktadır. Şekil 9 incelendiğinde en yüksek Tip I hata ortalamalarının LR; en düşük Tip I hata ortalamalarının ise Lord'un χ^2 tekniğinde olduğu görülmektedir. TB DMF türü için Tip I hata oranları üzerinde manidar etkiye sahip olan *'DMF yüzdesi'* yönteminin ikili karşılaştırmaları Tablo 12'de verilmiştir.

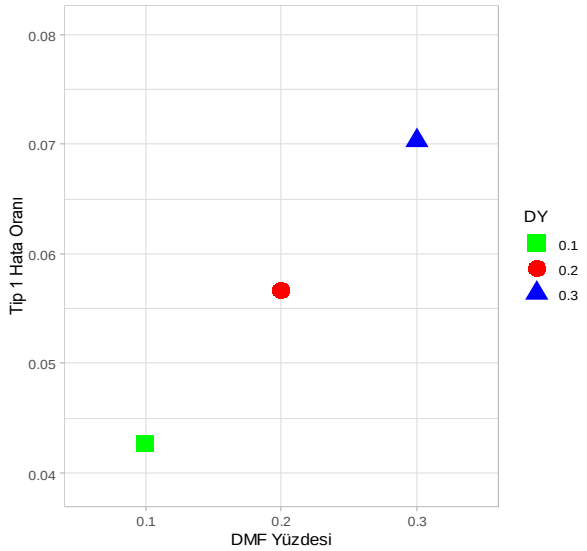
Tablo 12

TB DMF Türü İçin Tip I Hata Oranlarının DMF Yüzdelerine Göre İkili Karşılaştırmaları

DMF Yüzdeleri	Fark	Standart hata	(t) Tukey	p
%10 - %20	-0.0129		-5.223	.000**
%10 - %30	-0.0262	0.00247	-10.621	.000**
%20 - %30	-0.0133		-5.3	.000**

*p<.05 **p<.01

Tablo 12 incelendiğinde TB DMF türü için Tip I hata oranları üzerinde %10 - %20 (*t*=-5.223; *p*<.*01*), %10 - %20 (*t*=-10.621; *p*<.*01*) ve %10 - %20 (*t*=-5.399; *p*<.*01*) DMF yüzdeleri arasında manidar farklılıklar bulunmuştur. DMF yüzdesi arttıkça daha yüksek Tip I hata değerleri üretmiştir. TB DMF türü için DMF yüzdelerine ait Tip I hata oranları Şekil 10'da verilmiştir.



Şekil 10. DMF Yüzdesi Ana Etkisi için Tip I Hata Oranları

Şekil 10'de DMF yüzdesine göre Tip I hata değerleri yer almaktadır. Şekil 10 incelendiğinde DMF yüzdesi arttıkça Tip I hata oranlarının arttığı görülmektedir. En yüksek Tip I hata ortalamalarının %30; en düşük Tip I hata oranlarının ise %10 yönteminde olduğu görülmektedir. TB DMF türü için Tip I hata oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'Teknik'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 13'te verilmiştir.

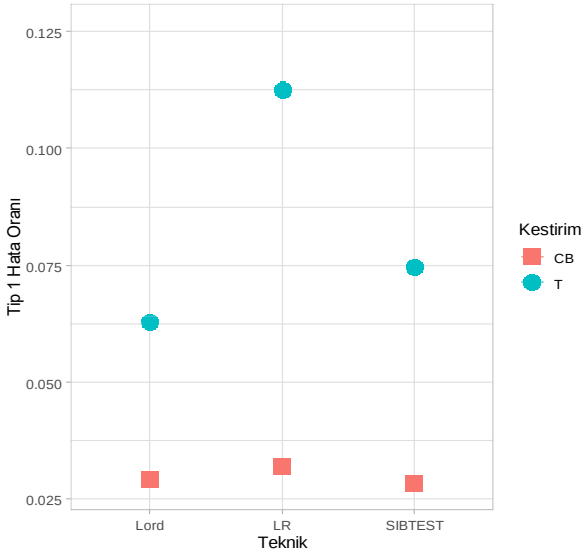
Tablo 13

TB DMF Türü İçin Tip I Hata Oranlarının 'Kestirim' x 'Teknik' İkili Etkileşime Göre İkili Karşılaştırmalar

KestirimxTeknik	Fark	Standart hata	Tukey (t)	p
ÇB Lord - TB Lord	-0.036267		-10.390	.000
ÇB Lord - ÇB LR	-0.001083		-0.310	.999
ÇB Lord - TB LR	-0.071767		-20.560	.000**
ÇBLord-ÇB SIBTEST	-0.000483		-0.138	.999
ÇB Lord - TB SIBTEST	-0.056017		-16.048	.000**
TB Lord - ÇB LR	0.035183		10.079	.000**
TB Lord - TB LR	-0.035500	0.00349	-10.170	.000**
TB Lord - ÇB SIBTEST	0.035783		10.251	.000**
TB Lord - TB SIBTEST	-0.019750		-5.658	.000**
ÇB LR - TB LR	-0.070683		-20.249	.000**
ÇB LR - ÇB SIBTEST	0.000600		0.172	.999**
ÇB LR - TB SIBTEST	-0.054933		-15.737	.000**
TB LR - ÇB SIBTEST	0.071283		20.421	.000**
TB LR - TB SIBTEST	0.015750		4.512	.001**
ÇBSIBTEST-TB SIBT.	-0.055533		-15.909	.000**

*p<.05 **p<.01; ÇB: Çok Boyutlu; TB: Tek Boyutlu

Tablo 13 incelendiğinde, TB DMF türü için Tip I hata oranları üzerinde ÇB Lord'un χ^2 - T Lord'un χ^2 (*t*=-10.390; *p<.01*), CB LR - T LR (*t*=-20.249; *p<.01*) ve ÇB SIBTEST-T SIBTEST (*t*=-15.909; *p<.01*) tüm yöntemler için tek ve çok boyutlu kestirimler arasında, çok boyutlu kestirim lehine manidar farklılıklar bulunmuştur. Çok boyutlu kestirim sonucu tüm teknikler için tek boyutlu kestirimden daha düşük Tip I hata oranları elde edilmiştir. TB DMF türü için Tip I hata oranları üzerinde ÇB Lord'un χ^2 - ÇB LR (*t*=-0.310; *p>.01*), ÇB Lord'un χ^2-ÇB SIBTEST (*t*=-0.138; *p>.01*) ve ÇB LR - ÇB SIBTEST (*t*=-15.909; *p>.01*) tüm çok boyutlu kestirimler için teknikler arasında manidar farklılıklar olmadığı görülmektedir. Diğer bir ifadeyle çok boyutlu kestirim sonucunda teknikler birbirine benzer çalışmaktadır. TB DMF türü için Tip I hata oranları üzerinde TB Lord'un χ^2 - TB LR (*t*=-10.170; *p<.01*), TB Lord'un χ^2 - TB SIBTEST (*t*=5.658; *p<.01*) ve TB LR - TB SIBTEST (*t*=4.512; *p<.01*) tüm tek boyutlu kestirimler için teknikler arasında manidar farklılıklar olduğu görülmektedir. TB DMF Türü için '*Kestirim*' x '*Teknik*' ikili etkileşimlerine ait Tip I hata değişimleri Şekil 11'de verilmiştir.



Şekil 11. Kestirim ile Teknik Etkileşimi için Tip I Hata Oranları

Şekil 11'de tekniklere göre tek ve çok boyutlu kestirimlerden elde edilen Tip I hata oranlarının değişimi yer almaktadır. Şekil 11 incelendiğinde çok boyutlu kestirimde Tip I hata oranlarının, tek boyutlu kestirimden daha düşük olduğu görülmektedir. Teknikler ayrı ayrı incelendiğinde; çok boyutlu kestirimde Tip I hata oranları açısından tek boyutlu kestirime göre en fazla değişimin LR tekniğinde; en az değişimin ise Lord'un χ^2 tekniğinde olduğu görülmektedir. Zwick vd. (1997) de gözlenen puana dayalı olan tekniklerin ölçme hataları ve testin güvenirliğini düşürebileceğini, dolayısıyla ortalamadan sapmış olan gerçek puanlara neden olabileceğini belirtmiştir. Bu kapsamda yapılan araştırmalarda, MTK modellerine uygun olarak üretilen verilerle gözlenen puana dayalı teknikler kullanılarak yapılan DMF analizlerinin şişirilmiş Tip I hataya sebep olduğu görülmektedir. Bu kapsamda çok boyutlu kestirimlerin bu probleme çözüm oluşturarak daha düşük Tip I hata oranları üretilmesine katkı sağladığı söylenebilir.

Literatür incelendiğinde; Tip I hata oranına göre LR ve Lord'un χ^2 teknikleri arasında manidar farklar bulunmamıştır (DeMars, 2009; Erdem Keklik, 2014; Gierl vd., 2000; Rogers ve Swaminathan, 1993; Vaughn ve Wang, 2010;). Bununla birlikte sonuçlar incelendiğinde LR Tekniğinin TB DMF'yi belirlemede etkili olduğu görülmektedir. Bu sonuçlar alanyazın tarafından desteklenmektedir (Jodoin ve Gierl, 2001; Pang vd.,1994). Bununla birlikte yapılan önceki çalışmalarda LR tekniği ile DMF belirleme sürecinde şişirilmiş Tip I hata oranlarının gözlendiği belirtilmiştir (Erdem Keklik, 2012; Penfield, 2009; Rogers ve Swaminathan, 1993). Bunun nedenlerinden biri pratik olarak manidar

olmayan bir maddenin manidar düzeyde DMF içeriyor olarak işaretlenmesi olarak belirtilmiştir (Jodoin ve Gierl, 1999). Şekil 11 incelendiğinde çok boyutlu kestirim sonucunda tekniklere ait Tip I hata oranlarının, tek boyutlu kestirime göre birbirine daha yaklaştığı görülmektedir. Bu durum '*Kestirim*' x '*Teknik*' ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumludur. TB DMF türü için Tip I hata oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'DMF yüzdesi'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 14'de verilmiştir.

Tablo 14

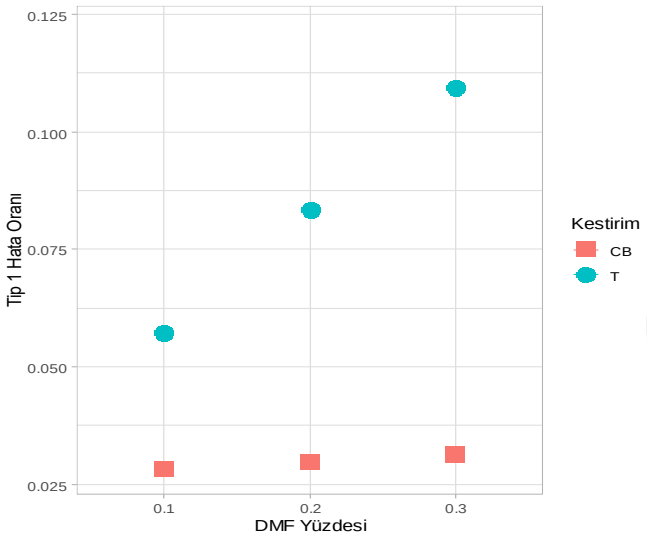
TB DMF Türü İçin Tip I Hata Oranlarının 'Kestirim' x 'DMF Yüzdesi' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxDMF Yüzdesi	Fark	Standart hata	Tukey (t)	p
ÇB %10 - ÇB %20	-0.001083		-0.344	.999
ÇB %10 - TB %20	-0.071767		-15.895	.000**
ÇB%10 - ÇB %30	-0.000483		-0.864	.952
ÇB %10 - TB %30	-0.056017		-23.009	.000**
TB %10 - ÇB %20	0.035183		8.508	.000**
TB %10 - TB %20	-0.035500		-7.043	.000**
TB %10 - ÇB %30	0.035783	.00349	7.988	.000**
TB %10 - TB %30	-0.019750		-14.157	.000**
ÇB %20 - TB %20	-0.070683		-15.551	.000**
ÇB%20 - ÇB %30	0.000600		-0.520	.995**
ÇB %20 - TB %30	-0.054933		-22.665	.000**
TB %20 - ÇB %30	0.071283		15.031	.000**
TB %20 - TB %30	0.015750		-7.114	.001**
ÇB %30 - TB %30	-0.055533		-22.145	.000**

*p<.05 **p<.01; CB: Çok Boyutlu; T: Tek Boyutlu

Tablo 14 incelendiğinde TB DMF türü için Tip I hata oranları üzerinde ÇB %10 - TB %10 (t=-8.852; p<*.01*), ÇB %20 - TB %20 (t=-15.551; p<*.01*) ve ÇB %30 - TB %30 (t=-22.145; p<*.01*) tüm DMF yüzdeleri için tek ve çok boyutlu kestirimler arasında, çok boyutlu kestirim lehine manidar farklılıklar bulunmuştur. Diğer bir ifadeyle '*Kestirim*' x '*DMF yüzdesi'* ikili etkileşime yönelik sonuçlar incelendiğinde, DMF yüzdesinin %10, %20 ve %30 olduğu durumların hepsinde çok boyutlu kestirim sonucu tek boyutlu kestirimden daha düşük Tip I hata oranları elde edilmiştir. TB DMF türü için Tip I hata oranları üzerinde ÇB %10 - ÇB %20 (t=-0.344; p>*.01*), ÇB%10 - ÇB %30 (t=-0.864; p>*.01*) ve ÇB %20 - ÇB %30 (t=-0.520; p>*.01*) tüm çok boyutlu kestirimler için DMF yüzdeleri arasında manidar farklılıklar olmadığı görülmektedir. Diğer bir ifadeyle çok boyutlu kestirim sonucunda DMF yüzdeleri birbirine benzer çalışmaktadır. TB DMF türü için Tip I hata oranları üzerinde TB %10 - TB %20 (t=-7.043; p<*.01*),

TB %10 - TB %30 (*t*=-14.157; *p*<.*01*) ve TB %20 - TB %30 (*t*=-7.114; *p*<.*01*) tüm tek boyutlu kestirimler için ise DMF yüzdeleri arasında manidar farklılıklar olduğu görülmektedir. TB DMF türü için *'Kestirim' x 'DMF yüzdesi'* ikili etkileşimlerine ait Tip I hata değişimleri Şekil 12'de verilmiştir.



Şekil 12. Kestirim ile DMF Yüzdesi Etkileşimi için Tip I Hata Oranları

Şekil 12'de DMF yüzdelerine göre tek ve çok boyutlu kestirimlerden elde edilen Tip I hata oranlarının değişimi yer almaktadır. Şekil 12 incelendiğinde çok boyutlu kestirimde Tip I hata oranlarının tek boyutlu kestirimden daha düşük olduğu görülmektedir. Alanyazındaki çalışmalar incelendiğinde bu çalışmada elde edilen sonuçlara paralel olarak testte bulunan DMF yüzdelerinin yetenek kestirimini etkilediği sonucuna ulaşılmıştır (Golia, 2010, Lee ve Zhang, 2017). DMF yüzdesine ait koşullar ayrı ayrı incelendiğinde çok boyutlu kestirimde Tip I hata oranları açısından tek boyutlu kestirime göre en fazla değişimin %30 koşulunda; en az değişimin ise %10 koşulunda olduğu görülmektedir. Diğer bir ifadeyle DMF yüzdesi oranlarındaki artış Tip I hata ortalamalarının artmasına neden olmuştur. DMF yüzdesi arttıkça tüm yöntemlerin Tip I hata oranlarının arttığı bulgusu, alanyazınla uyumludur (Fidalgo vd., 2000). Yapılan bu araştırmaya benzer olarak alanyazındaki çok sayıda araştırmada DMF gösteren madde sayısındaki artışın daha yüksek oranda Tip I hataya neden olduğu ortaya konmuştur (Atalay Kabasakal vd., 2014; Bauer vd., 2011; Finch, 2005; Guilera vd., 2013; Stark vd., 2006; Wang ve Yeh, 2003). Şekil 12 incelendiğinde çok boyutlu kestirim sonucunda

DMF yüzdelerine ait Tip I hata oranlarının, tek boyutlu kestirime göre birbirine daha yaklaştığı görülmektedir. Bu durum '*Kestirim*' x '*DMF yüzdesi*' ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumludur. TB DMF türü için Tip I hata oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'DMF büyüklüğü'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 15'te verilmiştir.

Tablo 15

TB DMF Türü İçin Tip I Hata Oranlarının 'DMF Büyüklüğü' İkili Etkileşime Göre İkili Karşılaştırmaları

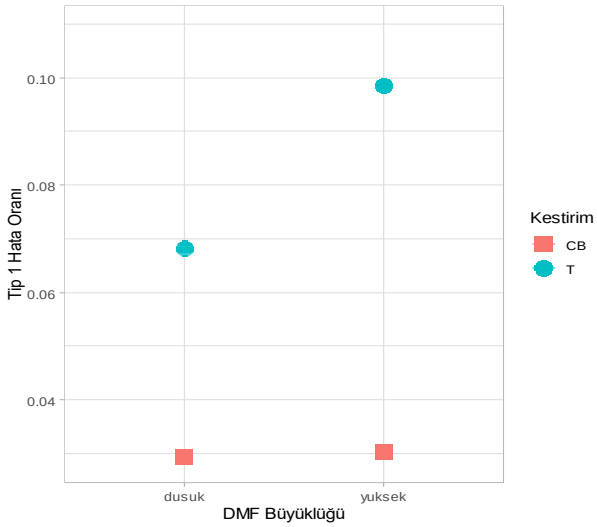
KestirimxDMF Büyüklüğü	Fark	Standart hata	t (Tukey)	p
ÇB düşük - TB düşük	-0.03909		-13.717	.000
ÇB düşük - ÇB yüksek	-0.00142		-0.499	.959
ÇB düşük - TB yüksek	-0.07065		-24.788	.000**
TB düşük - ÇB yüksek	0.03767	0.00285	13.218	.000**
TB düşük - TB yüksek	-0.03156		-11.072	.000**
ÇB yüksek - TB yüksek	-0.06923		-24.289	.000**

*p<.05 **p<.01; ÇB: Çok Boyutlu; TB: Tek Boyutlu

Tablo 15 incelendiğinde TB DMF türü için Tip I hata oranları üzerinde ÇB düşük - TB düşük (*t*=-18.600; *p<.01*) ve ÇB yüksek - TB yüksek (*t*=-19.407; *p<.01*) tüm DMF büyüklüğü için tek ve çok boyutlu kestirimler arasında, çok boyutlu kestirim lehine manidar farklılıklar bulunmuştur. Diğer bir ifadeyle '*Kestirim*' x '*DMF büyüklüğü'* ikili etkileşime yönelik sonuçlar incelendiğinde, DMF yüzdesinin düşük ve yüksek olduğu durumlarda, çok boyutlu kestirimin Tip I hata ortalamalarının tek boyutlu kestirimden manidar olarak farklılaştığı görülmektedir. '*Kestirim*' ile '*DMF büyüklüğü'* ikili etkileşime yönelik sonuçlar incelendiğinde, '*DMF büyüklüğü'*nün düşük ve yüksek olduğu durumların hepsinde, çok boyutlu kestirim, tek boyutlu kestirimden daha düşük Tip I hata oranları elde edilmiştir.

TB DMF türü için Tip I hata oranları üzerinde ÇB düşük - ÇB yüksek (*t*=0.499; *p>.01*) çok boyutlu kestirimler için boyut sayıları arasında manidar farklılıklar olmadığı görülmektedir. Diğer bir anlatımla çok boyutlu kestirim tek başına değerlendirildiğinde, DMF büyüklüğü koşulunun tüm düzeylerinde Tip I hata oranları arasındaki fark manidar değildir. Bu kapsamda çok boyutlu kestirim sonucunda boyut sayılarının birbirine benzer çalıştığı belirtilebilir. TB DMF türü için Tip I hata oranları üzerinde TB düşük - TB yüksek (*t*=-11.072; *p<.01*) tüm tek boyutlu kestirimler için boyut sayıları arasında

manidar farklılıklar olduğu görülmektedir. TB DMF türü için *'Kestirim' x 'DMF büyüklüğü'* ikili etkileşimlerine ait Tip I hata oranları Şekil 13'de verilmiştir.



Şekil 13. Kestirim ile DMF Büyüklüğü Etkileşimi için Tip I Hata Oranları

Şekil 13'te DMF büyüklüğüne göre tek ve çok boyutlu kestirimlerden elde edilen Tip I hata oranlarının değişimi yer almaktadır. Şekil 13 incelendiğinde genel olarak DMF büyüklüğü arttıkça Tip I hata oranının arttığı görülmektedir. Bu çalışmanın bulgularıyla paralel olacak şekilde alanyazında (Awour, 2008; Garrett, 2009; González-Romá vd., Kilmen, 2016; Kim, 2010; Magis ve De Boeck, 2012; Su ve Wang, 2005; 2006; Sünbül ve Sünbül, 2016; Tunç ve Kutlu, 2018) DMF büyüklüğü arttıkça Tip I oranlarının arttığı bulunmuştur. Buna ek olarak çok boyutlu kestirimde Tip I hata oranlarının tek boyutlu kestirimden daha düşük olduğu görülmektedir. Alanyazındaki çalışmalar incelendiğinde bu çalışmada elde edilen sonuçlara paralel olarak DMF büyüklüğünün yetenek kestirimi üzerinde etkisinin olduğu sonucuna ulaşılmıştır (Golia, 2014; Lee ve Zhang, 2017; Li ve Zumbo, 2009).

DMF büyüklüğüne ait koşullar ayrı ayrı incelendiğinde çok boyutlu kestirimde Tip I hata oranları açısından tek boyutlu kestirime göre en fazla değişimin yüksek; en az değişimin ise düşük '*DMF büyüklüğü'* koşulunda olduğu görülmektedir. Diğer bir ifadeyle DMF büyüklüğünün artması, çok boyutlu kestirimin tek boyutlu kestirime göre Tip I hata oranlarında sağladığı düşüşün artmasına neden olmuştur. Şekil 13 incelendiğinde çok boyutlu kestirim sonucunda DMF büyüklüklerine ait Tip I hata oranlarının, tek boyutlu kestirime göre birbirine daha yaklaştığı görülmektedir. Diğer bir

ifadeyle çok boyutlu kestirim sonucunda DMF büyüklükleri birbirine benzer çalışmaktadır. Bu durum '*Kestirim*' x '*DMF büyüklüğü'* ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumludur.

TB DMF Türü için; Tek Boyutlu ve Çok Boyutlu Kestirimlerinin, DMF Belirleme Tekniği, Boyut Sayısı, DMF Yüzdesi, DMF Büyüklüğü Koşullarındaki İstatistiksel Güç Oranlarının İncelenmesi

Bu inceleme kapsamında TB DMF türü için tek boyutlu ve çok boyutlu kestirimlere ait performanslar istatistiksel güç oranları bakımından incelenmiş, elde edilen sonuçlar Tablo 16'da verilmiştir.

Tablo 16

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

	Kareler Toplamı	*Sd*	*F*	*p*	η^2	ηp^2
Kestirim (K)	2.94881	1	5083.5855	.000	.826	.993
Teknik (T)	0.08406	2	72.4616	.000	.023	.819
DMF Yüzdesi (DY)	0.00261	2	2.2499	.000	.001	.123
Boyut Sayısı (BS)	0.04376	1	75.4376	.002	.012	.702
DMF Büyüklüğü (DB)	0.12161	1	209.6435	.000	.034	.868
K:T	0.09088	2	78.3336	.004	.025	.830
K:DY	0.00096	2	0.8238	.000	.000	.048
K:BS	0.03668	1	63.2264	.000	.010	.664
K:DB	0.14231	1	245.3368	.007	.040	.885
Hata	0.01856	32				

Tablo 16'da TB DMF türü için istatistiksel güç oranları yer almaktadır. Buna göre '*Kestirim*' ($F(1,32)=5083.5855$, $p < .001$, $\eta^2 = .826$) ana etkisinin yüksek düzey bir etkiye sahip olduğu görülmektedir. Bununla birlikte '*Teknik*' ($F(2,32)=72.4616$, $p<.001$, $\eta^2 =.023$), '*DMF yüzdesi*' ($F(2,32) = 2.2499$, $p < .001$, $\eta^2 = .001$), '*Boyut sayısı*' ($F(1,32) = 75.4376$, $p < .005$, $\eta^2 = .012$) ve '*DMF büyüklüğü'* ($F(1,32) = 209.6435$, $p<.001$, $\eta^2 = .034$) ana etkilerinin de manidar ancak düşük düzey bir etkiye sahip olduğu görülmektedir. '*Kestirim*' ile ikili etkileşimler incelendiğinde; '*Kestirim*' x '*Teknik*' ($F(2,32) = 78.3336$, $p < .005$, $\eta p^2= .830$), '*Kestirim*' x '*Boyut sayısı*' ($F(1,32)=63.2264$, $p < .05$, $\eta p^2= .664$)) ve '*Kestirim*' x '*DMF büyüklüğü'* ($F(1,32) = 245.3368$, $p < .05$, $\eta p^2 = .885$) ikili etkileşimlerinin manidar ve yüksek düzey bir etkiye sahip olduğu

görülmektedir. Buna karşın '*Kestirim*' x '*DMF yüzdesi*' ($F(2,32) = 0.8238$, $p < .001$, $\eta p^2 = .048$) ikili etkileşimlerinin ise manidar ancak orta düzey bir etkiye sahip olduğu görülmektedir. Bu kapsamda TB DMF türü dikkate alındığında, istatistiksel güç oranları üzerindeki en önemli ana etkinin '*Kestirim*' e ait olduğu belirtilebilir. '*Kestirim*' ana etkisinin istatistiksel güç oranları üzerinde yüksek düzey etkisi olduğu; '*Teknik*', '*Boyut sayısı*' ve '*DMF büyüklüğü'* ana etkilerinin ise düşük düzey etkileri olduğu görülmektedir. '*DMF yüzdesi*' ana etkisinin ise istatistiksel güç oranları üzerinde etkisi olmadığı saptanmıştır.

'Teknik', '*DMF büyüklüğü'* ve *'Boyut sayısı*' ana etkileri düşük düzey bir etkiye sahipken; '*Kestirim*' ile etkileşimlerini veren '*Kestirim*' ile *'Teknik' ve 'Kestirim'* ile *'DMF büyüklüğü'* ve *'Kestirim' ile* '*Boyut sayısı'* ikili etkileşimlerinin ise yüksek düzey etkiye sahip olduğu görülmektedir. Buna ek olarak TB DMF türü dikkate alındığında, istatistiksel güç oranları üzerinde en önemli etkileşim etkilerinin, '*Kestirim*' ile '*Teknik*', '*Kestirim*' ile *'DMF Büyüklüğü' ve 'Kestirim' ile* '*Boyut sayısı'*na ilişkin ikili etkileşimler olduğu belirtilebilir. TB DMF gösteren maddelerin istatistiksel güç oranları üzerinde '*Kestirim*' ile birlikte '*Teknik*', '*Boyut sa*yısı'' ve '*DMF büyüklüğü'* faktörlerinin birlikte oldukça etkili olduğu ifade edilebilir. Buna karşın '*Kestirim*' ile '*DMF yüzdesi*' faktörlerinin birlikte manidar bir etkiye sahip olduğu ancak bu etkinin düşük düzeyde olduğu görülmektedir. '*DMF yüzdesi*' ana faktörünün TB DMF gösteren maddelerin istatistiksel güç oranları üzerinde bir etkisi yok iken '*Kestirim*' ile ikili etkileşimlerinin istatistiksel güç oranları üzerinde düşük düzey bir etkiye sahip olduğu görülmektedir. Faktöriyel ANOVA sonucu manidar olan ve orta-yüksek düzey etkiye sahip olan ana faktörlere yönelik ileri incelemeler için; '*stats*' (R Core Team, 2019) paketinde yer alan '*TukeyHSD*' fonksiyonundan yararlanılarak Tukey testi gerçekleştirilmiştir. TB DMF türü için istatistiksel güç oranları üzerinde manidar etkiye sahip olan '*Kestirim'* yönteminin ikili karşılaştırmaları Tablo 17'de verilmiştir.

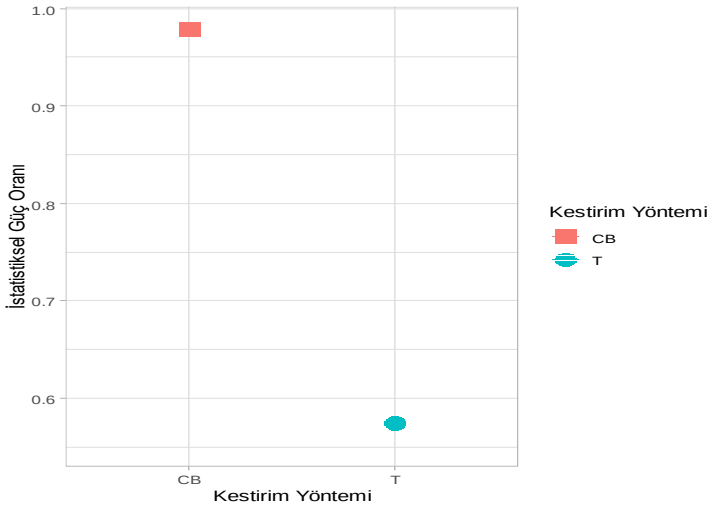
Tablo 17

TB DMF Türü İçin İstatistiksel Güç Oranlarının Kestirim Yöntemine Göre İkili Karşılaştırmaları

Kestirim yöntemi	Fark	Standart hata	t (Tukey)	p
ÇB – T	0.405	0.00568	71.299	.001**

*p<.05 **p<.01; ÇB: Çok Boyutlu; TB: Tek Boyutlu

Tablo 17 incelendiğinde, TB DMF türü için istatistiksel güç oranları incelendiğinde Çok Boyutlu (ÇB) kestirimin TB kestirimden manidar şekilde farklılaştığı görülmektedir (*t*=-26.858; *p*<*.01*). ÇB kestirim yöntemi, Tek Boyutlu (TB) kestirim yöntemine göre daha yüksek istatistiksel güç değerleri üretmiştir. TB DMF türü için 'Kestirim' faktörüne ait istatistiksel güç oranları değişimleri Şekil 14'de verilmiştir.



Şekil 14. Kestirim Ana Etkisi için Güç İstatistiksel Güç Oranları

Şekil 14'te görüldüğü gibi ÇB kestirimlerden elde edilen istatistiksel güç ortalamalarının TB kestirimlere göre düştüğü görülmektedir. Faktöriyel ANOVA sonucu manidar olan ve orta-yüksek düzey etkiye sahip olan ikili etkileşimlere yönelik ileri incelemeler için; '*stats*' (R Core Team, 2019) paketinde yer alan '*TukeyHSD*' fonksiyonundan yararlanılarak Tukey testi gerçekleştirilmiştir. TB DMF türü için istatistiksel güç oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'Teknik'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 18'de verilmiştir.

Tablo 18

TB DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'Teknik' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxTeknikler	Fark	Standart hata	t (Tukey)	p
ÇB Lord - TB LR	0.387083		31.613	.000**
ÇBLord -ÇB SIBTEST	0.000917	.00122	0.075	.999
ÇB Lord - TB SIBTEST	0.332083		27.121	.000**
TB Lord - ÇB SIBTEST	-0.538667		-43.993	.000**

*p<.05 **p<.01; ÇB: Çok Boyutlu; TB: Tek Boyutlu

(Devam ediyor)

Tablo 18 (Devam)

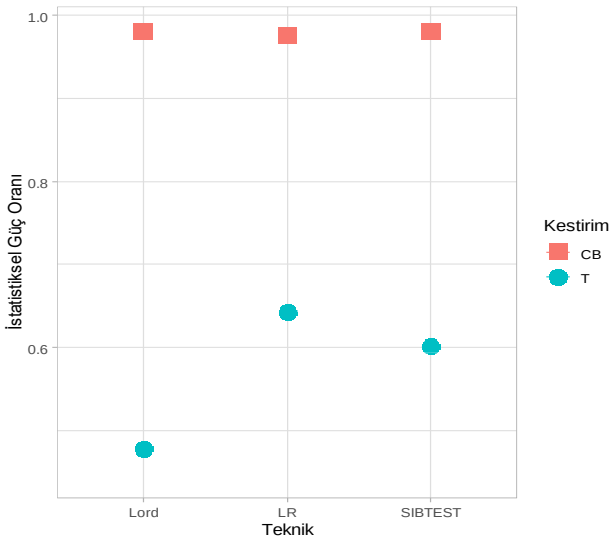
TB DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'Teknik' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxTeknikler	Fark	Standart hata	t (Tukey)	p
TB Lord - ÇB SIBTEST	-0.538667		-43.993	.000**
TB Lord - TB SIBTEST	-0.207500		-16.947	.000**
TB Lord-TB SIBTEST	-0.207500		-16.947	.000**
ÇB LR - TB LR	0.382667		31.253	.000**
ÇB LR - ÇB SIBTEST	-0.003500	.00122	-0.286	.999**
ÇB LR - TB SIBTEST	0.327667		26.761	.000**
TB LR - ÇB SIBTEST	-0.386167		-31.538	.000**
TB LR - TB SIBTEST	-0.055000		-4.492	.001**
ÇBSIBTEST-TB SIB.	0.331167		27.046	.000**

*p<.05 **p<.01; CB: Çok Boyutlu; T: Tek Boyutlu

Tablo 18 incelendiğinde, TB DMF türü için istatistiksel güç oranları üzerinde ÇB Lord'un χ^2 - TB Lord'un χ^2 (t=44.068; $p<.01$), ÇB LR - TB LR (t=31.253; $p<.01$) ve ÇB SIBTEST - TB SIBTEST (t=27.046; $p<.01$) tüm teknikler için tek ve ve çok boyutlu kestirimler arasında, çok boyutlu kestirim lehine manidar farklılıklar bulunmuştur. ÇB kestirim sonucu tüm metotlarda TB kestirimden daha yüksek istatistiksel güç oranları elde edilmiştir. TB DMF türü için istatistiksel güç oranları üzerinde ÇB Lord'un χ^2 - ÇB LR (t=0.361; $p>.01$), ÇB Lord'un χ^2 - ÇB SIBTEST (t=0.075; $p>.01$) ve ÇB LR - ÇB SIBTEST (t=-0.2869; $p>.01$) tüm ÇB kestirimler için teknikler arasında manidar farklılıklar olmadığı görülmektedir. Diğer bir ifadeyle ÇB kestirim sonucunda teknikler birbirine benzer çalışmaktadır.

TB DMF türü için istatistiksel güç oranları üzerinde TB Lord'un χ^2 - TB LR (t=-12.455; $p<.01$), TB Lord'un χ^2 - TB SIBTEST (t=-16.947; $p<.01$) ve TB LR - TB SIBTEST (t=-4.492; $p<.01$) tüm tek boyutlu kestirimler için teknikler arasında manidar farklılıklar olduğu görülmektedir. Diğer bir ifadeyle TB kestirim sonucunda teknikler birbirinden farklı çalışmaktadır. TB DMF türü için '*Kestirim*' x '*Teknik*' ikili etkileşimlerine ait istatistiksel güç oranları Şekil 15'te verilmiştir.



Şekil 15. Kestirim ile Teknik Etkileşimi için İstatistiksel Güç Oranları

Şekil 15'te tekniklere göre tek ve çok boyutlu kestirimlerden elde edilen istatistiksel güç oranları değişimi yer almaktadır. Şekil 15 incelendiğinde çok boyutlu kestirimde istatistiksel güç oranlarının tek boyutlu kestirimden daha yüksek olduğu görülmektedir. Teknikler ayrı ayrı incelendiğinde; çok boyutlu kestirimde istatistiksel güç oranları açısından tek boyutlu kestirime göre en fazla değişimin Lord'un χ^2 tekniğinde; en az değişimin ise LR tekniğinde olduğu görülmektedir. Şekil 15 incelendiğinde, çok boyutlu kestirim sonucunda tekniklere ait istatistiksel güç oranlarının, tek boyutlu kestirime göre birbirine daha yaklaştığı görülmektedir. Bu durum '*Kestirim*' x '*Teknik*' ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumludur. TB DMF türü için istatistiksel güç oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'Boyut sayısı'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 19'da verilmiştir.

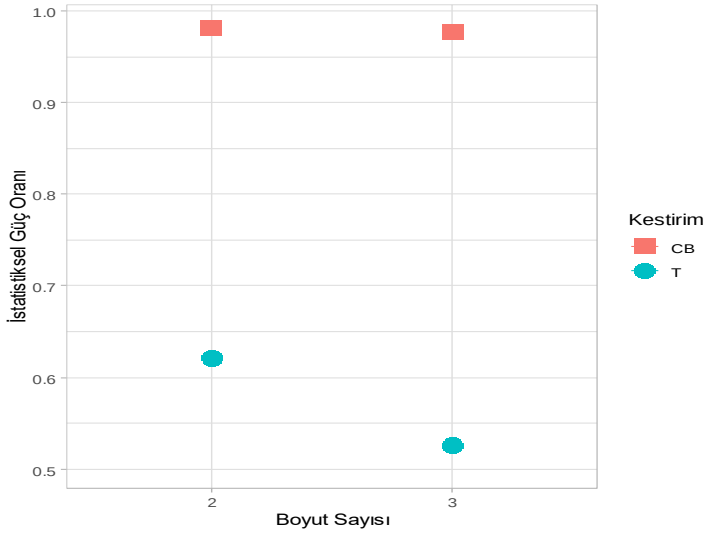
Tablo 19

TB DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'Boyut Sayısı' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxBoyut Sayısı	Fark	Standart hata	t (Tukey)	p
ÇB BS2 - TB BS2	0.36978		36.987	.000
ÇB BS2 - ÇB BS3	0.00561		0.561	.943
ÇB BS2 - TB BS3	0.47144		47.156	.000**
TB BS2 - ÇB BS3	-0.36417	0.01	-36.426	.000**
TB BS2 - TB BS3	0.10167		10.169	.000**
ÇB BS3 - TB BS3	0.46583		46.595	.000**

*p<.05 **p<.01

Tablo 19 incelendiğinde, TB DMF türü için istatistiksel güç oranları üzerinde ÇB BS2 - TB BS2 (t=36.987; p<.01) ve ÇB BS3 - TB BS3 (t=46.595; p<.01) tüm boyut sayıları için tek ve çok boyutlu kestirimler arasında, ÇB kestirim lehine manidar farklılıklar bulunmuştur. Diğer bir ifadeyle '*Kestirim*' x '*Boyut sayısı'* ikili etkileşimine yönelik sonuçlar incelendiğinde, DMF yüzdesinin iki ve üç olduğu durumlarda, çok boyutlu kestirimin istatistiksel güç oranlarının tek boyutlu kestirimden manidar olarak farklılaştığı görülmektedir. '*Kestirim*' ile *'Boyut sayısı'* ikili etkileşime yönelik sonuçlar incelendiğinde, *'Boyut sayısı'* nın iki ve üç olduğu durumların hepsinde, çok boyutlu kestirimde, tek boyutlu kestirimden daha yüksek istatistiksel güç oranları elde edilmiştir. TB DMF türü için istatistiksel güç oranları üzerinde ÇB BS2 - ÇB BS3 (t=0.561; p>.01) ÇB kestirimler için boyut sayıları arasında manidar farklılıklar olmadığı görülmektedir. Diğer bir ifadeyle ÇB kestirim sonucunda boyut sayıları birbirine benzer çalışmaktadır. TB DMF türü için istatistiksel güç oranları üzerinde TB BS2 - TB BS3 (t=10.169; p<.01) tüm TB kestirimler için boyut sayıları arasında manidar farklılıklar olduğu görülmektedir. TB DMF türü için *'Kestirim' x 'Boyut sayısı'* ikili etkileşimlerine ait istatistiksel güç oranları değişimleri Şekil 16'da verilmiştir.



Şekil 16. Kestirim ile Boyut Sayısı Etkileşimi için İstatistiksel Güç Oranları

Şekil 16'da boyut sayılarına göre tek ve çok boyutlu kestirimlerden elde edilen istatistiksel güç oranlarının değişimi yer almaktadır. Şekil 16 incelendiğinde çok boyutlu kestirimde istatistiksel güç oranlarının tek boyutlu kestirimden daha yüksek olduğu görülmektedir. '*Boyut sayısı*'na ait koşullar ayrı ayrı incelendiğinde çok boyutlu

kestirimde istatistiksel güç oranları açısından tek boyutlu kestirime göre en fazla değişimin üç boyutlu; en az değişimin ise iki boyutlu koşulda olduğu görülmektedir. Diğer bir ifadeyle boyutluluğun artması, çok boyutlu kestirimin tek boyutlu kestirime göre istatistiksel güç oranlarında sağladığı artışın artmasını sağlamıştır. Şekil 16 incelendiğinde çok boyutlu kestirim sonucunda boyut sayılarına ait istatistiksel güç oranlarının, tek boyutlu kestirime göre birbirine daha yaklaştığı görülmektedir. Diğer bir ifadeyle çok boyutlu kestirim sonucunda boyutlar birbirine benzer çalışmaktadır. Bu durum '*Kestirim*' x '*Boyut sayısı*' ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumludur. TB DMF türü için istatistiksel güç oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'DMF büyüklüğü'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 20'de verilmiştir.

Tablo 20

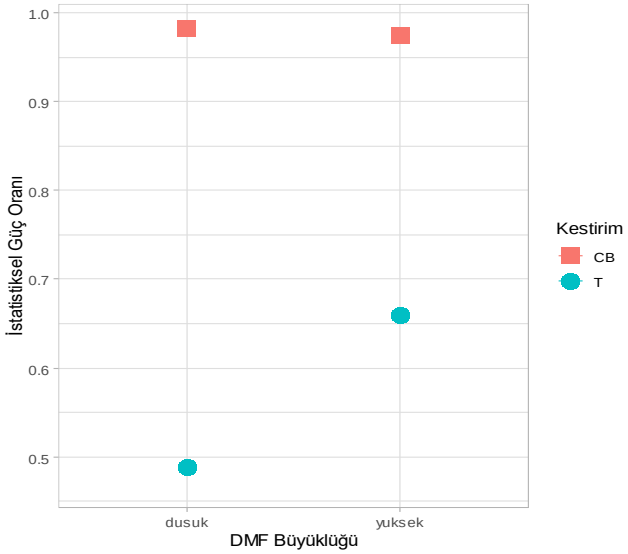
TB DMF Türü İçin Tip I Hata Oranlarının 'DMF Büyüklüğü' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxDMF Büyüklüğü	Fark	Standart hata	(t) Tukey	p
ÇB düşük - TB düşük	0.49139		49.151	.000**
ÇB düşük - ÇB yüksek	0.00439		0.439	.971
ÇB düşük - TB yüksek	0.34861		34.870	.000**
TB düşük - ÇB yüksek	-0.48700	0.01	-48.712	.000**
TB düşük - TB yüksek	-0.14278		-14.281	.000**
ÇB yüksek - TB yüksek	0.34422		34.431	.000**

*p<.05 **p<.01

Tablo 20 incelendiğinde, TB DMF türü için istatistiksel güç oranları üzerinde ÇB düşük - TB düşük (t=49.151; $p<.01$) ve ÇB yüksek - TB yüksek (t=34.431; $p<.01$) tüm DMF büyüklükleri için tek ve ve çok boyutlu kestirimler arasında, ÇB kestirim lehine manidar farklılıklar bulunmuştur. Diğer bir ifadeyle '*Kestirim*' x '*DMF büyüklüğü'* ikili etkileşime yönelik sonuçlar incelendiğinde, DMF yüzdesinin düşük ve yüksek olduğu durumlarda, çok boyutlu kestirimin istatistiksel güç oranlarının tek boyutlu kestirimden manidar olarak farklılaştığı görülmektedir. '*Kestirim*' ile '*DMF büyüklüğü'* ikili etkileşime yönelik sonuçlar incelendiğinde, '*DMF büyüklüğü'*nün düşük ve yüksek olduğu durumların hepsinde, çok boyutlu kestirimde, tek boyutlu kestirimden daha yüksek istatistiksel güç oranları elde edilmiştir. TB DMF türü için Tip I hata oranları üzerinde ÇB düşük - ÇB yüksek (t=0.439; $p>.01$) çok boyutlu kestirimler için boyut sayıları arasında manidar farklılıklar olmadığı görülmektedir. Diğer bir anlatımla çok

boyutlu kestirim tek başına değerlendirildiğinde, DMF büyüklüğü koşulunun tüm düzeylerinde çok boyutlu kestirim için istatistiksel güç oranları arasındaki farklılıklar manidar değildir. Bu kapsamda çok boyutlu kestirim sonucunda boyut sayılarının birbirine benzer çalıştığı belirtilebilir. TB DMF türü için istatistiksel güç oranları üzerinde TB düşük - TB yüksek (t=-14.281; p<.01) tüm tek boyutlu kestirimler için boyut sayıları arasında manidar farklılıklar olduğu görülmektedir. TB DMF türü için *'Kestirim' x 'DMF büyüklüğü'* ikili etkileşimlerine ait istatistiksel güç oranları Şekil 17'de verilmiştir.



Şekil 17. Kestirim ile DMF Büyüklüğü Etkileşimi için İstatistiksel Güç Oranları

Şekil 17'de DMF büyüklüklerine göre tek ve çok boyutlu kestirimlerden elde edilen istatistiksel güç oranlarının değişimi yer almaktadır. Şekil 17 incelendiğinde çok boyutlu kestirimde istatistiksel güç oranlarının tek boyutlu kestirimden daha yüksek olduğu görülmektedir. Alanyazında (Awour, 2008; González-Romá vd., 2006; Hidalgo ve Lopez-Pina, 2004; Kilmen, 2016; Kim, 2010; Magis ve De Boeck, 2012; Wyse ve Mapuranga, 2009) bu araştırmanın bulguları ile paralel olacak şekilde DMF büyüklüğü arttıkça istatistiksel güç oranlarının arttığı sonucuna varılmıştır. '*DMF büyüklüğü'ne* ait koşullar ayrı ayrı incelendiğinde çok boyutlu kestirimde istatistiksel güç oranları açısından tek boyutlu kestirime göre en fazla değişimin düşük; en az değişimin ise yüksek '*DMF büyüklüğü'* koşulunda olduğu görülmektedir. Diğer bir ifadeyle DMF büyüklüğündeki artış çok boyutlu kestirimin istatistiksel güç oranları üzerinde yaptığı düzeltmenin azalmasına neden olmuştur. Şekil 17 incelendiğinde çok boyutlu kestirim

sonucunda DMF büyüklüklerine ait istatistiksel güç oranlarının, tek boyutlu kestirime göre birbirine daha yaklaştığı görülmektedir. Diğer bir ifadeyle çok boyutlu kestirim sonucunda DMF büyüklükleri birbirine benzer çalışmaktadır. Bu durum '*Kestirim*' x '*DMF büyüklüğü'* ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumludur.

TBO DMF Türü için; Tek Boyutlu ve Çok Boyutlu Kestirimlerinin, DMF Belirleme Tekniği, Boyut Sayısı, DMF Yüzdesi, DMF Büyüklüğü Koşullarındaki Tip I Hata Oranlarının İncelenmesi

Bu inceleme kapsamında TBO DMF türü için tek boyutlu ve çok boyutlu kestirimlere ait performanslar Tip I hata oranları bakımından incelenmiş, elde edilen sonuçlar Tablo 21'de verilmiştir.

Tablo 21

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

	Kareler Toplamı	*Sd*	*F*	*p*	η^2	ηp^2
Kestirim (K)	0.052802	1	7222.372	.000	.571	.957
Teknik (T)	0.004030	2	27.5625	.000	.044	.633
DMF Yüzdesi (DY)	0.008249	2	56.4128	.000	.089	.779
Boyut Sayısı (BS)	0.000209	1	2.8555	.101	.002	.082
DMF Büyüklüğü (DB)	0.004894	1	66.9403	.000	.053	.676
K:T	0.003570	2	24.4190	.000	.039	.604
K:DY	0.006459	2	44.1742	.000	.070	.734
K:BS	0.000024	1	0.3256	.572	.003	.010
K:DB	0.004086	1	55.8906	.000	.044	.636
Hata	0.002339	32				

Tablo 21'de TBO DMF türü için Tip I hata oranları yer almaktadır. Buna göre '*Kestirim*' ($F(1,32) = 7222.372$, $p < .001$, $\eta^2 = .571$) ana etkilerinin manidar ve yüksek düzey bir etkiye sahip olduğu görülmektedir. Bununla birlikte '*DMF yüzdesi*' ana etkisinin ($F(1,32) = 56.4128$, $p < .001$, $\eta^2 = .089$) ise manidar ve orta düzey bir etkiye sahip olduğu görülmektedir. Buna karşın '*Teknik*' ($F(2,32) = 27.5625$, $p < .001$, $\eta^2 = .044$) ve '*DMF büyüklüğü'* ($F(1,32) = 66.9403$, $p < .001$, $\eta^2 = .053$) ana etkilerinin ise manidar ve düşük düzey bir etkiye sahip olduğu görülmektedir. '*Boyut sayısı*' ($F(1,32) = 2.8555$, $p > .05$, $\eta^2 = .002$) ana etkisinin ise manidar bir etkiye sahip olmadığı görülmektedir. '*Kestirim*' ile ikili etkileşimler incelendiğinde; '*Kestirim*' x *'Teknik'* ($F(2,32) = 24.4190$,

$p < .001$, $\eta p^2 = .604$), '*Kestirim*' x *'DMF yüzdesi'* ($F(2,32)=44.1742$, $p < .001$, $\eta p^2 = .734$) ve '*Kestirim*' x '*DMF büyüklüğü'* ($F(1,32)=55.8906$, $p < .001$, $\eta p^2 = .636$) ikili etkileşimlerinin manidar ve yüksek düzey bir etkiye sahip olduğu görülmektedir. '*Kestirim*' x '*Boyut sayısı'* ($F(1,32) = 0.3256$, $p>.005$, $\eta p^2=.010$) ikili etkileşiminin ise manidar bir etkiye sahip olmadığı görülmektedir.

TBO DMF türü dikkate alındığında, Tip I hata üzerindeki en önemli ana etkinin '*Kestirim*'e ait olduğu belirtilebilir. Benzer şekilde '*Teknik*', '*DMF yüzdesi*' ve '*DMF büyüklüğü'* ana etkilerinin de Tip I hata oranları üzerinde orta düzey etkiye sahip olduğu görülmektedir. Buna karşın; '*Boyut sayısı'* faktörünün hem ana etkileri hem de '*Kestirim*' ile etkileşimlerinin manidar olmaması nedeniyle, TBO DMF gösteren maddelerin Tip I hata oranları üzerinde bu faktörün etkili olmadığı ifade edilebilir.

TBO DMF türü dikkate alındığında, Tip I hata üzerinde neredeyse tüm etkileşim etkileri orta düzey etkiye sahip olmakla birlikte; en iyi ikili etkileşim etkisinin '*Kestirim*' ile '*DMF yüzdesi*' arasında olduğu belirtilebilir. Bunu sırasıyla '*Kestirim*' ile '*DMF büyüklüğü'* ve '*Kestirim'* ile *'Teknik'* izlemektedir. Bu doğrultuda TBO DMF gösteren maddelerin Tip I hata oranları üzerinde '*Kestirim*' ile '*DMF yüzdesi*', '*DMF büyüklüğü'* ve *'Teknik'* faktörlerinin birlikte etkili olduğu ifade edilebilir. '*Teknik*' ana etkisi manidar olmakla birlikte düşük düzey bir etkiye sahipken; '*Kestirim*' ile etkileşimlerini veren '*Kestirim*' ile '*Teknik*' ikili etkileşimlerinin ise manidar ve yüksek düzey etkiye sahip olduğu görülmektedir. Bu doğrultuda TBO DMF gösteren maddelerin Tip I hata oranları üzerinde '*Kestirim*' ile '*Teknik*' faktörlerinin birlikte daha etkili olduğu ifade edilebilir.

'*DMF yüzdesi*' ana etkisi manidar ve orta düzey etkili iken; '*Kestirim*' ile etkileşimlerini veren '*Kestirim*' ile '*DMF yüzdesi*' ikili etkileşimlerinin ise manidar ve yüksek düzey bir etkiye sahip olduğu görülmektedir. Bu doğrultuda TBO DMF gösteren maddelerin Tip I hata oranları üzerinde '*Kestirim*' ile '*DMF yüzdesi*' faktörlerinin birlikte daha etkili olduğu ifade edilebilir. '*Boyut sayısı'* ana etkisi manidar değilken; '*Kestirim*' ile etkileşimlerini veren '*Kestirim*' ile '*Boyut sayısı'* ikili etkileşimlerinin de manidar olmadığı görülmektedir. Bu doğrultuda TBO DMF gösteren maddelerin Tip I hata oranları üzerinde '*Kestirim*' ile '*Boyut sayısı'* faktörlerinin birlikte etkili olmadığı ifade edilebilir.

'DMF büyüklüğü' ana etkisi manidar olmakla birlikte düşük düzey bir etkiye sahipken; '*Kestirim*' ile etkileşimlerini veren '*Kestirim*' ile '*DMF büyüklüğü'* ikili etkileşimlerinin ise manidar ve yüksek düzey etkiye sahip olduğu görülmektedir. Bu doğrultuda TBO DMF gösteren maddelerin Tip I hata oranları üzerinde '*Kestirim*' ile

'*DMF büyüklüğü'* faktörlerinin birlikte daha etkili olduğu ifade edilebilir. Faktöriyel ANOVA sonucu manidar olan ve orta-yüksek düzey etkiye sahip olan ana faktörlere yönelik ileri incelemeler için; '*stats*' (R Core Team, 2019) paketinde yer alan '*TukeyHSD*' fonksiyonundan yararlanılarak Tukey testi gerçekleştirilmiştir. TBO DMF türü için Tip I hata oranları üzerinde manidar etkiye sahip olan '*Kestirim'* yönteminin ikili karşılaştırmaları Tablo 22'de verilmiştir.

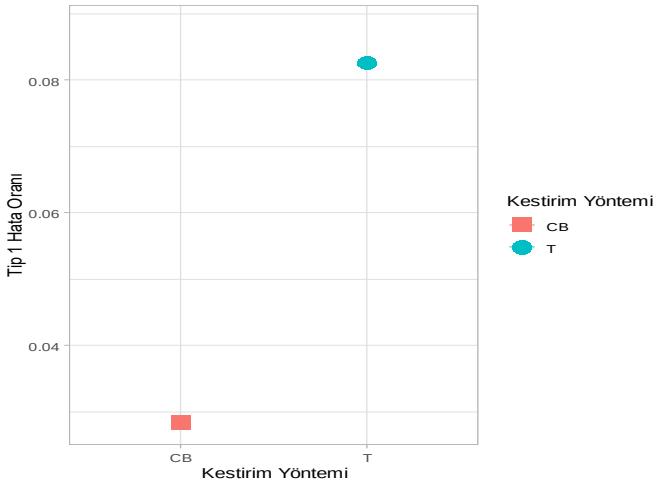
Tablo 22

TBO DMF Türü İçin Tip I Hata Oranlarının Kestirim Yöntemine Göre İkili Karşılaştırmaları

Kestirim yöntemi	Fark	Standart hata	(t) Tukey	p
ÇB – T	-0.0542	0.002	-26.874	.001**

*p<.05 **p<.01; CB: Çok Boyutlu; T: Tek Boyutlu

Tablo 22 incelendiğinde TB DMF Türü için Tip I hata oranları üzerinde Çok Boyutlu (ÇB) kestirim yöntemi lehine ve manidardır (t=-26.874; $p<.01$). ÇB kestirim yöntemi, Tek Boyutlu (TB) kestirim yöntemine göre daha düşük Tip 1 hata değerleri üretmiştir. TB DMF türü için kestirim yöntemlerine ait Tip I hata değişimleri Şekil 18'de verilmiştir.



Şekil 18. Kestirim Ana Etkisi için Tip I Hata Oranları

Şekil 18'de tek ve çok boyutlu kestirimlerden elde edilen Tip I hata oranlarının değişimi yer almaktadır. Şekil 18 incelendiğinde, çok boyutlu kestirimlerden elde edilen

Tip I hata oranlarının tek boyutlu kestirimlere göre düştüğü görülmektedir. TBO DMF türü için Tip I hata oranları üzerinde manidar etkiye sahip olan *'DMF yüzdesi'* yönteminin ikili karşılaştırmaları Tablo 23'de verilmiştir.

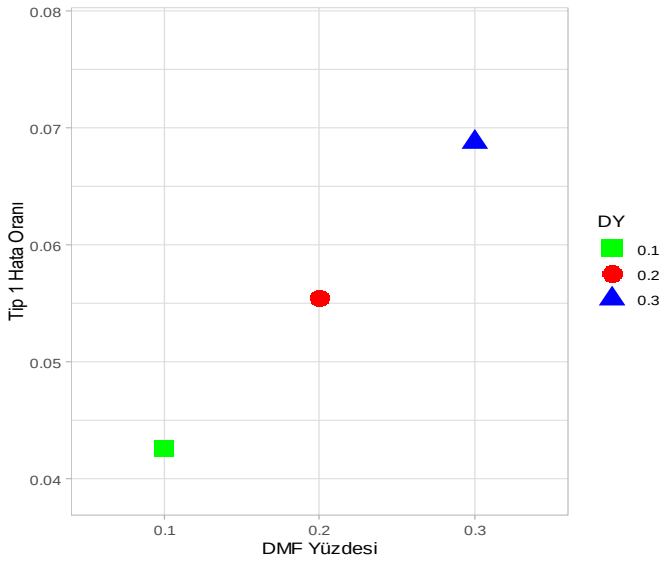
Tablo 23

TBO DMF Türü İçin Tip I Hata Oranlarının DMF Yüzdelerine Göre İkili Karşılaştırmaları

DMF Yüzdeleri	Fark	Standart hata	t (Tukey)	p
%10 - %20	-0.0129	0.00247	-5.223	.000**
%10 - %30	-0.0262	0.00247	-10.621	.000**
%20 - %30	-0.0133	0.00247	-5.399	.000**

*p<.05 **p<.01

Tablo 23 incelendiğinde, TBO DMF türü için Tip I hata oranları üzerinde %10 - %20 (*t*=--5.223; *p*<*.01*), %10 - %20 (*t*=-10.621; *p*<*.01*) ve %10 - %20 (*t*=-5.399; *p*<*.01*) DMF yüzdeleri arasında manidar farklılıklar bulunmuştur. DMF yüzdesi arttıkça daha yüksek Tip I hata değerleri üretmiştir. TBO DMF Türü için DMF yüzdelerine ait Tip I hata değişimleri Şekil 19'da verilmiştir.



Şekil 19. DMF Yüzdesi Ana Etkisi için Tip I Hata Oranları

Şekil 19'da DMF yüzdelerinden elde edilen Tip I hata oranlarının değişimi yer almaktadır. Şekil 19 incelendiğinde '*DMF yüzdesi'* arttıkça Tip I hata oranlarının arttığı görülmektedir. Magis ve De Boeck (2011) de yaptıkları simülasyon çalışmasında DMF

gösteren madde sayısı arttıkça Tip I hata oranlarının önemli ölçüde büyüdüğünü, özellikle büyük örneklemlerde bu büyümenin çok fazla olduğunu ve arındırma işleminin bile bu büyümeyi kontrol altına alamadığını belirtmişlerdir.

En yüksek Tip I hata ortalamalarının %30; en düşük Tip I hata ortalamalarının ise %10 yönteminde olduğu görülmektedir. Faktöriyel ANOVA sonucu manidar olan ve orta-yüksek düzey etkiye sahip olan ikili etkileşimlere yönelik ileri incelemeler için; '*stats*' (R Core Team, 2019) paketinde yer alan '*TukeyHSD*' fonksiyonundan yararlanılarak Tukey testi gerçekleştirilmiştir. TB DMF türü için Tip I hata oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'Teknik'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 24'te verilmiştir.

Tablo 24

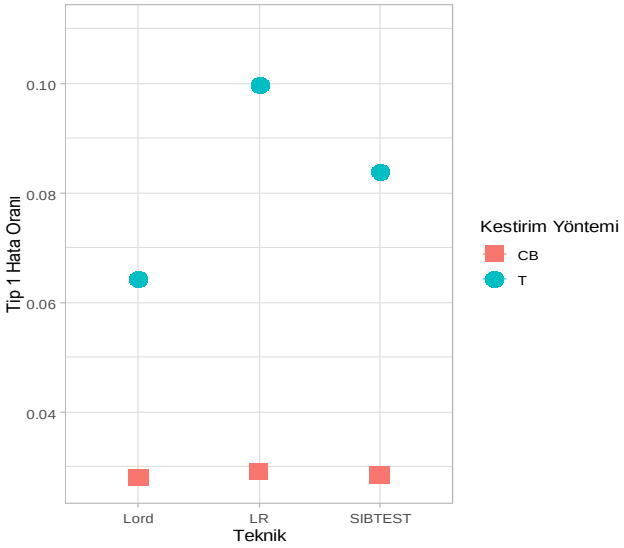
TBO DMF Türü İçin Tip I Hata Oranlarının 'Kestirim' x 'Teknik' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxTeknik	Fark	Standart hata	t (Tukey)	p
ÇB Lord - TB Lord	-0.036267		-10.390	.000
ÇB Lord - ÇB LR	-0.001083		-0.310	.999
ÇB Lord - TB LR	-0.071767		-20.560	.000**
ÇBLord - ÇB SIBTEST	-0.000483		-0.138	.999
ÇB Lord - TB SIBTEST	-0.056017		-16.048	.000**
TB Lord - ÇB LR	0.035183		10.079	.000**
TB Lord - TB LR	-0.035500	0.00349	-10.170	.000**
TB Lord - ÇB SIBTEST	0.035783		10.251	.000**
TB Lord - TB SIBTEST	-0.019750		-5.658	.000**
ÇB LR - TB LR	-0.070683		-20.249	.000**
ÇB LR - ÇB SIBTEST	0.000600		0.172	.999**
ÇB LR - TB SIBTEST	-0.054933		-15.737	.000**
TB LR - ÇB SIBTEST	0.071283		20.421	.000**
TB LR - TB SIBTEST	0.015750		4.512	.001**
ÇBSIBTEST-TB SIBT.	-0.055533		-15.909	.000**

*p<.05 **p<.01; CB: Çok Boyutlu; T: Tek Boyutlu

Tablo 24 incelendiğinde, TB DMF türü için Tip I hata oranları üzerinde ÇB Lord - TB Lord'un χ^2 (*t*=-10.390; *p*<*.01*), ÇB LR - TB LR (*t*=-20.249; *p*<*.01*) ve ÇB SIBTEST - TB SIBTEST (*t*=-15.909; *p*<*.01*) tüm yöntemler için tek ve çok boyutlu kestirimler arasında, ÇB kestirim lehine manidar farklılıklar bulunmuştur. ÇB kestirim sonucu tüm teknikler için tek boyutlu kestirimden daha düşük Tip I hata oranları elde edilmiştir. TBO DMF türü için Tip I hata oranları üzerinde ÇB Lord - ÇB LR (*t*=-0.310; *p*>*.01*), ÇB Lord'un χ^2 - ÇB SIBTEST (*t*=-0.138; *p*>*.01*) ve ÇB LR - ÇB SIBTEST (*t*=-15.909; *p*>*.01*) tüm ÇB kestirimler için teknikler arasında manidar farklılıklar olmadığı

görülmektedir. Diğer bir ifadeyle çok boyutlu kestirim sonucunda teknikler birbirine benzer çalışmaktadır. TBO DMF türü için Tip I hata oranları üzerinde TB Lord'un χ^2 - TB LR (t=10.170; p<.01), TB Lord'un χ^2 - TB SIBTEST (t=-5.658; p<.01) ve TB LR - TB SIBTEST (t=4.512; p<.01) tüm tek boyutlu kestirimler için teknikler arasında manidar farklılıklar olduğu görülmektedir. TB DMF türü için '*Kestirim*' x '*Teknik*' ikili etkileşimlerine ait Tip I hata oranları Şekil 20'de verilmiştir.



Şekil 20. Kestirim ile Teknik Etkileşimi için Tip I Hata Oranları

Şekil 20'de tekniklere göre tek ve çok boyutlu kestirimlerden elde edilen Tip I hata oranları değişimi yer almaktadır. Şekil 20 incelendiğinde çok boyutlu kestirimde Tip I hata oranlarının tek boyutlu kestirimden daha düşük olduğu görülmektedir. Teknikler ayrı ayrı incelendiğinde; çok boyutlu kestirimde Tip I hata oranları açısından tek boyutlu kestirime göre en fazla değişimin LR tekniğinde; en az değişimin ise Lord'un χ^2 tekniğinde olduğu görülmektedir. Bulut ve Suh (2017) da LR'nin TBO DMF için kullanılabileceğini belirtmişlerdir.

Şekil 20 incelendiğinde çok boyutlu kestirim sonucunda tekniklere ait Tip I hata oranlarının, tek boyutlu kestirime göre birbirine daha yaklaştığı görülmektedir. Bu durum '*Kestirim*' x '*Teknik*' ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumludur. Cohen ve Kim (1993) de çalışmasında bu çalışmaya benzer olarak yeterli örneklem büyüklüğünde ve yeterince uzun testlerde farklı tekniklerde benzer sonuçların elde edileceğini ortaya koymuştur. TBO DMF türü için Tip I hata oranları üzerinde

manidar etkiye sahip olan *'Kestirim' x 'DMF yüzdesi'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 25'te verilmiştir.

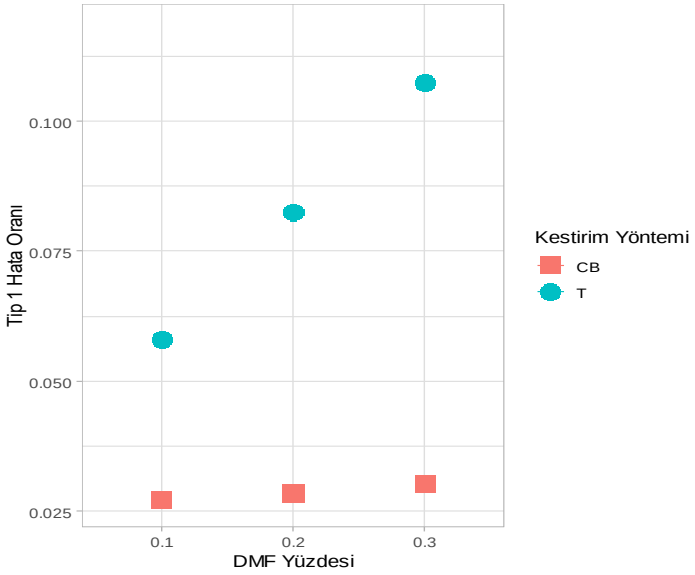
Tablo 25

TBO DMF Türü İçin Tip I Hata Oranlarının 'Kestirim' x 'DMF Yüzdesi' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxDMF Yüzdesi	Fark	Standart hata	t (Tukey)	p
ÇB %10 - TB %10	-0.03090		-8.852	.000
ÇB %10 - ÇB %20	-0.00120		-0.344	.999
ÇB %10 - TB %20	-0.05548		-15.895	.000**
ÇB%10 - ÇB %30	-0.00302		-0.864	.952
ÇB %10 - TB %30	-0.08032		-23.009	.000**
TB %10 - ÇB %20	0.02970		8.508	.000**
TB %10 - TB %20	-0.02458	0.00349	-7.043	.000**
TB %10 - ÇB %30	0.02788		7.988	.000**
TB %10 - TB %30	-0.04942		-14.157	.000**
ÇB %20 - TB %20	-0.05428		-15.551	.000**
ÇB%20 - ÇB %30	-0.00182		-0.520	.995**
ÇB %20 - TB %30	-0.07912		-22.665	.000**
TB %20 - ÇB %30	0.05247		15.031	.000**
TB %20 - TB %30	-0.02483		-7.114	.001**
ÇB %30 - TB %30	-0.07730		-22.145	.000**

*p<.05 **p<.01; ÇB: Çok Boyutlu; TB: Tek Boyutlu

Tablo 25 incelendiğinde TBO DMF türü için Tip I hata oranları üzerinde ÇB %10 - TB %10 (*t*=-8.852; *p*<*.01*), ÇB %20 - TB %20 (*t*=-15.551; *p*<*.01*) ve ÇB %30 - TB %30 (*t*=-22.145; *p*<*.01*) olmak üzere tüm DMF yüzdeleri için tek ve çok boyutlu kestirimler arasında, ÇB kestirim lehine manidar farklılıklar bulunmuştur. Diğer bir ifadeyle '*Kestirim*' x '*DMF yüzdesi'* ikili etkileşime yönelik sonuçlar incelendiğinde, DMF yüzdesinin %10, %20 ve %30 olduğu durumların hepsinde çok boyutlu kestirim sonucu tek boyutlu kestirimden daha düşük Tip I hata oranları elde edilmiştir. TBO DMF türü için Tip I hata oranları üzerinde ÇB %10 - ÇB %20 (*t*=0.344; *p*>*.01*), ÇB %10 - ÇB %30 (*t*=-0.864; *p*>*.01*) ve ÇB %20 - ÇB %30 (*t*=-0.520; *p*>*.01*) tüm çok boyutlu kestirimler için DMF yüzdeleri arasında manidar farklılıklar olmadığı görülmektedir. Diğer bir ifadeyle çok boyutlu kestirim sonucunda DMF yüzdeleri birbirine benzer çalışmaktadır. TBO DMF türü için Tip I hata oranları üzerinde TB %10 - TB %20 (*t*=-7.043; *p*<*.01*), TB %10 - TB %30 (*t*=-14.157; *p*<*.01*) ve TB %20 - TB %30 (*t*=-7.114; *p*<*.01*) olmak üzere tüm tek boyutlu kestirimler için DMF yüzdeleri arasında manidar farklılıklar olduğu görülmektedir. TBO DMF türü için *'Kestirim' x 'DMF yüzdesi'* ikili etkileşimlerine ait Tip I hata oranları Şekil 21'de verilmiştir.



Şekil 21. Kestirim ile DMF Yüzdesi Etkileşimi için Tip I Hata Oranları

Şekil 21'de DMF yüzdelerine göre tek ve çok boyutlu kestirimlerden elde edilen Tip I hata oranları yer almaktadır. Şekil 21 incelendiğinde çok boyutlu kestirimde Tip I hata oranlarının tek boyutlu kestirimden daha düşük olduğu görülmektedir. TBO DMF türü için '*DMF yüzdesi'* oranlarındaki artış Tip I hata oranlarında çok boyutlu kestirimine ait düzeltmenin artmasını sağlamıştıtr. '*DMF yüzdesi'* arttıkça tüm yöntemlerin Tip I hata oranlarının arttığı bulgusu, alanyazındaki sonuçlarla uyumludur (Fidalgo ve ark., 2000; Finch, 2005). Yapılan bu çalışmaya benzer olarak alanyazındaki çok sayıda çalışmada DMF gösteren madde sayısındaki artışın daha yüksek oranda Tip I hataya neden olduğu ortaya konmuştur (Atalay Kabasakal vd., 2014; Bauer vd., 2011; De Boeck, 2012; Finch, 2005; Guilera vd., 2013; Stark vd., 2006; Tunç ve Kutlu, 2018; Wang ve Yeh, 2003). Şekil 21 incelendiğinde çok boyutlu kestirim sonucunda DMF yüzdelerine ait Tip I hata oranlarının, tek boyutlu kestirime göre birbirine daha yaklaştığı görülmektedir. Bu durum '*Kestirim*' x '*DMF yüzdesi*' ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumludur. TBO DMF türü için tip I hata oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'DMF büyüklüğü'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 26'da verilmiştir.

Tablo 26

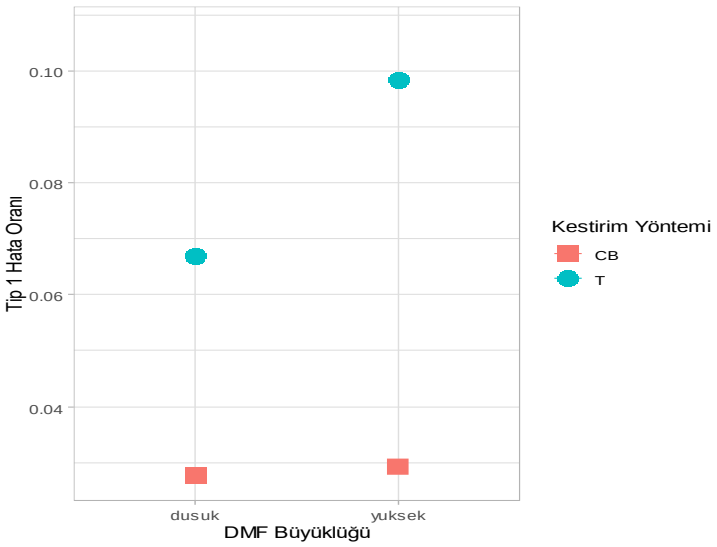
TBO DMF Türü İçin Tip I Hata Oranlarının 'DMF Büyüklüğü' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxDMF Büyüklüğü	Fark	Standart hata	t (Tukey)	p
ÇB düşük - TB düşük	-0.03909		-13.717	.000
ÇB düşük - ÇB yüksek	-0.00142		-0.499	.958
ÇB düşük - TB yüksek	-0.07065		-24.788	.000**
TB düşük - ÇB yüksek	0.03767	0.00285	13.218	.000**
TB düşük - TB yüksek	-0.03156		-11.072	.000**
ÇB yüksek - TB yüksek	-0.06923		-24.289	.000**

*p<.05 **p<.01; CB: Çok Boyutlu; T: Tek Boyutlu

Tablo 26 incelendiğinde TB DMF türü için Tip I hata oranları üzerinde ÇB düşük - TB düşük (t=-18.600; $p<.01$) ve ÇB yüksek - TB yüksek (t=-19.407; $p<.01$) olmak üzere tüm DMF büyüklükleri için tek ve çok boyutlu kestirimler arasında, çok boyutlu kestirim lehine manidar farklılıklar bulunmuştur. Diğer bir ifadeyle '*Kestirim*' x '*DMF büyüklüğü'* ikili etkileşimine yönelik sonuçlar incelendiğinde, DMF yüzdesinin düşük ve yüksek olduğu durumlarda, çok boyutlu kestirimin Tip I hata ortalamalarının tek boyutlu kestirimden manidar olarak farklılaştığı görülmektedir. '*Kestirim*' ile '*DMF büyüklüğü'* ikili etkileşimine yönelik sonuçlar incelendiğinde, '*DMF büyüklüğü'*nün düşük ve yüksek olduğu durumların hepsinde, çok boyutlu kestirimde, tek boyutlu kestirimden daha düşük Tip I hata oranları elde edilmiştir.

TBO DMF türü için Tip I hata oranları üzerinde ÇB düşük - ÇB yüksek (t=0.499; $p>.01$) olmak üzere çok boyutlu kestirimler için DMF büyüklükleri arasında manidar farklılıklar olmadığı görülmektedir. Diğer bir anlatımla çok boyutlu kestirim tek başına değerlendirildiğinde, DMF büyüklüğü koşulunun tüm düzeylerinde Tip I hata oranları arasındaki fark manidar değildir. Diğer bir ifadeyle boyut sayılarının birbirine benzer çalışmaktadır. TBO DMF türü için Tip I hata oranları üzerinde TB düşük - TB yüksek (t=-11.072; $p<.01$) olmak üzere tüm tek boyutlu kestirimler için boyut sayıları arasında manidar farklılıklar olduğu görülmektedir. TBO DMF türü için *'Kestirim' x 'DMF büyüklüğü'* ikili etkileşimlerine ait Tip I hata oranları Şekil 22'de verilmiştir.



Şekil 22. Kestirim ile DMF Büyüklüğü Etkileşimi için Tip I Hata Oranları

Şekil 22'de DMF büyüklüklerine göre tek ve çok boyutlu kestirimlerden elde edilen Tip I hata oranlarının değişimi yer almaktadır. DMF büyüklüğü arttıkça hem tek hem de çok boyutlu kestirimlerde Tip I hata oranlarının arttığı bulgusu alanyazında (Kim, 2010; Kilmen, 2016; Magis ve De Boeck, 2012; Tunç ve Kutlu, 2018) desteklenmektedir. Şekil 25 incelendiğinde çok boyutlu kestirimde Tip I hata oranlarının tek boyutlu kestirimden daha düşük olduğu; DMF büyüklüğünün artmasının, çok boyutlu kestirimin tek boyutlu kestirimine göre Tip I hata ortalamalarında sağladığı düşüşü arttırdığı görülmektedir. '*Kestirim*' x '*DMF büyüklüğü'* ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumlu olarak çok boyutlu kestirim sonucunda DMF büyüklüklerinin birbirine benzer çalıştığı görülmektedir.

TBO DMF Türü için; Tek Boyutlu ve Çok Boyutlu Kestirimlerin, DMF Belirleme Tekniği, Boyut Sayısı, DMF Yüzdesi, DMF Büyüklüğü Koşullarındaki İstatistiksel Güç Oranlarının İncelenmesi

Bu inceleme kapsamında TBO DMF türü için tek boyutlu ve çok boyutlu kestirimlere ait performanslar istatistiksel güç oranları bakımından incelenmiş, elde edilen sonuçlar Tablo 27'de verilmiştir.

Tablo 27

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

	Kareler Toplamı	*Sd*	*F*	*p*	η^2	ηp^2
Kestirim (K)	3.14211	1	3492.9950	.000	.090	.991
Teknik (T)	0.13606	2	75.6244	.000	.012	.825
DMF Yüzdesi (DY)	0.01153	2	6.4083	.005	.000	.286
Boyut Sayısı (BS)	0.05179	1	57.5717	.000	.000	.643
DMF Büyüklüğü (DB)	0.08618	1	95.8058	.000	.008	.749
K:T	0.14143	2	78.6093	.000	.014	.830
K:DY	0.00206	2	6.8753	.332	.000	.066
K:BS	0.04152	1	46.1567	.000	.006	.591
K:DB	0.09746	1	108.3449	.000	.010	.772
Hata	0.02879	32				

Tablo 27'de TBO DMF türü için istatistiksel güç oranlarına ait sonuçlar yer almaktadır. Buna göre '*Kestirim*' ($F(1,32) = 3492.9950$, $p < .001$, $\eta^2 = .090$) ana etkisinin manidar ve orta düzey bir etkiye sahip olduğu görülmektedir. Buna ek olarak '*Teknik*' ($F(2,32) = 75.6244$, $p<.001$, $\eta^2= .012$) ve '*DMF büyüklüğü'* ($F(1,32) =95.8058$, $p < .001$, $\eta^2 = .008$) ana etkilerinin ise manidar olduğu ancak düşük düzey bir etkiye sahip olduğu görülmektedir. Buna karşın '*DMF yüzdesi*' ($F(2,32) = 6.4083$, $p<.05$, $\eta^2=.000$) ve '*Boyut sayısı'* ($F(1,32) = 57.5717$, $p < .001$, $\eta^2 = .000$) ana etkilerinin de manidar olduğu ancak bir etkiye sahip olmadığı görülmektedir.

'*Kestirim*' ile ikili etkileşimler incelendiğinde; '*Kestirim*' x '*Teknik*' ($F(2,32)=78.3336$, $p < .001$, $\eta p^2 = .830$), '*Kestirim*' x '*Boyut sayısı'* ($F(1,32) = 46.1567$, $p <.001$, $\eta p^2 = .591$) ve '*Kestirim*' x '*DMF büyüklüğü'* ($F(1,32) = 108.3449$, $p < .001$, $\eta p^2=.772$) ikili etkileşimlerinin manidar ve yüksek düzey bir etkiye sahip oldukları görülmektedir. Buna karşın '*Kestirim*' x '*DMF yüzdesi*' ($F(2,32) = 6.8753$, $p > .05$, $\eta p^2=.066$) ikili etkileşimlerinin ise manidar ancak orta düzey bir etkiye sahip olduğu görülmektedir. TBO DMF türü dikkate alındığında, istatistiksel güç oranları üzerindeki en önemli ana etkinin '*Kestirim*'e ait olduğu belirtilebilir. '*Kestirim*' ana etkisinin istatistiksel güç oranları üzerinde yüksek düzey; '*Teknik*' ana etkisinin ise düşük düzey etkiye sahip olduğu belirtilebilir. Buna karşın '*Boyut sayısı'* ve '*DMF büyüklüğü'* ana etkilerinin ise istatistiksel güç oranları üzerinde etkiye sahip olmadığı görülmektedir. TBO DMF türü dikkate alındığında, istatistiksel güç oranları üzerinde en önemli etkileşim etkilerinin, '*Kestirim*' ile '*Teknik*', '*Kestirim*' ile '*Boyut sayısı'* ve '*Kestirim*' ile '*DMF büyüklüğü'* ne ilişkin ikili etkileşimler olduğu belirtilebilir. Bu doğrultuda TBO DMF gösteren maddelerin istatistiksel güç oranları üzerinde '*Kestirim*' ile birlikte *'Teknik'*,

‘*Boyut sayısı’* ve ‘*DMF büyüklüğü’* faktörlerinin birlikte oldukça etkili olduğu ifade edilebilir. Buna karşın ‘*Kestirim*’ ile ‘*DMF yüzdesi*’ faktörlerinin birlikte etkileşimlerinin manidar olmadığı görülmektedir.

‘*Teknik*’ ana faktörünün TBO DMF gösteren maddelerin istatistiksel güç oranları üzerinde düşük düzey bir etkisi varken ‘*Kestirim*’ ile ikili etkileşimlerinin istatistiksel güç oranları üzerinde yüksek düzey bir etkiye sahip olduğu görülmektedir. ‘*DMF yüzdesi*’ ana faktörünün TBO DMF gösteren maddelerin istatistiksel güç oranları üzerinde bir etkisi olmadığı gibi ‘*Kestirim*’ ile ikili etkileşimlerinin de istatistiksel güç oranları üzerinde manidar bir etkisi olmadığı görülmektedir. ‘*Boyut sayısı’* ve ‘*DMF büyüklüğü’* ana faktörlerinin ise TBO DMF gösteren maddelerin istatistiksel güç oranları üzerinde bir etkisi yok iken; ‘*Kestirim*’ ile ikili etkileşimlerin istatistiksel güç oranları üzerinde yüksek düzey bir etkiye sahip olduğu görülmektedir. Faktöriyel ANOVA sonucu manidar olan ve orta-yüksek düzey etkiye sahip olan ana faktörlere yönelik ileri incelemeler için; ‘*stats*’ (R Core Team, 2019) paketinde yer alan ‘*TukeyHSD*’ fonksiyonundan yararlanılarak Tukey testi gerçekleştirilmiştir. TBO DMF türü için istatistiksel güç oranları üzerinde manidar etkiye sahip olan ‘*Kestirim’* yönteminin ikili karşılaştırmaları Tablo 28’de verilmiştir.

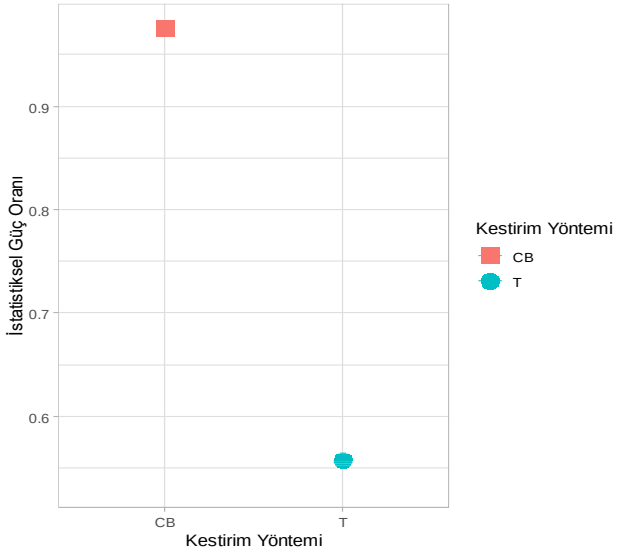
Tablo 28

TBO DMF Türü İçin İstatistiksel Güç Oranlarının Kestirim Yöntemine Göre İkili Karşılaştırmaları

Kestirim yöntemi	Fark	Standart hata	t (Tukey)	p
ÇB – T	0.418	0.00707	59.102	.001**

*p<.05 **p<.01; ÇB: Çok Boyutlu; TB: Tek Boyutlu

Tablo 28 incelendiğinde TBO DMF Türü için istatistiksel güç oranları üzerinde Çok Boyutlu (CB) kestirim yöntemi lehine ve manidardır (t=59.102; p<.01). ÇB kestirim yöntemi, Tek Boyutlu (TB) kestirim yöntemine göre daha yüksek istatistiksel güç değerleri üretmiştir. TBO DMF türü için ‘Kestirim’ faktörüne ait istatistiksel güç oranları Şekil 23’te verilmiştir.



Şekil 23. Kestirim Ana Etkisi için İstatistiksel Güç Oranları

Şekil 23'te tek ve çok boyutlu kestirimlerden elde edilen istatistiksel güç oranlarının değişimi yer almaktadır. Şekil 23 incelendiğinde çok boyutlu kestirimlerden elde edilen istatistiksel güç oranlarının tek boyutlu kestirimlere göre düştüğü; çok boyutlu kestirimlerin DMF'yi kontrol altına almada, tek boyutlu kestirimlerden daha etkili olduğu söylenbilir. Elde edilen bu bulgunun, tezin çıkış fikri olan, uygulanmış bir testte madde çıkarmadan çok boyutlu kestirim yapılarak düzeltilmiş toplam puan elde etme fikrini desteklediği şeklinde yorumlanabilir. Faktöriyel ANOVA sonucu manidar olan ve orta-yüksek düzey etkiye sahip olan ikili etkileşimlere yönelik ileri incelemeler için; '*stats*' (R Core Team, 2019) paketinde yer alan '*TukeyHSD*' fonksiyonundan yararlanılmış Tukey testi gerçekleştirilmiştir. TBO DMF Türü için istatistiksel güç oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'Teknik'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 29'da verilmiştir.

Tablo 29

TBO DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'Teknik' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxTeknik	Fark	Standart hata	t (Tukey)	p
ÇB Lord - TB Lord	0.539583		44.068	.000
ÇB Lord - ÇB LR	0.004417		0.361	.999
ÇB Lord - TB LR	0.387083		31.613	.000**
ÇB Lord - ÇB SIBTEST	0.000917	0.0122	0.075	.100
ÇB Lord - TB SIBTEST	0.332083		27.121	.000**
TB Lord - ÇB LR	-0.535167		-43.707	.000**
TB Lord - TB LR	-0.152500		-12.455	.000**

(Devam ediyor)

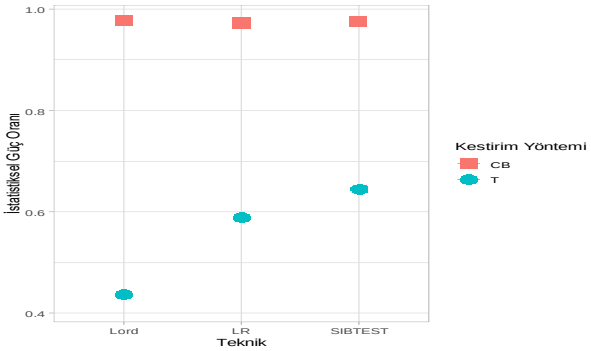
Tablo 29 (Devam)

TBO DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'Teknik' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxTeknik	Fark	Standart hata	t (Tukey)	p
TB Lord - ÇB SIBTEST	-0.538667		-43.993	.000**
TB Lord - TB SIBTEST	-0.207500		-16.947	.000**
ÇB LR - TB LR	0.382667		31.253	.000**
ÇB LR - ÇB SIBTEST	-0.003500	0.0122	-0.286	.999**
ÇB LR - TB SIBTEST	0.327667		26.761	.000**
TB LR - ÇB SIBTEST	-0.386167		-31.538	.000**
TB LR - TB SIBTEST	-0.055000		-4.492	.001**
ÇBSIBTEST-TB SIBT.	0.331167		27.046	.000**

*p<.05 **p<.01; ÇB: Çok Boyutlu; TB: Tek Boyutlu

Tablo 29 incelendiğinde TBO DMF türü için istatistiksel güç oranları üzerinde ÇB Lord'un χ^2 - TB Lord (t=44.068; p<.01), ÇB LR - TB LR (t=31.253; p<.01) ve ÇB SIBTEST - TB SIBTEST (t=27.046; p<.01) olmak üzere tüm yöntemler için tek ve çok boyutlu kestirimler arasında, ÇB kestirim lehine manidar farklılıklar bulunmuştur. Çok boyutlu kestirim sonucu tüm tekniklerde tek boyutlu kestirimden daha yüksek istatistiksel güç oranları elde edilmiştir. TBO DMF türü için istatistiksel güç oranları üzerinde ÇB Lord'un χ^2 - ÇB LR (t=0.361; p>.01), ÇB Lord'un χ^2 - ÇB SIBTEST (t=0.075; p>.01) ve ÇB LR - ÇB SIBTEST (t=-0.2869; p>.01) olmak üzere tüm çok boyutlu kestirimler için yöntemler arasında manidar farklılıklar olmadığı görülmektedir. Tablo 29 incelendiğinde çok boyutlu kestirim sonucunda tekniklerin birbirine benzer çalıştığı görülmektedir. TBO DMF türü için istatistiksel güç oranları üzerinde TB Lord'un χ^2 - TB LR (t=-12.455; p<.01), TB Lord'un χ^2 - TB SIBTEST (t=-16.947; p<.01) ve TB LR - TB SIBTEST (t=4.492; p<.01) tüm tek boyutlu kestirimler için teknikler arasında manidar farklılıklar olduğu görülmektedir. TBO DMF türü için '*Kestirim*' x '*Teknik*' ikili etkileşimlerine ait istatistiksel güç oranları değişimleri Şekil 24'te verilmiştir.



Şekil 24. Kestirim ile Teknik Etkileşimi için İstatistiksel Güç Oranları

Şekil 24'te tekniklere göre tek ve çok boyutlu kestirimlerden elde edilen istatistiksel güç oranlarının değişimi yer almaktadır. Şekil 24 incelendiğinde çok boyutlu kestirimde istatistiksel güç oranlarının tek boyutlu kestirimden daha yüksek olduğu; çok boyutlu kestirimde istatistiksel güç oranları açısından tek boyutlu kestirime göre en fazla değişimin Lord'un χ^2 yönteminde; en az değişimin ise LR yönteminde olduğu görülmektedir.

Şekil 24 incelendiğinde çok boyutlu kestirim sonucunda tekniklere ait istatistiksel güç oranlarının, tek boyutlu kestirime göre birbirine daha yaklaştığı görülmektedir. Bu durum '*Kestirim*' x '*Teknik*' ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumludur. TBO DMF türü için istatistiksel güç oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'DMF yüzdesi'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 30'da verilmiştir.

Tablo 30

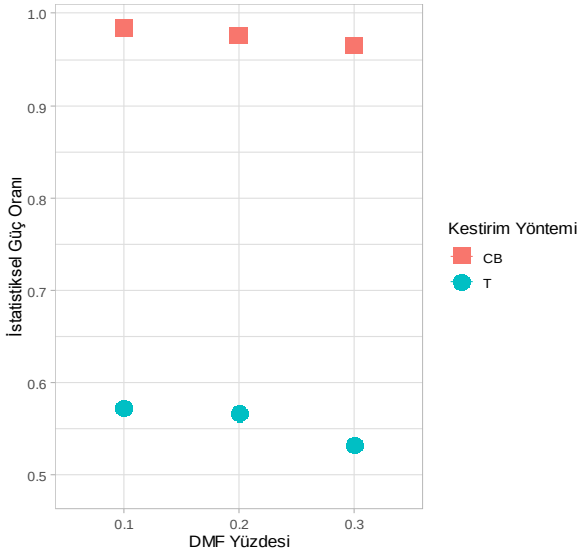
TBO DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'DMF Yüzdesi' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxDMF Yüzdesi	Fark	Standart hata	t (Tukey)	p
ÇB %10 - TB %10	0.41050		33.526	.000**
ÇB %10 - ÇB %20	0.00717		0.585	.991
ÇB %10 - TB %20	0.41717		34.070	.000**
ÇB %10 - ÇB %30	0.01842		1.504	.664
ÇB %10 - TB %30	0.45133		36.861	.000**
TB %10 - ÇB %20	-0.40333		-32.940	.000**
TB %10 - TB %20	0.00667		0.544	.994
TB %10 - ÇB %30	-0.39208	0.00349	-32.022	.000**
TB %10 - TB %30	0.04083		3.335	.242
ÇB %20 - TB %20	0.41000		33.485	.000**
ÇB %20 - ÇB %30	0.01125		0.919	.939
ÇB %20 - TB %30	0.44417		36.275	.000**
TB %20 - ÇB %30	-0.39875		-32.566	.000**
TB %20 - TB %30	0.03417		2.790	.085
ÇB %30 - TB %30	0.43292		35.356	.000**

*p<.05 **p<.01; ÇB: Çok Boyutlu; TB: Tek Boyutlu

Tablo 30 incelendiğinde TBO DMF türü için istatistiksel güç oranları üzerinde ÇB %10 - TB %10 (t=33.526; $p<.01$), ÇB %20 - TB %20 (t=34.070; $p<.01$) ve ÇB %30 - TB %30 (t=35.356; $p<.01$) olmak üzere tüm DMF yüzdeleri için tek ve çok boyutlu kestirimler arasında, çok boyutlu kestirim lehine manidar farklılıklar bulunmuştur. Diğer bir ifadeyle '*Kestirim*' x '*DMF yüzdesi'* ikili etkileşime yönelik sonuçlar incelendiğinde, DMF yüzdesinin %10, %20 ve %30 olduğu durumların hepsinde çok boyutlu kestirim

sonucu tek boyutlu kestirimden daha yüksek istatistiksel güç oranları elde edilmiştir. TBO DMF türü için istatistiksel güç oranları üzerinde ÇB %10 - ÇB %20 (t=0.585; p>.*01*), ÇB%10 - ÇB %30 (t=1.504; p>.*01*) ve ÇB%20 - ÇB %30 (t=0.919; p>.*01*) tüm çok boyutlu kestirimler için DMF yüzdeleri arasında manidar farklılıklar olmadığı görülmektedir. Diğer bir ifadeyle çok boyutlu kestirim sonucunda DMF yüzdeleri birbirine benzer çalışmaktadır. TBO DMF türü için istatistiksel güç oranları üzerinde TB %10 - TB %20 (t=0.544; p>.*01*), TB %10 - TB %30 (t=3.335; p>.*01*) ve TB %20 - TB %30 (t=2.790; p>.*01*) olmak üzere tüm tek boyutlu kestirimler için DMF yüzdeleri arasında manidar farklılıklar olmadığı görülmektedir. Diğer bir ifadeyle tek boyutlu kestirim sonucunda DMF yüzdeleri birbirine benzer çalışmaktadır. TBO DMF türü için *'Kestirim' x 'DMF yüzdesi'* ikili etkileşimlerine ait istatistiksel güç oranları Şekil 25'te verilmiştir.



Şekil 25. Kestirim ile DMF Yüzdesi Etkileşimi için İstatistiksel Güç Oranları

Şekil 25'te DMF yüzdelerine göre tek ve çok boyutlu kestirimlerden elde edilen istatistiksel güç oranlarının değişimi yer almaktadır. Şekil 25 incelendiğinde çok boyutlu kestirimde istatistiksel güç oranlarının tek boyutlu kestirimden daha yüksek olduğu; tek boyutlu kestirime göre en fazla değişimin %30 koşulunda; en az değişimin ise %20 koşulunda olduğu görülmektedir. Bununla birlikte çok boyutlu kestirimin düzeltmesi açısından bu koşulun düzeyleri arasında fazla farklılık olmadığı; DMF yüzdesi oranlarındaki artışın çok boyutlu kestirimin daha iyi düzeltme yapmasını sağladığı

görülmüştür. TBO DMF türü için istatistiksel güç oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'Boyut sayısı'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 31'de verilmiştir.

Tablo 31

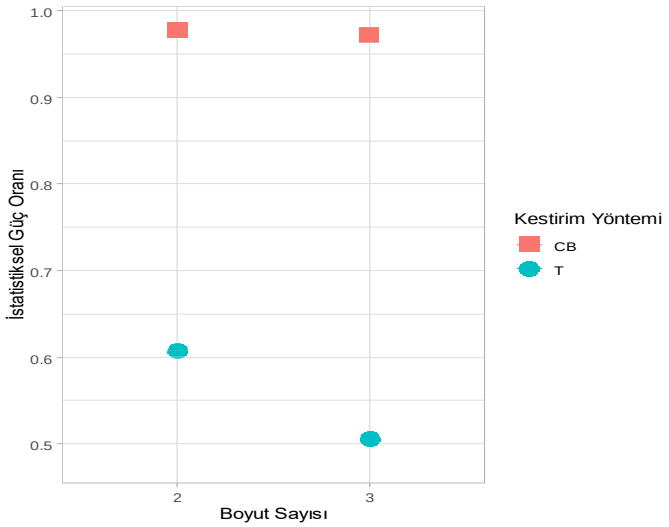
TBO DMF Türü İçin İstatistiksel Güç Oranlarının 'Kestirim' x 'Boyut Sayısı' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxBoyut Sayısı	Fark	Standart hata	t (Tukey)	p
ÇB BS2 - TB BS2	0.36978		36.987	.000
ÇB BS2 - ÇB BS3	0.00561		0.561	.942
ÇB BS2 - TB BS3	0.47144		47.156	.000**
TB BS2 - ÇB BS3	-0.36417	0.01	-36.426	.000**
TB BS2 - TB BS3	0.10167		10.169	.000**
ÇB BS3 - TB BS3	0.46583		46.595	.000**

*p<.05 **p<.01; ÇB: Çok Boyutlu; TB: Tek Boyutlu

Tablo 31 incelendiğinde TBO DMF türü için istatistiksel güç oranları üzerinde ÇB BS2 - TB BS2 (t=36.987; $p<.01$) ve ÇB BS3 - TB BS3 (t=46.595; $p<.01$) olmak üzere tüm boyut sayıları için tek ve çok boyutlu kestirimler arasında, ÇB kestirim lehine manidar farklılıklar bulunmuştur. Diğer bir ifadeyle '*Kestirim*' x '*Boyut sayısı'* ikili etkileşimine yönelik sonuçlar incelendiğinde, DMF yüzdesinin iki ve üç olduğu durumlarda, çok boyutlu kestirimin istatistiksel güç oranlarının tek boyutlu kestirimden manidar olarak farklılaştığı görülmektedir. '*Kestirim*' ile *'Boyut sayısı'* ikili etkileşime yönelik sonuçlar incelendiğinde, *'Boyut sayısı'*nın iki ve üç olduğu durumların hepsinde çok boyutlu kestirimde, tek boyutlu kestirimden daha yüksek istatistiksel güç oranları elde edilmiştir.

TBO DMF Türü için istatistiksel güç oranları üzerinde ÇB BS2 - ÇB BS3 (t=0.561; $p>.01$) olmak üzere çok boyutlu kestirimler için boyut sayıları arasında manidar farklılıklar olmadığ; boyut sayılarının birbirine benzer çalıştığı görülmektedir. TBO DMF türü için istatistiksel güç oranları üzerinde TB BS2 - TB BS3 (t=10.169; $p<.01$) olmak üzere tüm tek boyutlu kestirimler için boyut sayıları arasında manidar farklılıklar olduğu görülmektedir. TBO DMF türü için *'Kestirim' x 'Boyut sayısı'* ikili etkileşimlerine ait istatistiksel güç oranları değişimleri Şekil 26'da verilmiştir.



Şekil 26. Kestirim ile Boyut Sayısı Etkileşimi için İstatistiksel Güç Oranları

Şekil 26'da boyut sayılarına göre tek ve çok boyutlu kestirimlerden elde edilen istatistiksel güç oranlarının değişimi yer almaktadır. Şekil 26 incelendiğinde çok boyutlu kestirimde istatistiksel güç oranlarının tek boyutlu kestirimden daha yüksek olduğu; tek boyutlu kestirime göre en fazla değişimin üç boyutlu; en az değişimin ise iki boyutlu koşulda olduğu görülmektedir. Şekil 26 incelendiğinde çok boyutlu kestirim sonucunda boyut sayılarına ait istatistiksel güç oranlarının, tek boyutlu kestirime göre birbirine daha yaklaştığı görülmektedir. '*Kestirim*' x '*Boyut sayısı*' ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumlu olarak çok boyutlu kestirim sonucunda boyutlar birbirine benzer çalışmaktadır. TBO DMF türü için istatistiksel güç oranları üzerinde manidar etkiye sahip olan *'Kestirim' x 'DMF büyüklüğü'* ikili etkileşime yönelik Tukey Post-Hoc testi sonuçları Tablo 32'de verilmiştir.

Tablo 32

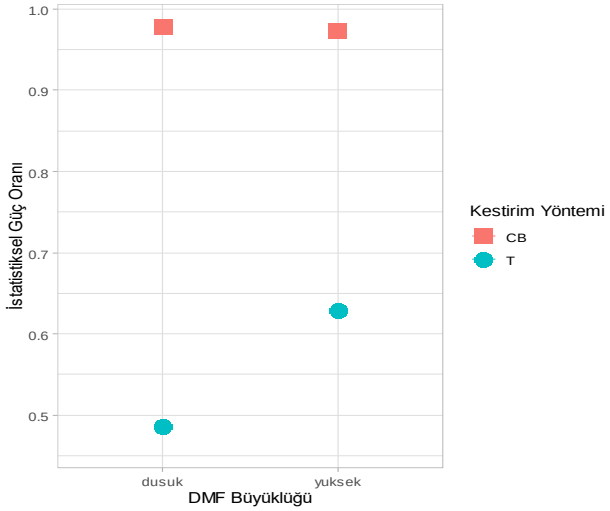
TBO DMF Türü İçin Tip I Hata Oranlarının 'DMF Büyüklüğü' İkili Etkileşime Göre İkili Karşılaştırmaları

KestirimxDMF Büyüklüğü	Fark	Standart hata	t (Tukey)	p
ÇB düşük - TB düşük	0.49139		49.151	.000**
ÇB düşük - ÇB yüksek	0.00439		0.439	.971
ÇB düşük - TB yüksek	0.34861		34.870	.000**
TB düşük - ÇB yüksek	-0.48700	0.01	-48.712	.000**
TB düşük - TB yüksek	-0.14278		-14.281	.000**
ÇB yüksek - TB yüksek	0.34422		34.431	.000**

*p<.05 **p<.01; ÇB: Çok Boyutlu; TB: Tek Boyutlu

Tablo 32 incelendiğinde TBO DMF türü için Tip I hata oranları üzerinde ÇB düşük - TB düşük (t=49.151; p<.01) ve ÇB yüksek - TB yüksek (t=34.431; p<.01) tüm DMF büyüklükleri için tek ve çok boyutlu kestirimler arasında, çok boyutlu kestirim lehine manidar farklılıklar bulunmuştur. Diğer bir ifadeyle '*Kestirim*' x '*DMF büyüklüğü'* ikili etkileşimine yönelik sonuçlar incelendiğinde, DMF büyüklüğünün düşük ve yüksek olduğu durumlarda, çok boyutlu kestirimin istatistiksel güç oranlarının tek boyutlu kestirimden manidar olarak farklılaştığı görülmektedir. '*Kestirim*' ile '*DMF büyüklüğü'* ikili etkileşimine yönelik sonuçlar incelendiğinde, '*DMF büyüklüğü'*nün düşük ve yüksek olduğu durumların hepsinde, çok boyutlu kestirimde, tek boyutlu kestirimden daha yüksek istatistiksel güç oranları elde edilmiştir.

TBO DMF türü için Tip I hata oranları üzerinde ÇB düşük - ÇB yüksek (t=0.439; p>.01) olmak üzere çok boyutlu kestirimler için boyut sayıları arasında manidar farklılıklar olmadığı görülmektedir. Diğer bir anlatımla çok boyutlu kestirim tek başına değerlendirildiğinde, DMF büyüklüğü koşulunun tüm düzeylerinde istatistiksel güç oranları arasında manidar farklılıklar yoktur. Bu kapsamda çok boyutlu kestirim sonucunda boyut sayılarının birbirine benzer çalıştığı belirtilebilir. TBO DMF türü için Tip I hata oranları üzerinde TB düşük - TB yüksek (t=-14.281; p<.01) olmak üzere tüm tek boyutlu kestirimler için boyut sayıları arasında manidar farklılıklar olduğu görülmektedir. TBO DMF türü için *'Kestirim' x 'DMF büyüklüğü'* ikili etkileşimlerine ait istatistiksel güç oranları Şekil 27'de verilmiştir.



Şekil 27. Kestirim ile DMF Büyüklüğü Etkileşimi için İstatistiksel Güç Oranları

Şekil 27'de DMF büyüklüklerine göre tek ve çok boyutlu kestirimlerden elde edilen istatistiksel güç oranları yer almaktadır. Şekil 27 incelendiğinde çok boyutlu kestirimde istatistiksel güç oranlarının tek boyutlu kestirimden daha yüksek olduğu bulgusu alanyazında (González-Romá vd., 2006; Hidalgo ve Lopez-Pina, 2004; ; Kilmen, 2016; Su ve Wang, 2005) desteklenmektedir. '*DMF büyüklüğü'*ne ait koşullar ayrı ayrı incelendiğinde çok boyutlu kestirimde istatistiksel güç oranları açısından tek boyutlu kestirime göre en fazla değişimin düşük; en az değişimin ise yüksek '*DMF büyüklüğü'* koşulunda olduğu görülmektedir. DMF büyüklüğündeki artışın çok boyutlu kestirimin istatistiksel güç oranları üzerinde yaptığı düzeltmenin azalmasına neden olduğu söylenebilir. '*Kestirim*' x '*DMF büyüklüğü'* ikili etkileşimlerine yönelik post-hoc karşılaştırmaları ile uyumlu olarak çok boyutlu kestirim sonucunda DMF büyüklükleri birbirine benzer çalışmaktadır.

Tek ve Çok Boyutlu Madde Tepki Kuramı Yetenek Parametre Kestirimlerine Ait Uyum İstatistiklerinin Değişiminin İncelenmesi

Yetenek parametrelerinin kestiriminde R dilinde yer alan *"mirt"* (Chalmers, 2012) paketinden yararlanılnıştır. Her bir koşul için ayrı ayrı, tek ve çok boyutlu kestirimler yapılmıştır. Model 1, tek; Model 2, iki ve Model 3, üç boyut için mirt kestirimlerini göstermektedir. Bu kestirimlere ait uyum istatistiklerinden AIC, BIC ve Log-olabilirlik değerleri, her bir koşulx100 tekrar için elde edilmiş; elde edilen değerlerin ortalamalarına ait sonuçlar Tablo 33'te verilmiştir.

Tablo 33

Tüm Koşular için Tek Boyutlu Kestirimlere Yönelik Uyum İstatistikleri Ortalamaları

		100 Tekrar İçin		
Üretilme Koşulu	Kestirimlerde Tan. Boyut Sayısı	AIC Ortalaması	BIC Ortalaması	Log-olabilirlik Ortalaması
İki Boyutlu				
1.	Model 1	92437,72	93075,30	-46098,9
	Model 2	90048,65	90199,72	-45045,3
	Model 3	90480,87	91740,10	-45003,4
2.	Model 1	90017,46	90655,05	-44888,8
	Model 2	87013,16	88272,40	-43269,6
	Model 3	88463,33	89414,39	-44052,7
3.	Model 1	89493,03	90130,61	-44626,5
	Model 2	86033,48	87184,54	-43337,7
	Model 3	87054,70	88313,93	-43290,3

(Devam ediyor)

Tablo 33 (Devam)

Tüm Koşular için Tek Boyutlu Kestirimlere Yönelik Uyum İstatistikleri Ortalamaları

		100 Tekrar İçin		
Üretilme Koşulu	Kestirimlerde Tan. Boyut Sayısı	AIC Ortalaması	BIC Ortalaması	Log-olabilirlik Ortalaması
4.	Model 1	88301,65	88939,23	-44533,4
	Model 2	84719,72	83670,78	-42859,4
	Model 3	85746,63	87005,86	-42818,8
5.	Model 1	92112,04	92749,63	-45936,0
	Model 2	90590,80	91541,87	-45616,4
	Model 3	91596,15	92855,38	-45561,0
6.	Model 1	92618,37	93255,96	-46189,2
	Model 2	91204,14	92155,20	-45923,1
	Model 3	92229,64	93488,87	-45877,8
7.	Model 1	94154,46	94792,05	-47047,8
	Model 2	80951,79	86902,85	-45449,1
	Model 3	90968,81	92228,04	-45247,4
8.	Model 1	91982,18	92619,77	-45871,1
	Model 2	87125,99	88077,05	-43884,0
	Model 3	88162,15	89421,39	-43844,1
9.	Model 1	92628,47	93266,05	-46194,2
	Model 2	87304,50	88255,56	-43973,2
	Model 3	88336,84	89596,07	-43931,4
10.	Model 1	92401,05	93038,64	-46080,5
	Model 2	85667,94	86619,01	-43155,0
	Model 3	86683,38	87942,61	-43104,7
11.	Model 1	94121,67	94759,26	-46940,8
	Model 2	91722,51	92673,58	-46182,3
	Model 3	92743,04	94002,27	-46134,5
12.	Model 1	94063,96	94701,55	-46912,0
	Model 2	91558,97	92510,04	-46100,5
	Model 3	92596,29	93855,52	-46061,1
Üç Boyutlu				
13.	Model 1	91033,24	91670,83	-45396,6
	Model 2	90030,82	91105,76	-44836,4
	Model 3	89846,52	80981,88	-44686,3
14.	Model 1	89225,63	89863,21	-44492,8
	Model 2	88262,25	89213,31	-43952,1
	Model 3	86126,41	87385,64	-43826,2
15.	Model 1	89917,57	90555,16	-44838,8
	Model 2	88098,46	89049,53	-43870,2
	Model 3	86663,56	87922,80	-43594,8
16.	Model 1	88341,01	88978,60	-44050,5
	Model 2	86405,74	87356,81	-43023,9
	Model 3	85120,94	86380,17	-42823,5
17.	Model 1	94003,91	94641,50	-47454,1
	Model 2	93783,65	94734,72	-46712,8
	Model 3	92774,08	93033,32	-46650,0
18.	Model 1	91119,39	91756,97	-45439,7
	Model 2	90921,04	91872,11	-45281,5
	Model 3	89054,73	89213,97	-45240,4
19.	Model 1	91948,75	92586,34	-45854,4
	Model 2	90132,22	91083,28	-44887,1
	Model 3	89077,68	89236,91	-45176,6
20.	Model 1	94167,67	94805,26	-46963,8
	Model 2	91832,10	92783,17	-45737,1
	Model 3	90256,61	91515,85	-45391,3

(Devam ediyor)

Tablo 33 (Devam)

Tüm Koşular için Tek Boyutlu Kestirimlere Yönelik Uyum İstatistikleri Ortalamaları

		100 Tekrar İçin		
Üretilme Koşulu	Kestirimlerde Tan. Boyut Sayısı	AIC Ortalaması	BIC Ortalaması	Log-olabilirlik Ortalaması
21.	Model 1	94715,44	95353,02	-47237,7
	Model 2	91409,71	92360,77	-45525,9
	Model 3	89425,71	90684,95	-44975,9
22.	Model 1	94202,01	94839,60	-46981,0
	Model 2	90206,94	91158,01	-44924,5
	Model 3	87964,30	89223,53	-44245,1
23.	Model 1	93231,29	93868,88	-46495,6
	Model 2	88688,18	89639,24	-44165,1
	Model 3	85930,72	87189,96	-43228,4
24.	Model 1	91142,53	91780,12	-45451,3
	Model 2	86211,53	87162,60	-42926,8
	Model 3	83105,98	84365,22	-41816,0

*Modeller arasında yapılan model karşılaştırması sonucunda LRT (likelihoodratio test) istatistiğinin p değerleri manidardır

Tablo 33 incelendiğinde, ilk 12 koşulda, iki boyutlu; son 12 koşulda ise üç boyutlu verilerin üretildiği koşullara ait sonuçların yer aldığı görülmektedir. Bu koşullarda, her bir koşul için, modellere ait AIC, BIC ve log-Lik değerlerinin 100 tekrara ait ortalamaları alınarak ilgili modelin karşısında verilmiştir. Bu ortalama değerler incelendiğinde genel olarak iki boyutlu üretilen verilerde, bu değerlerin Model 2'de daha düşük olduğu; üç boyutlu üretilen verilerde ise bu değerlerin Model 3'te daha düşük olduğu görülmektedir. Koşullar ayrı ayrı ele alındığında; 1., 2., 9., 11., 20. ve 21. koşullarda, 100 tekrardan ikisinin beklenilen boyut dışında bir boyuta daha iyi uyum sağladığı görülmektedir. 5., 8. ve 12. koşullarda ise 100 tekrardan üçünün beklenilen boyut dışında bir boyuta daha iyi uyum sağladığı görülmektedir. Bununla birlikte 10. koşulda 100 tekrardan dördünün; 14. koşulda ise 100 tekrardan birinin beklenilen boyut dışında bir boyuta daha iyi uyum sağladığı görülmektedir. Bu kapsamda tüm sonuçlar birlikte değerlendirildiğinde genel olarak iki boyutlu üretilen verilerde, iki boyutlu modelin; üç boyutlu üretilen verilerde ise üç boyutlu modelin veriye daha iyi uyum sağladığı; 2400 tekrardan 26 tanesinin beklenilen boyut dışında bir boyuta daha iyi uyum sağladığı görülmektedir. Çalışma kapsamında üretilen veri setleri üzerinde çok boyutlu kestirimler yapılmıştır. Elde edilen bu yeni veri setleri üzerinde R dilinde yer alan *"mirt"*(Chalmers, 2012) paketinden yararlanılarak tekrar her bir koşul için ayrı ayrı model kestirimleri yapılmıştır. Bu kestirimlere ait uyum istatistikleri, her bir koşulx100 tekrar için elde edilerek ortalamalarına ait sonuçlar Tablo 34'te verilmiştir.

Tablo 34

Tüm Koşular için Çok Boyutlu Kestirimlere Yönelik Uyum İstatistikleri Ortalamaları

Üretilme Koşulu	Kestrimlerde Tan. Boyut Sayısı	AIC Ortalaması	BIC Ortalaması	Log-olabilirlik Ortalaması
1.	Model 1	89378,02	90008,59	-44528,4
	Model 2	86122,48	87178,48	-43337,7
	Model 3	85054,70	88313,93	-43290,3
2.	Model 1	88215,56	88927,21	-44522,7
	Model 2	84734,65	83776,88	-42864,3
	Model 3	83456,63	82455,83	-42823,3
3.	Model 1	92222,14	92678,56	-45945,1
	Model 2	91567,81	91551,84	-45613,1
	Model 3	90136,15	90185,38	-45551,1
4.	Model 1	92567,43	93155,86	-46192,1
	Model 2	92223,14	93457,32	-45942,2
	Model 3	91116,64	92188,87	-45877,8
5.	Model 1	91238,98	92134,89	-45678,9
	Model 2	90968,81	92228,04	-45247,4
	Model 3	89167,16	91851,41	-45123,2
6.	Model 1	91982,18	94701,66	-45871,1
	Model 2	88224,74	92510,12	-44841,2
	Model 3	87125,99	92134,34	-44454,0
7.	Model 1	92628,47	94641,50	-46194,2
	Model 2	88336,84	94231,78	-43930,2
	Model 3	87324,50	94123,43	-43973,2
8.	Model 1	92401,12	93542,48	-46080,5
	Model 2	86683,38	93334,67	-43164,1
	Model 3	85512,15	93227,28	-43023,6
9.	Model 1	94121,67	93533,32	-46940,8
	Model 2	92743,04	93197,34	-46134,5
	Model 3	91234,12	93167,27	-46134,5
10.	Model 1	94063,96	93234,12	-46912,0
	Model 2	91558,97	93109,54	-46098,6
	Model 3	91122,29	92098,66	-46078,2
11.	Model 1	93003,91	92454,89	-47454,1
	Model 2	93783,65	92132,84	-46712,8
	Model 3	92774,08	92095,74	-46650,0
12.	Model 1	92596,29	92316,57	-46178,2
	Model 2	91237,33	91734,43	-46123,4
	Model 3	90089,78	91503,72	-46036,3
13.	Model 1	92123,34	92230,57	-45458,7
	Model 2	91237,33	91111,59	-45423,4
	Model 3	90089,78	91042,53	-45336,3
14.	Model 1	92345,67	91477,43	-45254,4
	Model 2	91344,22	91255,63	-45250,2
	Model 3	90934,02	90424,29	-45231,3
15.	Model 1	92118,23	91689,83	-44439,7
	Model 2	91678,12	91074,22	-45281,5
	Model 3	89054,73	89218,65	-45240,4
16.	Model 1	91566,45	91075,49	-45854,4
	Model 2	90172,33	91083,34	-44887,1
	Model 3	89199,68	89247,90	-45176,6
17.	Model 1	91937,75	91086,98	-45854,4
	Model 2	90222,22	91001,78	-44887,1
	Model 3	89088,68	89115,89	-44776,6

(Devam ediyor)

Tablo 34 (Devam)

Tüm Koşular için Çok Boyutlu Kestirimlere Yönelik Uyum İstatistikleri Ortalamaları

Üretilme Koşulu	Kestrimlerde Tan. Boyut Sayısı	AIC Ortalaması	BIC Ortalaması	Log-olabilirlik Ortalaması
18.	Model 1	91123,56	90667,54	-44451,2
	Model 2	86220,48	88932,45	-44126,7
	Model 3	86105,75	88345,33	-44016,1
19.	Model 1	89054,73	89675,99	-44040,3
	Model 2	88345,23	88789,53	-43950,2
	Model 3	86245,74	87659,75	-43623,4
20.	Model 1	88578,18	89639,24	-43988,1
	Model 2	87975,36	89223,53	-43978,1
	Model 3	87666,30	88113,47	-43978,0
21.	Model 1	88567,18	88798,24	-42567,1
	Model 2	87795,76	87345,94	-42333,4
	Model 3	85678,72	86179,82	-42110,2
22.	Model 1	86642,57	87987,46	-42451,3
	Model 2	86456,53	87101,63	-42326,7
	Model 3	85467,97	86549,22	-42216,4
23.	Model 1	86546,56	86984,65	-42121,3
	Model 2	86199,45	86886,54	-42120,4
	Model 3	85998,98	85681,22	-42119,3
24.	Model 1	86642,75	85193,55	-41548,3
	Model 2	86123,85	84234,79	-41543,8
	Model3	85469,12	83325,12	-41536,9

*Modeller arasında yapılan model karşılaştırması sonucunda LRT (likelihoodratio test) istatistiğinin p değerleri manidardır.

Tablo 34 incelendiğinde, ilk 12 koşulda, iki boyutlu; son 12 koşulda ise üç boyutlu verilere ait sonuçların yer aldığı görülmektedir. Bu koşullarda, her bir koşul için, modellere ait AIC, BIC ve log-Lik değerlerinin 100 tekrara ait ortalamaları alınarak ilgili modelin karşısında verilmiştir. Bu ortalama değerler incelendiğinde iki boyutlu üretilen verilerde, bu değerlerin Model 2'de daha düşük olduğu; üç boyutlu üretilen verilerde ise bu değerlerin Model 3'te daha düşük olduğu görülmektedir. Bu kapsamda tüm sonuçlar birlikte değerlendirildiğinde iki boyutlu üretilen verilerde, iki boyutlu modelin; üç boyutlu üretilen verilerde ise üç boyutlu modelin veriye daha iyi uyum sağladığı görülmektedir.

Tablo 33 ve Tablo 34 birlikte incelendiğinde, Tablo 33'te genel olarak iki boyutlu üretilen verilerde, iki boyutlu modelin; üç boyutlu üretilen verilerde ise üç boyutlu modelin veriye daha iyi uyum sağladığı; 2400 tekrardan 26 tanesinin beklenilen boyut dışında bir boyuta daha iyi uyum sağladığı görülmekte iken bu durumun çok boyutlu kestirimler üzerinden elde edilen sonuçlarda düzeldiği; tüm tekrarlarda iki boyutlu üretilen verilerde, iki boyutlu modelin; üç boyutlu üretilen verilerde ise üç boyutlu modelin veriye daha iyi uyum sağladığı görülmektedir. Bu durum çok boyutlu kestirimler

üzerinden elde edilen sonuçların uyum istatistiklerini olumlu yönde etkilediğini göstermektedir.

Gerçek Veri Setinde Tek Boyutlu ve Çok Boyutlu Kestirimlerin Lojistik Regresyon, SIBTEST ve Lord'un χ^2 Tekniklerinin Tip I Hata ve İstatistiksel Güç Oranlarının İncelenmesi

Kestirim yöntemi ve tekniklere göre Tip I hata oranlarının nasıl farklılaştığını belirlemek için tek yönlü MANOVA yapılmıştır. Analiz sonucunda kestirim yöntemi ve teknik değişkenleri birlikte ele alındığında, Tip I hata oranları üzerinde manidar farklılıklar oluşturduğu belirlenmiştir, [F(1,4)=10.596, p=0.05367, Pillai=0.87599; np^2=78]. Değişkenler için sonuçlar ayrı ayrı incelendiğinde, kestirim yöntemleri değişkeninin Tip I hata oranları açısından manidar farklılıklar oluşturduğu belirlenmiştir, [F(1,4)=11.804, p=0.0264]. Değişkenler için sonuçlar ayrı ayrı incelendiğinde, teknik değişkeninin ise Tip I hata oranları açısından manidar farklılıklar oluşturmadığı belirlenmiştir, [F(1,4) = 0.1369, p=0.7301].

Kestirim yöntemi ve tekniklere göre istatistiksel güç oranlarının nasıl farklılaştığını belirlemek için tek yönlü MANOVA yapılmıştır. Analiz sonucunda kestirim yöntemi ve teknik değişkenleri birlikte ele alındığında, Tip I hata oranları üzerinde manidar farklılıklar oluşturduğu belirlenmiştir, [F(1,4) = 11.325, p=0.04, Pillai=0.88304; np^2=88]. Değişkenler için sonuçlar ayrı ayrı incelendiğinde, kestirim değişkeninin istatistiksel güç oranları açısından manidar farklılıklar oluşturduğu belirlenmiştir, [F(1,4)=14.626, p=0.01871]. Değişkenler için sonuçlar ayrı ayrı incelendiğinde, teknik değişkeninin ise istatistiksel güç oranları açısından manidar farklılıklar oluşturmadığı belirlenmiştir [F(1,4) = 0.4346, p=0.5462].

BÖLÜM 5

SONUÇ VE ÖNERİLER

Bu bölümde araştırmanın bulgularından elde edilen sonuçlara yer verilmiştir. Bulgulara ait sonuçlar verilirken en önemli bulgulardan başlamak üzere sonuçlar verilmiştir. Gerçek veri uygulamasından elde edilen sonuçlar da ilgili sonuçları destekleyecek şekilde kullanılmıştır. Daha sonra elde edilen tüm sonuçlar kapsamında uygulayıcılara ve araştırmacılara önerilerde bulunulmuştur.

Sonuçlar

Bu çalışmanın sonuçları, çalışmanın bulgularına bağlı olarak ayrı başlıklar halinde sunulmuştur. İlk başlık altında DMF türleri birlikte ele alınarak; DMF türlerine göre tek boyutlu ve çok boyutlu kestirimlere yönelik yapılan DMF analizlerine ait Tip I hata ve istatistiksel güç oranları ve buna bağlı sonuçlar açıklanmaktadır. İkinci başlık altında tek ve çok boyutlu kestirimlere yönelik farklı koşullar altında Tip I hata ve istatistiksel güç oranlarının incelenmekte ve üçüncü başlık altında ise yetenek parametre kestirimlerine ait uyum istatistiklerine değinilmektedir.

Tek Boyutlu ve Çok Boyutlu Kestirimlere Yönelik Yapılan DMF Analizlerine Ait Tip I hata ve İstatistiksel Güç Oranlarının İncelenmesi

Hem TB hem de TBO DMF türü için, ÇB kestirim yöntemi ile TB kestirim yöntemine göre daha düşük Tip 1 hata değerleri ve daha yüksek istatistiksel güç değerleri elde edilmiştir. Tek ve çok boyutlu kestirim yöntemleri arasındaki farkların manidar ve yüksek düzey etkiye sahip olduğu sonucuna ulaşılmıştır. Gerçek veri ile yapılan inceleme sonucunda da tek ve çok boyutlu kestirimlere ait Tip I hata ve istatistiksel güç oranları arasında, çok boyutlu kestirim lehine manidar farklılıklar olduğu görülmüştür. Bu sonuçlar birlikte dikkate alındığında beklendiği üzere çok boyutlu kestirimlerin, tek boyutlu kestirimlere göre düzeltilmiş puan elde edilmesinde etkili olduğunun gerçek verilerle desteklendiği görülmektedir.

Hem TB hem de TBO DMF türü için, Tip I hata oranları üzerinde DMF Yüzdesi ve DMF Büyüklüğü arttıkça tek ve çok boyutlu kestirimlerin Tip 1 hata oranları üzerindeki manidar farklılıkları ve bu farklılıklara yönelik etki büyüklükleri artmaktadır. Diğer bir ifadeyle TB ve TBO DMF türü için DMF Yüzdesi ve DMF büyüklüğü arttıkça çok boyutlu kestirim daha fazla düzeltme sağlamaktadır. Sadece TBO DMF türü için ise DMF yüzdesine ait farklılıkların manidar ancak etkisiz olduğu görülmüştür. İstatistiksel güç oranları için ise hem TB hem de TBO DMF türü için DMF yüzdesi, DMF büyüklüğü ve boyut sayısı arttıkça daha fazla düzeltme sağlamaktadır.

Hem TB hem de TBO DMF türü için; çok boyutlu kestirimde Tip I hata oranları açısından tek boyutlu kestirime göre en fazla değişimin LR tekniğinde; en az değişimin ise Lord'un χ^2 tekniğinde olduğu görülmektedir. Buna karşın çok boyutlu kestirimde istatistiksel güç oranları açısından ise tek boyutlu kestirime göre en fazla değişimin Lord'un χ^2 tekniğinde; en az değişimin ise LR tekniğinde olduğu görülmektedir. Bununla birlikte hem TB hem de TBO DMF türü için tüm koşullarda tek boyutlu kestirimde bu koşulların alt düzeyleri arasında Tip I hata ve istatistiksel güç oranları arasındaki farklılıklar belirginken, çok boyutlu kestirimde bu farklılıklar oldukça azalmakta; düzeyler birbirine benzer çalışmaktadır.

Sonuçlar genel olarak değerlendirildiğinde; hem TB hem de TBO DMF türü için; çok boyutlu kestirimlerden elde edilen veri setleri üzerinde yapılan DMF analizleri sonucunda, tek boyutlu kestirimlerden elde edilen veri setleri üzerinde yapılan DMF analizlerinden daha düşük Tip 1 hata değerleri ve daha yüksek istatistiksel güç oranları elde edildiği görülmektedir. TB ve TBO DMF türü için, DMF yüzdesi, DMF büyüklüğü, boyut sayısı arttıkça sağlanan düzeltmenin arttığı sonucuna ulaşılmıştır.

Tek ve Çok Boyutlu Kestirimlere Yönelik Farklı Koşullar Altında Tip I Hata ve İstatistiksel Güç Oranlarının İncelenmesi

DMF Yüzdesi, DMF büyüklüğü, boyut sayısı arttıkça tüm tekniklerde Tip I hata oranlarının arttığı görülmektedir. LR'nun TBO DMF; SIBTEST'in ise TB DMF'nin belirlenmesinde daha etkili olduğu, DMF yüzdesi arttıkça bu etkinin arttığı belirlenmiştir. Lord'un χ^2'nin ise hem TB hem de TBO DMF'nin belirlenmesinde etkili olduğu sonucuna varılmıştır.

DMF yüzdesi, DMF büyüklüğü, boyut sayısı arttıkça genel olarak istatistiksel güç oranın düştüğü görülmektedir. DMF türü bakımından incelendiğinde ise diğer koşullardan farklı olarak tüm tekniklerin benzer şekilde çalışmadığı görülmektedir. LR ve Lord'un χ^2 için TB; SIBTEST için ise TBO DMF'nin belirlenmesinde elde edilen istatistiksel güç oranlarının daha yüksek olduğu görülmektedir.

Sonuçlar genel olarak değerlendirildiğinde çok boyutlu kestirimlerden elde edilen Tip I hata oranlarının tek boyutlu kestirimlerden düşük; istatistiksel güç oranlarının ise daha yüksek olduğu görülmektedir. Diğer bir ifadeyle çok boyutlu kestirim, tek boyutlu kestirime göre puanlar üzerinde düzeltme sağlamaktadır. Bu durum ÇBMTK'nın birinci yetenek düzeylerine göre sabitleme yaparak düzeltme sağladığını ortaya koymaktadır. Tek boyutlu kestirim sonuçları dikkate alındığında Tip I hata oranındaki şişmeye bağlı olarak, Tip I hata oranı 0.10 değerinin üstünde olan pek çok koşul için testlerin gücünün rapor edilmesi tavsiye edilmemektedir. Bu kapsamda çok boyutlu kestirim üzerinden sonuçlar elde edilmesinin bu duruma çözüm olabileceği; Boyut sayısı, DMF yüzdesi, DMF büyüklüğü arttıkça ve TBO da daha yüksek düzeltme sağladığı ortaya konmuştur.

Tek ve Çok Boyutlu Kestirimlere Yönelik Yetenek Parametre Kestirimlerine Ait Uyum İstatistiklerinin İncelenmesi

Tek boyutlu kestirimde genel olarak iki boyutlu üretilen verilerin iki boyutlu; üç boyutlu üretilen verilerin ise üç boyutlu yapıya daha iyi uyum sağladığı; 2400 tekrardan 26 tanesinin beklenilen boyut dışında bir boyuta daha iyi uyum sağladığı görülmektedir. Çok boyutlu kestirimlerde ise tüm iki boyutlu üretilen verilerin, iki boyutlu yapıya; üç boyutlu üretilen verilerin ise üç boyutlu yapıya daha iyi uyum sağladığı belirlenmiştir. Tek boyutlu kestirimde 2400 tekrardan 26 tanesinin beklenilen boyut dışında bir boyuta daha iyi uyum sağladığı görülmekte iken bu durumun çok boyutlu kestirimler üzerinden elde edilen sonuçlarda düzeldiği görülmüştür. Bu durum çok boyutlu kestirimler üzerinden elde edilen sonuçların uyum istatistiklerini olumlu yönde etkilediğini göstermektedir.

Öneriler

Yapılan bu araştırmadan elde edilen bulgular doğrultusunda yapılabilecek olan uygulamaya dönük öneriler ve araştırmacılara yönelik öneriler aşağıda sunulmuştur.

Uygulayıcılara Öneriler

1. Tek boyutluluğu sağlanamayan testlerde, ÇBMTK ile yetenek kestirimi yapmaları, elde edilen birinci yetenek parametrelerini dikkate alarak DMF analizlerini gerçekleştirmeleri önerilmektedir.
2. Özellikle boyut sayısı, DMF yüzdesi, DMF büyüklüğü fazla olan koşullarda tek boyutlu yetenek kestirimine göre daha fazla düzeltme sağlayan çok boyutlu yetenek kestiriminin tercih edilmesi önerilmektedir.
3. TB DMF'nin belirlenmesinde; daha düşük Tip I hata oranları veren LR ve daha yüksek istatistiksel güç oranları veren Lord'un χ^2 tekniğinun kullanılması önerilmektedir.
4. TBO DMF'nin belirlenmesinde; daha düşük Tip I hata oranları veren Lord'un χ^2 ve daha yüksek istatistiksel güç oranları veren LR tekniğinun kullanılması önerilmektedir.

Araştırmacılara Öneriler

1. Bu çalışmada yetenek parametreleri normal dağılıma göre sabitlenmiştir. Yetenek parametrelerine ait dağılımların da sonuçlar üzerinde etkili olabileceği düşünülmektedir. Bu kapsamda yetenek parametrelerine ait dağılımlar da değişimlenerek; farklı yetenek dağılımlarında çalışma tekrarlanabilir.
2. Bu çalışmada tek ve çok boyutlu kestirimlerin DMF puanlarını düzeltme açısından farkları incelenmiştir. İleri çalışmalarda farklı koşullar altında tek ve çok boyutlu kestirimlerin Değişen Test Fonksiyonu ve Bilgi Fonksiyonu üzerindeki etkileri incelenebilir.
3. Bu çalışma kapsamında TB ve TBO DMF koşulları incelenmiştir. Çapraz TBO DMF de dikkate alınarak çalışma tekrarlanabilir.

KAYNAKLAR

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: dynamics of ability determinants. *Journal Of Applied Psychology, 77*(5), 598-614.

Ackerman, T. A. (1989). Unidimensional IRT calibration of compensatory and non compensatory multidimensional items. *Applied Psychological Measurement, 13*(2), 113-127.

Ackerman, T. A. (2005). Multidimensional item response theory modeling. In A. Maydeu Olivares ve J. J. McArdle (Eds.), *Contemporary Psychometrics*, 3-26. Mahwah, NJ, Lawrence Erlbaum.

Ackerman, T. A., Gierl, M. J. ve Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice, 22*(3), 37-51. doi: 10.1111/j.1745-3992.2003.tb00136.

Akalın, Ş. Y. ve Koç, N. T. D. (2014). *Kamu Personeli Seçme Sınavı Genel Yetenek Testinin Madde Yanlılığı Açısından İncelenmesi* (Yayınlanmamış doktora tezi). Ankara Üniversitesi. Eğitim Bilimleri Enstitüsü, Ankara.

Akcan, R. (2018). *2016 Lisans Yerleştirme Sınavı İngilizce Testinin Madde Yanlılığı Açısından İncelenmesi* (Yayınlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA ve NCME) (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Anderson, P. (1999). Perspective: Complexity theory and organization science. *Organization Science, 10*(3), 216-232.

Angoff, W. H. ve Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement, 10*(2), 95-106. doi: 10.1111/j.1745-3984.1973.tb00787.

Ansley, T. N. ve Forsyth R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two dimensional data. *Applied Psychological Measurement, 9*(1), 37-48.

Arslan, M. (2020). *TEOG Sınavının Yabancı Dil Alt Testine Ait Maddelerin Yanlılığının İncelenmesi* (Yayınlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Atalay, K., Gök, B., Kelecioğlu, H. ve Arsan, N. (2012). Değişen Madde Fonksiyonunun Belirlenmesinde Kullanılan Farklı Yöntemlerin Karşılaştırılması: Bir Simülasyon Çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *43*(43), 270-281.

Atar, B. (2007). *Differential item functioning analyses for mixed response data using IRT likelihood-ratio test, logistic regression, and GLLAMM procedures* (Unpublished doctoral dissertation). Florida State University, USA.

Atar, B. ve Kamata, A. (2011). Comparison of ırt likelihood ratio test and logistic regression DIF detection procedures. H*acettepe University Journal of Education*, *41*, 36-47.

Avsar, A. S. ve Tavsancil, E. (2017). Examination of polytomous items' psychometric properties according to nonparametric item response theory models in different test conditions. *Educational Sciences*, *17*(2), 493-514.

Awuor, R. A. (2008). *Effect of unequal sample sizes on the power of DIF detection: An IRT based Monte Carlo study with SIBTEST and Mantel-Haenszel procedures.* State University, Virginia Polytechnic Institute, Virginia.

Ayva Yörü, F. G. ve Atar, H. Y. (2019). Determination of differential item functioning (DIF) according to SIBTEST, Lord's [Chi-squared], Raju's area measurement and breslow-day methods. *Journal of Pedagogical Research*, *3*(3), 139-150.

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, (MD): ERIC Clearinghouse on Assessment and Evaluation.

Baker, F. B. ve Kim, S. (2004). Item response theory: *Parameter estimation techniques* (2nd ed.). New York: Marcel Dekker

Bauer, D., Holzer, M., Kopp, V. ve Fischer, M. R. (2011). Pick-N multiple choice-exams: a comparison of scoring algorithms. *Advances in Health Sciences Education*, *16*, 211-221.

Bobcock, B.G.E.(2009). *Estimating a noncompensatory IRT model using a modified metropolis algorithm* (Unpublished Doctoral Dissertation). The University of Minesota, Minesota.

Bock, R. D. ve Aitkin M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, *46*(4), 443-459. doi: 10.1007/BF02293801

Bock, R. D. ve Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, *6*(4), 431-444. doi: 10.1177/014662168200600405

Bolt, D. M. (2002). A monte carlo comparison of parametric and nonparametric polytomous DIF detection methods. *Applied Measurement in Education*, *15*(2), 113-141. Retrieved from https://eric.ed.gov/?id=EJ654822

Bolt, D. M. ve 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. ve Stout, W. (1996). Differential item functioning: Its multidimensional model and resulting SIBTEST detection procedure. *Behaviormetrika, 23,* 67-95.

Boughton, K. A., Gierl, M. J. ve Khaliq, S. N. (2000, May). Differential bundle functioning on mathematics and science achievement tests: a small step toward understanding differential performance. In *Annual Meeting of the Canadian Society for Studies in Education (CSSE), Edmonton, Alberta, Canada*, 24-27.

Bradley, J. V. (1978). Robustness? *The British Journal of Mathematical and Statistical Psychology*, *31*(2), (pp.144–152). doi: 10.1111/j.2044-8317.1978.tb00581.

Bulut, O. ve Suh, Y. (2017, October). Detecting multidimensional DIF with the multiple indicators multiple causes model, the item response theory likelihood ratio test, and logistic regression. 5*1*(2), (pp. 1-14). doi: 10,3389/feduc.2017.00051

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

Cao, M., Tay, L. ve Liu, Y. (2017). A monte carlo study of an iterative wald test procedure for DIF analysis. *Educational and Psychological Measurement, 77*(1), 104–118. doi: 10.1177/0013164416637104.

Cappaert, K. (2014). *Dissecting the impact of DIF/DBF on ability estimation and person fit.* (Unpublished doctoral dissertation). The University of Wisconsin-Milwaukee in Educational Psychology.

Carroll, H. C. (2015). *An examination of the improved wald test for differential item functioning detection with multiple groups* (Unpublished doctoral dissertation). University of Kansas, Kansas city.

Carter, N. T. ve Zickar, M. J. (2011). The influence of dimensionality on parameter estimation accuracy in the generalized graded unfolding model. *Educational and Psychological Measurement, 71*(5), 765-788. doi: 10.1177/0013164410387594

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. doi: 10.18637/jss.v048.i06

Clark, P. C. ve LaHuis, D. M. (2012). An examination of power and type I errors for two differential item functioning indices using the graded response model. *Organizational Research Methods*, *15*(2), 229-246. doi: 10.1177/1094428111403815

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale. Lawrence Erlbaum Associates, Publishers.

Cohen, A. S. ve Kim, S. H. (1993). A comparison of Lord's χ^2 and Raju's area measures in detection of. *Applied Psychological Measurement, 17*(1), 39-52. doi: 10.1177/014662169301700109

Cohen, A. S., Kim, S. H. ve Wollack, J. A. (1996). An investigation of the likelihood ratio test for detection of differential item functioning. *Applied Psychological Measurement, 20*(1), 15–26. doi: 10.1177/014662169602000102

Cromwell, S. (2002). Primer on ways to explore item bias. *Paper presented at the 25th Annual Meeting of the Southwest Educational Research Association,* 14-16. February, Austin, Texas, USA.

Çakıcı-Eser, D. (2015). *Çok Boyutlu Madde Tepki Kuramının Farklı Modellerinden Çeşitli Koşullar Altında Kestirilen Parametrelerin İncelenmesi* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Dainis, A. M. (2008). *Methods for identifying differential item and test functioning: An investigation of Type I error rates and power* (Unpublished Doctoral issertation). James Madison University Assessment and Measurement, Harrisonburg, Virgina.

de la Tore, J. ve Patz, R.J. (2005). Making the most of what we have: a practical application of multidimensional item response theory ın test scoring. *Journal of Educational and Behavioral Statistics, 30*(3), 295-311.

DeMars, C. E. (2016). Partially compensatory multidimensional item response theory models: Two alternate model forms. *Educational and Psychological Measurement, 76*(2), 231-257.

DeMars, C. E. (2009). Modification of the mantel-haenszel and logistic regression DIF procedures to incorporate the SIBTEST regression correction. *Journal of Educational and Behavioral Statistics*, *34*, 149-170. doi: 10.3102/1076998607313923

DeMars, C. E. ve Lau, A. (2011). Differential item functioning detection with latent classes: how accurately can we detect who is responding differentially?. *Educational and Psychological Measurement, 71*(4), 597-616.

Doğan, N. (2002). *Klasik Test Kuramı Ve Örtük Özellikler Kuramının Örneklemler Bağlamında Karşılaştırılması* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Dorans, N. J. ve Holland, P. W. (1993). DIF detection and description: mantel- haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning*, *92*(10), (pp. 5-66). Hillsdale, NJ: Erlbaum.

Douglas, J. A., Roussos, L. A. ve Stout, W. (1996). Item-bundle DIF hypothesis testing: Identifying suspect bundles and assessing their differential functioning. *Journal of Educational Measurement*, *33*(4), 465-484.

Duncan, S. C. (2006). *Improving the prediction of differential item functioning: A comparison of the use of an effect size for logistic regression DIF and MantelHaenszel DIF methods* (Unpublished Doctoral Dissertation), Texas A&M University College Station Educational Psychology, Texas.

Ellis, B. B., Raju, N. S. (2003). Test and ıtem bias: what they are, what they aren't, and how to detect them. *ERIC Institute Education Sciences.* Retrieved from: https://files.eric.ed.gov/fulltext/ED480042.pdf.

Embretson, S.E., Reise, S.P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates, Publishers.

Erdem Keklik, D. (2014). Değişen Madde Fonksiyonunu Belirlemede Mantel- Haenszel ve Lojistik Regresyon Tekniklerinin Karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *5*(2), 12- 25.

Erdem, B. (2015). Ortaöğretime Geçişte Kullanılan Ortak Sınavların Değişen Madde Fonksiyonu Açısından Kitapçık Türlerine Göre Farklı Yöntemlerle İncelenmesi (Yayınlanmamış Yüksek Lisans Tezi) Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara

Fidalgo, A. M., Mellenbergh G. J. ve Muniz, J. (2000). Effects of amount of DIF, test length, and purification type on robustness and power of Mantel-Haenszel procedures. *Methods of Psychological Research Online*, *5*(3), 43-53.

Field A. (2005). Discovering Statistics Using SPSS. London: SAGE Publications. *Anuario de psicología/The UB Journal of psychology*, *37*(1), 195-196.

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*(4), 278-295. doi.org/10.1177/01466216052757

Finch, H. (2010). Item parameter estimation for the MIRT model: Bias and precision of confirmatory factor analysis-based models. *Applied Psychological Measurement*, *34*(1), 10-26. doi.org/10.1177/0146621609336112

Finch, W. H. ve French, B. F. (2007). Detection of crossing differential item functioning: A comparison of four methods. *Educational and Psychological Measurement*, *67*(4), 565-582. doi: 10.1177/0013164406296975

Finch, W. H. ve French, B. F. (2008). Anomalous Type I error rates for identifying one type of differential item functioning in the presence of the other. *Educational and Psychological Measurement*, *68*(5), 742–759. doi: 10.1177/0013164407313370

Flora, D. B. ve Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466-478.

Flowers, C.P., Raju, N.S. ve Oshima, T. C. (2002). A comparison of measurement equivalence methods based on confirmatory factor analysis and item response theory. *Paper presented at the Annual Meeting of the National Council on Measurement in Education*, New Orleans.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., ... ve Heiberger, R. (2012). Package 'car'. *Vienna: R Foundation for Statistical Computing*, *5*(17), 16-26.

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

Fritz, C. O., Morris, P. E. ve Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology*: *General*, *141*(1), 2–18. doi: 10.1037/a0024338

Furlow, C. F., Raiford Ross, T. ve Gagné, 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.

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

George, D. ve Mallery, P. (2020). IBM SPSS statistics 26 step by step: *A simple guide and reference.* Routledge: Taylor & Francis Group.

Gierl, M. J., Jodoin, M. G., ve Ackerman, T. A. (2000). Performance of mantel-haenszel, simultaneous item bias test, and logistic regression when the proportion of DIF items is large. *The Annual Meeting of the American Educational Research Association*, New Orleans, Louisiana, USA.

Girl, M. J. ve Khaliq, S. N. (2001). Identifying sources of differential item and bundle functioning on translated achievement tests: *A confirmatory analysis. Journal of Educational Measurement*, *38*(2), 164-187. doi: 10.1111/j.1745-3984.2001.tb01121.

Golia, S. (2010). The assessment of DIF on Rasch measures with an application to job satisfaction. *Electronic Journal of Applied Statistical Analysis: Decision Support Systems and Services Evaluation*, *1*(1), 16-25.

Golia, S. (2014) Assessing the impact of uniform and nonuniform differential item functioning items on Rasch measure: The polytomous case. *Computational Statistics*, *30*(2), 441-461.

González-Romá, V., Hernandez, A. ve Gomez-Benito, J. (2006). Power and Type I error of the mean and covariance structure analysis model for detecting differential item functioning in graded response items. *Multivariate Behavioral Research, 41*(1), 29-53.

Gray, C. D., & Kinnear, P. R. (2012). *IBM SPSS statistics 19 made simple. New* York: Psychology Press, Taylor & Francis Group.

Guilera, G., Gómez-Benito, J., Hidalgo, M. D. ve Sánchez-Meca, J. (2013). Type I error and statistical power of the Mantel-Haenszel procedure for detecting DIF: a meta-analysis. *Psychological Methods, 18*(4), 553-571. doi: 10.1037/a0034306

Hair, J. F., Risher, J. J., Sarstedt, M., ve Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2-24.

Hambleton, R. K. (1989). *Principles and selected applications of item response theory*. Boston: Springer Science & Business Media.

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

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

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

Happach, R. M. ve Tilebein, M. (2015). Simulation as research method: Modeling social interactions in management science. In C. Misselhorn (Eds.), *Collective agency and cooperation in natural and articial systems*, *122*(2), 239–259. New York: Springer. doi:10.1007/978-3-319-15515-9_13

Harwell, M. R. (1997). Analyzing the results of Monte Carlo studies in item response theory. *Educational and Psychological Measurement*, *57*(2), 266-279.

Harwell, M. (2018). A strategy for using bias and RMSE as outcomes in Monte Carlo studies in statistics. *Journal of Modern Applied Statistical Methods, 17*(2), 29–38. doi: 10.22237/jmasm/1551907966

Harwell, M. R., Kohli, N. ve Peralta, Y. (2017). Experimental design and data analysis in computer simulation studies in the behavioral sciences. *Journal of Modern Applied Statistical Methods, 16*(2), 3-28. doi: 10.22237/jmasm/1509494520

Harwell, M., Stone, C. A., Hsu, T. C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement*, *20*(2), 101-125.

Hidalgo, 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*(6), 903–915. doi: 10.1177/0013164403261769

Hoaglin, D. C. ve Andrews, D. F. (1975). The reporting of computation-based results in statistics. *The American Statistician*, *29*, 122-126. doi: 10.2307/2683438

Holland, P.W. (1985). On the study of differential item performance without IRT. *Proceedings of the 27th Annual Conference of the Military Testing Association, 1*, 282-287, San Diego.

Holmes Finch, W. ve French, B. F. (2007). Detection of crossing differential item functioning: A comparison of four methods. *Educational and Psychological Measurement, 67*(4), 565- 582.

Howell, D. C. (2014). *Fundamental statistics for the behavioral sciences.* Belmont, CA: Cengage Learning.

Hulin, C. L., Lissak, R. I. ve Drasgow, F. (1982). Recovery of two-and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement*, *6*(3), 249-260.

İnal, H. (2018). *Test Eşitlemede Grup Değişmezliğinin Farklı Simülasyon Koşulları Altında İncelenmesi* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara

Jacoby, W. G. (1991). *Data theory and dimensional analysis*, Sage Publication, California.

Jodoin, M. G., Gierl, M. J. (2001). Evaluating type 1 error and power rates using an effect size measure with the logistic regression procedure for dıf detection. *Applied Psychological Measurement, 14*(4), 329-349. doi: 10.1207/S15324818AME1404_2

Johnson Frotman, K. A.(2007). *The evaluation of new criteria for polytomous DIF in the DFIT framework* (Unpublished Doctoral Dissertation). llinois Institute of Technology, Chicago.

Kabasakal, A.., Arsan, N., Gök, B. ve Kelecioglu, H. (2014). Comparing performances (Type I error and Power) of IRT likelihood ratio SIBTEST and Mantel-Haenszel methods in the determination of differential item functioning. *Educational Sciences: Theory and Practice*, *14*(6), 2186-2193.

Kan, A. ve Bulut, O. (2014). Crossed random-effect modeling: examining the effects of teacher experience and rubric use in performance assessments. *Eurasian Journal of Educational Research*, *57*, 1-27.

Kart Arslan, A. (2019). *Karma Yapıdaki Geniş Ölçekli Naep Fen Bilimleri Testinin Boyutluluk Yapısının Çok Boyutlu Madde Tepki Kuramına Göre İncelenmesi* (Yayınlanmış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Kelecioğlu, H. (2001). Örtük Özellikler Teorisindeki b ve a Parametreleri ile Klasik Test Teorisindeki p ve r İstatistikleri Arasındaki İlişki. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *20*(20), 104-110.

Kilmen, S. (2016). Effect of DIF magnitudes, focal group sample size, and DIF ratio on the performance of SIBTEST. *International Journal of Social Sciences and Education, 6*(1), 91-98.

Kim, J. (2010). *Controlling type I error rate in evaluating differential item functioning for four DIF methods: Use of three procedures for adjustment of multiple item testing* (Unpublished Doctoral Dissertation). Georgia State University, Department of Educational Policy Studies, Atlanta.

Koğar, H. (2014). *Madde Tepki Kuramının Farklı Uygulamalarından Elde Edilen Parametrelerin ve Model Uyumlarının Örneklem Büyüklüğü ve Test Uzunluğu Açısından Karşılaştırılması* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Köse, İ. A. (2010). *Madde Tepki Kuramına Dayalı Tek Boyutlu Ve Çok Boyutlu Modellerin Test Uzunluğu ve Örneklem Büyüklüğü Açısından Karşılaştırılması* (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Kreiter, C.D. (1993). *An emprical ınvestigation of compensatory and noncompensatory test ıtems in simulated and real data (*Unpublished Doctoral Dissertation). The University of Iowa, Iowa.

Krisci, L., Hsu, T. ve Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement, 25*, 146-162.

Kristjansson, E., Aylesworth, R., McDowell, I. ve Zumbo, B. D. (2005). A comparison of four methods for detecting DIF in ordered response items. *Educational and Psychological Measurement, 65*(6), 935-953. doi: 10.1177/0013164405275668

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*(863), (pp.1-12). doi: 10.3389/fpsyg.2013.00863

Lee, H. ve Geisinger, K. F. (2016). The matching criterion purification for differential item functioning analyses in a large-scale assessment. *Educational and Psychological Measurement, 76*(1), 141-163. doi: 10.1177/0013164415585166

Lee, Y.-H. ve Zhang, J. (2017). Effects of differential item functioning on examinees' test performance and reliability of test. *International Journal of Testing, 17*(1), 23-54.

Lenth, R., Singmann, H., Love, J., Buerkner, P., ve Herve, M. (2019). Package 'emmeans'. *R package version, 8*(5), 10-17.

Levy, R. (2006). *Posterior predictive model checking for multidimensionality in item response theory and bayesian network (*Unpublished Doctoral Dissertation). Maryland Üniversitesi, Maryland.

Li, Y., Brooks, G. P. ve Johanson, G. A. (2012). Item discrimination and type I error in the detection of differential item functioning. *Educational and Psychological Measurement, 72*(5), 847-861.

Liaw, Y. L. (2015). When can Multidimensional Item Response Theory (MIRT) models be a solution for differential ıtem functioning (DIF). *A Monte Carlo Simulation Study. Washington: University of Washington*.

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

Lord, F.M. (1953). A theory of test scores. *Psychometric Monographs* (Whole No. 7), Richmond, VA: William Byrd Press.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Mahwah, New Jersey: Routledge

Lord, F. M. ve Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Welsley Publishing Company.

Luecht, R. M. ve Miller, T. R. (1992). Unidimensional calibrations and interpretations of composite traits for multidimensional tests. *Applied Psychogical Measurement, 16*, 279-292.

MacDonald, P. ve Paunonen, S.V. (2002). A Monte Carlo comparison of item an person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement, 62*, 921-943.

Magis, D. ve De Boeck, P. (2011). Identification of differential item functioning in multiple-group settings: A multivariate outlier detection approach. *Multivariate Behavioral Research, 46*, 733–755. doi:10.1080/00273171.2011.606757.

Magis, D. ve De Boeck, P. (2012). A robust outlier approach to prevent type I error inflation in differential item functioning. *Educational and Psychological Measurement, 72*(2), 291-311.

Magis, D., Béland, S., Tuerlinckx, F. ve De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*(3), 847-862. doi: 10.3758/BRM.42.3.847

Mair, P. (2018). *Modern psychometrics with R*. Cham: Springer International Publishing.

Mazor, K. M., Hambleton, R. K. ve Clauser, B. E. (1998). Multidimensional DIF analyses: the effects of matching on unidimensional subtest scores. *Appl. Psychol. Measure. 22*, 357–367. doi:10.1177/014662169802200404

Meade, A. W ve Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods, 7*(4), 361-388.

Meara, K., Robin, F. ve Sireci, S. G. (2000). Using multidimensional scaling to assess the dimensionality of dichotomous item data. *Multivariate Behavioral Research, 35*(2), 229-259.

Millsap, R. E. ve Everson, H. T. (1993). Methodology review: Statistical approaches for assesing measurement bias. *Applied Psychological Measurement, 17*(4), 297-334. doi: 10.1177/014662169301700401

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *ETS Research Report Series*, 1-30.

Narayanan, P., Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous ıtem bias procedures for detecting differential. *Applied Psychological Measurement, 18*(4), 315-328.

Norouzian, R. ve Plonsky, L. (2018). Correlation and simple linear regression in applied linguistics. *The Palgrave handbook of applied linguistics research methodology*, 395-421.

Oshima, T. C., Raju, N. S. ve Flowers, C. P. (1997). Development and demonstration of multidimensional IRT-based ınternal measures of differential functioning of ltems and tests. *Journal of Educational Measurement, 34*(3), 253-272.

Osterlind, S. J. ve Everson, H. T. (2009). *Differential item functioning* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Özdemir, B. (2015). A comparison of IRT-based methods for examining differential item functioning in TIMSS 2011 Mathematics subtest. *Procedia Social and Behavioral Sciences, 174*, 2075-2083. doi: 10.1016/j.sbspro.2015.02.004

Özer-Özkan, Y. (2012). *Öğrenci Başarılarının Belirlenmesi Sınavından (ÖBBS) Klasik Test Kuramı, Tek Boyutlu ve Çok Boyutlu Madde Tepki Kuramı Modelleri İle Kestirilen Başarı Puanlarının Karşılaştırılması* (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Özyıldırım Gümüş, F. (2018). *Teog matematik alt testinde değişen madde fonksiyonunun coğrafi bölgelere göre incelenmesi* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara. Erişim adresi https://tez.yok.gov.tr

Pang, X. L vd. (1994, April). Performance of Mantel-Haenszel and Logistic Regression DIF Procedures over Replications Using Real Data. *Paper presented at the Annual Meeting of the American Educational Research Association*, New Orleans, LA. Retrieved from: http://eric.ed.gov/?id=ED374153

Penfield, R. D. (2009). DIFAS: Differential item functioning analysis system (Version 5.0). *Computer Software*.

Penfield, R. D. ve Camilli, G. (2006). 5 Differential item functioning and item bias. *Handbook of Statistics, 26*, (pp.125-167).

Penfield, R. D. ve Camilli, G. (2007). Rao and S. Sinharay eds. *Differential item functioning and item bias* in *Handbook of Statistics, 26*, (pp.125-167).

Potenza, M. T. ve Dorans, N. J. (1995). DIF Assessment For Polytomously Scored Items: A Framework For Classification and Evaluation. *Applied Psychological Measurement*, *19*, 23–3.

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, *53*(4), 495-502. doi: 10.1007/BF02294403

Raju, N. S., van der Linden, W. J. ve Fleer, P. F. (1995). An IRT-based internal measure of test bias with application of differential item functioning. *Applied Psychological Measurement, 19*(4), 353–368. doi: 10.1177/014662169501900405

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, *9*(4), 401-412.

Reckase, M. D. (2009). Multidimensional item response theory models. In *Multidimensional item response theory*. Springer, New York, NY.

Reckase, M. D., Ackerman, T. A., & Carlson, J. E. (1988). Building a unidimensional test using multidimensional items. *Journal of Educational Measurement*, *25*(3), 193-203.

Reckase, M. D., Ackerman, T. A., & Carlson, J.E. (1988). Building a unidimensional test using multidimensional items. *Journal of Educational Measurement*, 25, 193–204.

R Development Core Team (2019). R: *A language and environment for statistical computing.* Vienna, Austria. URL. Retrieved from: http://www.R-project.org/.

Revelle, W. (2022). psych: Procedures for psychological, psychometric, and personality research. doi: https://cran.r-project.org/web/packages/psych/index.html

Reyhanlıoğlu Keçeoğlu, Ç. (2018). *Parametrik ve Parametrik Olmayan Madde Tepki Kuramında Farklı Örneklem Büyüklüklerine ve Boyutluluklarına Göre Parametre Değişmezliğinin İncelenmesi* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Robitzsch, A. (2022). sirt: Supplementary item response theory models. Erişim adresi: https://cran.r-project.org/web/packages/sirt/index.html

Rockoff, D. (2018). *A randomization test for the detection of differential item functioning* (Unpublished doctoral dissertation). The University of Arizona, Arizona.

Rogers, W. T. ve Ndalichako, J. (2000). Number-right, item-response, and finite-state scoring: Robustness with respect to lack of equally classifiable options and item option independence. *Educational And Psychological Measurement, 60*(1), 5-19.

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*(2), 215-230.

Satıcı, D. K. ve Özer Özkan, Y. (2017). Temel Eğitimden Ortaöğretime Geçiş Sınavının (2014-Kasım) cinsiyet açısından madde yanlılığının incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, *13*(1), 254-274.

Schilling, S. ve Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika, 70,* 533-555.

Seungho Yang, M. A.(2007). *A comparison of unidimensional and multidimensional rasch models using parametrer estimates and fit indices when assumption of unidimensionality is violated (*Unpublished Doctoral Dissertation). The Ohio State University, The Ohio State.

Shealy, R. ve Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DTF. *Psychometrika, 58,* 159–194. doi:10.1007/BF02294572

Sheng, Y. (2005). *Bayesian analysis of hierarchical IRT models: compa- ring and combining the unidimensional and multi-unidimension IRT models* Unpublished Doctoral Dissertation). University of Missouri Columbia, Columbia.

Shepard, L., Camilli, G. ve Williams, D. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement, 22*(2), 77–105. doi:10.1111/j.1745-3984.1985.tb01050.

Sireci, S. G. ve Allalouf, A. (2003). Appraising item equivalence across multiple languages and cultures. *Language Testing*, *20*(2), 148-166.

Sireci, S. G., Patsula, L. ve Hambleton, R. K. (2005). Ronald K. Hambleton, Peter F. Merenda, Charles D. Spielberger eds. Statistical methods for identifying flaws in the test adaptation process. In *Adapting educational and psychological tests for cross-cultural assessment*, (pp.93-115).

Spencer, G.S. (2004). *The Strength of Multidimensional item response theory in exporing consrtuct space that is multidimensional and corralated (*Unpublished Doctoral Dissertation). Brigam Young University, Brigam.

Spray, J.A., Davey, T.C., Reckase, M.D., Ackerman, T.A. ve Carlson, J.E. (1990). Comparison of Two Logistic Multidimensional Item Response Theory Models. *Research Report ONR90-8.* Retrieved from: http://ericea.net/ft/tamu/Efa.htm

Stark, S., Chernyshenko, O. S. ve Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, *91*(6), 1292-1312.

Steinberg, L. ve Thissen, D. (2006). Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods, 11*(4), 402-422.

Steinberg, L., Thissen, D. ve Wainer, H. (2000). Dorans N., Eignor D., Flaugher R., Green B.F., Mislevy R.J. eds. Validity. In *Computerized Adaptive Testing*, (pp.85-229), Routledge.

Stephens-Bonty, T. A. (2008). *Using three different categorical data analysis techniques to detect differential item functioning* (Unpublished Doctoral Dissertation). Georgia State University College of Education, Atlanta.

Stout, W., Li, H., Nandakumar, R. ve Bolt, D. (1997). MULTISIB – a procedure to investigate DIF when a test is intentionally multidimensional. *Appl. Psychol. Measure. 21*, 195–213. doi:10.177/01466216970213001

Soysal, S. (2017). *Toplam Test Puanı ve Alt Test Puanlarının Kestiriminin Hiyerarşik Madde Tepki Kuramı Modelleri ile Karşılaştırılması* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Su, Y. H. ve Wang, W. C. (2005). Efficiency of the mantel, generalized mantel-haenszel, and logistic discriminant function analysis methods in detecting differential item functioning for polytomous items. *Applied Measurement in Education, 18*(4), 313-350.

Suh, Y. ve 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 Özelliklerine Sahip Yapılarda, Madde Parametrelerinin Değişmezliğinin Klasik Test Teorisi, Tek Boyutlu Madde Tepki Kuramı Ve Çok Boyutlu Madde Tepki Kuramı Çerçevesinde İncelenmesi* (Yayımlanmamış Doktora Tezi). Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü, Mersin.

Sünbül, Ö. ve Ömür Sünbül, S. (2016). Değişen Madde Fonksiyonunun Belirlenmesinde Kullanılan Yöntemlerde Tip I Hata ve Güç Çalışması. *İlköğretim Online, 15*(3), 882- 897.

Swaminathan, H. ve Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*(4), 361-370. doi: 10.1111/j.1745-3984.1990.tb00754.

Sympson, J.B. (1978). *A model for testing with multidimensional items*. In D. J. Weiss (Ed.), Proceedings of the 1977 Computerized Adaptive Testing Conference, (pp.82-98). Minneapolis: University of Minnesota, Department of Psychology, Computerized Adaptive Testing Laboratory.

Şahin, A. (2012). *Madde Tepki Kuramı'nda Test Uzunluğu Ve Örneklem Büyüklüğünün Model Veri Uyumu, Madde Parametreleri Ve Standart Hata Değerlerine Etkisinin İncelenmesi* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Şahin, A. ve Anıl, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice, 17,* 321–335. doi: 10.12738/estp.2017.1.0270

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using multivariate statistics 6,* (pp.497-516). Boston, MA: Pearson

Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement, 27*, 159-203.

Thissen, D. (2001). *IRTLRDIF v.2.0.b:* Software for the computation of the statistics ınvolved in ıtem response theory likelihood-ratio tests for differential ıtem functioning.

Thissen, D., Steinberg, L., ve Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning, (4),* (pp.67-113). Hillside, NJ: Erlbaum.

Tunç, E. B. (2016). *İki ve Çok Kategorili Puanlanan Maddelerde Değişen Madde Fonksiyonlarının Karşılaştırılması* (Yayınlanmamış Doktora tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara. Erişim adresi: https://tez.yok.gov.tr

Tunç, E. B. ve Kutlu, Ö. (2018). İki ve Çok Kategorili Puanlanan Maddelerde Değişen Madde Fonksiyonlarının Karşılaştırılması. *Başkent University Journal of Education*, *5*(1), 40-50.

Vandenberghe, C. N. ve Gierl, M. J. (2001). *Differential bundle functioning on three achievement tests:* a comparison of aboriginal and non-aboriginal examinees, *71*(2), (pp.143-150).

Vaughn, B. K. ve Wang, Q. (2010). DIF trees: Using classification trees to detect differential item functioni ng. *Educational and Psychological Measurement, 70*, 941-952.

Walker, C. M. ve Göçer Şahin, S. (2020). Using differential item functioning to test for interrater reliability in constructed response items. *Educational and Psychological Measurement, 80*(4), 808-820.

Walker, C. M., Azen, R. ve Schmitt, T. (2006). Statistical versus substantive dimensionality the effect of distributional differences on dimensionality assessment using DIMTEST. *Educational and Psychological Measurement, 66*(5), 721-738.

Walker, C. M., Zhang, B. ve 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.

Walker, C. M., Zhang, B., Banks, K. ve Cappaert, K. (2012). Establishing effect size guidelines for interpreting the results for differential bundle functioning analyses using SIBTEST. *Educational and Psychological Measurement*, *72*(3), 415-434.

Wells, C. S., Subkoviak, M. J. ve Serlin, R. C. (2002). The effect of item parameter drift on examinee ability estimates. *Applied Psychological Measurement, 26*(1), 77-87.

Wang, W. C., Chen, P. H. ve Cheng, Y. Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological methods*, *9*(1), 116-136. doi.org/10.1037/1082-989X.9.1.116

Wang, W. C. ve Su, Y. H. (2004). Effects of average signed area between two item characteristic curves and test purification procedures on the DIF detection via the Mantel–Haenszel method. *Applied Measurement in Education*, *17*(2), 113–144. doi: 10.1207/s15324818ame1702_2

Wang, W., Tay, L., ve Drasgow, F. (2013). Detecting differential item functioning of polytomous items for an ideal point response process. *Applied Psychological Measurement*, *37*(4), 316-335.

Wang, W. C., ve Yeh, Y. L. (2003). Effects of anchor item methods on differential item functioning detection with the likelihood ratio test. *Applied Psychological Measurement, 27*(6), 479-498. doi: 10.1177/0146621603259902

Winkler, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical association*, *62*(319), 776-800.

Woods, C. M. ve Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement*, *35*(5), 339-361.

Wyse, A. E. ve Mapuranga, R. (2009). Differential item functioning analysis using Rasch item information functions. *International Journal of Testing*, *9*(4), 333-357.

Yıldırım, H. (2015). *2012 Yılı Seviye Belirleme Sınavı Matematik Alt Testinin Madde Yanlılığı Açısından İncelenmesi* (Yayımlanmamış Yüksek Lisans Tezi). Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Yıldıztekin, B. (2020). *Test Boyutluluğunun Değişen Madde Fonksiyonuna Etkisinin Farklı Koşullar Altında İncelenmesi* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Zeng, L. (1989). *Robustness of unidimensional latent trait models when applied to multidimensional data* (Unpublished Doctoral Dissertation). Georgia University, Athens.

Zhang, B. (2008). Application of unidimensional item response models to tests with items sensitive to secondary dimensions. *The Journal of Experimental Education, 77*(2), 147-166.

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

Zwick, R., Thayer, D. T. ve Mazzeo, J. (1997). Descriptive and inferential procedures for assessing differential item functioning in polytomous items. *Applied Measurement in Education, 10*(4), 321-344. doi: 10.1207/s15324818ame1004_2

EKLER

EK 1. Simülasyon Koşullarına Yönelik Veri Üretimi

```
#Simülasyon Koşullarına Yönelik Veri Üretimi

library(mirt)
set.seed (1234)

ref.theta < - rmvnorm ( 750, mean = c ( 0, 0 ) ,
                  sigma = matrix ( c ( 1, .5, .5, 1 ), 2 ) )

foc.theta < - rmvnorm ( 750, mean = c ( 0, -1 ) ,
                  sigma = matrix ( c ( 1, .5, .5, 1 ), 2 ) )

Q < - matrix ( 0, nrow = 60, ncol = 2 )

  Q [ - y, 1 ] < - 1
  Q [ y, 2 ] < - 1

prob < - dat < - matrix ( 0, nrow = 1500, ncol = I )

 for ( ii in 1:I ) {
   prob [ ,ii ] < - ( stats::plogis ( a [ ii, 1 ] * ( theta1- b [ ii, 1 ] ) ) ) ^ Q [ ii, 1 ]
   prob [ ,ii ] < - ( stats::plogis ( a [ ii, 2 ] * ( theta2- b [ ii, 2 ] ) ) ) ^ Q [ ii, 2 ] }
                  dat [ prob > matrix ( stats:: runif ( 1500 * I  ), 1500, I ) ] < - 1

ref.theta < - rmvnorm ( 750, mean = c ( 0, 0, 0 ) ,
                  sigma = matrix ( c ( 1, .5, .5, .5, 1, .5, .5, .5, 1 ), 3 ) )

foc.theta < - rmvnorm ( 750, mean = c ( 0, -1, -1 ) ,
                  sigma = matrix ( c ( 1, .5, .5, .5, 1, .5, .5, .5, 1 ), 3 ) )

Q < - matrix ( 0, nrow = 60, ncol = 3 )

  Q [ - c ( y, z ), 1 ] < - 1
  Q [ y, 2 ] < - 1
  Q [ z, 3 ] < - 1

 prob < - dat < - matrix ( 0, nrow = 1500, ncol = I )

   for ( ii in 1:I ) {

     prob [ ,ii ] < - ( stats::plogis ( a [ ii, 1 ] * ( theta1- b [ ii, 1 ] ) ) ) ^ Q [ ii, 1 ]
     prob [ ,ii ] < - ( stats::plogis ( a [ ii, 2 ] * ( theta2- b [ ii, 2 ] ) ) ) ^ Q [ ii, 2 ]
     prob [ ,ii ] < - ( stats::plogis ( a [ ii, 3 ] * ( theta3- b [ ii, 3 ] ) ) ) ^ Q [ ii, 3 ] }

                  dat [ prob > matrix ( stats:: runif ( 1500 * I ), 1500, I ) ] < - 1
```

EK 2. Çok Boyutlu Kestirimden Elde Edilen Birinci Yetenek Parametresi Değerleri

Çok Boyutlu Kestirimden Elde Edilen Birinci Yetenek Parametresi Değerleri		
	1. Koşul	4. Koşul
1	-1,88259	-2,06500
2	-1,95345	-1,94394
3	-1,58144	-1,92435
4	-1,71869	-1,96906
5	-1,78055	-1,66482
6	-1,77296	-1,97166
7	-1,83277	-1,68911
8	-1,65586	-1,79540
9	-1,76754	-1,77723
…	…	…
751	0,33169	0,138266
752	0,75134	0,014821
754	0,93721	0,376628
755	0,76157	0,236421
756	1,16713	0,431832
757	0,45927	-0,400102
758	0,24283	0,089910
769	0,41109	0,219016
760	0,21958	0,043170
…	…	…
1491	2,99951	2,037348
1492	2,07628	2,256026
1493	2,58629	1,873812
1495	1,40564	1,420151
1496	1,73066	2,071136
1497	2,06512	2,094341
1498	-0,33939	1,661593
1499	- 0,44762	1,860013
1500	-0,37302	2,042422

EK 3. LR, SIBTEST ve Lord'un $\chi 2$ Yöntemleri DMF İncelemesi Sonuçları

Maddeler	LR_Stat.	LR_p	SIBTEST_Stat.	SIBTEST_p	Lord'un $\chi 2$ _Stat	Lord'un $\chi 2$ _p
M1	114.7744.	0.0000 ***	46.0249	0.0000 ***	103.2776	0.0000 ***
M2	0.3476	0.5555	0.0194	0.8891	0.0516	0.8202
M3	0.0553	0.8141	0.0329	0.8561	0.0046	0.9461
M4	0.0502	0.8227	0.2149	0.6429	0.0017	0.9667
M5	5.2316	0.0222 *	0.2581	0.6114	2.7815	0.0954
M6	0.0376	0.8463	0.6001	0.4385	0.0804	0.7768
M7	0.3965	0.5289	1.6189	0.2032	0.6615	0.4160
M8	5.3748	0.0204 *	12.1707	0.0005 ***	5.2310	0.0222 *
M9	0.3078	0.5790	0.0391	0.8432	0.1286	0.7198
M10	0.0168	0.8970	0.0453	0.8314	0.0059	0.9389
M11	0.1937	0.6599	0.2493	0.6176	0.0444	0.8331
M12	4.5743	0.0325 *	1.0877	0.2970	2.4475	0.1177
M13	0.4990	0.4800	1.6681	0.1965	0.6453	0.4218
M14	0.3750	0.5403	0.2793	0.5972	0.5259	0.4683
M15	0.5028	0.4783	1.1744	0.2785	0.6685	0.4136
M16	72.1633	0.0000 ***	18.9127	0.0000 ***	67.3763	0.0000 ***
M17	0.4533	0.5008	0.3098	0.5778	0.1892	0.6636
M18	1.0202	0.3125	0.4163	0.5188	1.0.-986	0.2946
M19	0.1481	0.7004	0.6540	0.4187	0.4225	0.5157
M20	0.9656	0.3258	0.2157	0.6423	1.1101	0.2921
M21	72.1485	0.0000 ***	13.0746	0.0003 ***	58.9604	0.0000 ***
M22	0.2657	0.6063	0.2125	0.6448	0.0102	0.9194
M23	0.5711	0.4498	0.0380	0.8455	0.1646	0.6850
M24	0.2829	0.5948	0.0456	0.8309	0.5350	0.4645
M25	0.0115	0.9145	0.0241	0.8767	0.0339	0.8540
M26	0.2525	0.6153	0.0112	0.9157	0.3296	0.5659
M27	5.7180	0.0168 *	2.0647	0.1507	3.0021	0.0832 .
M28	0.0306	0.8612	0.1364	0.7118	0.0092	0.9237
M29	0.0045	0.9465	0.1416	0.7067	0.0749	0.7843
M30	0.0359	0.8498	0.0863	0.7690	0.0001	0.9934
M31	0.1158	0.7336	0.0051	0.9429	0.2301	0.6314
M32	0.3699	0.5430	2.3265	0.1272	0.5450	0.4604
M33	0.7057	0.4009	0.0950	0.7580	0.3957	0.5293
M34	1.0335	0.3093	0.0888	0.7657	0.3091	0.5782
M35	0.7620	0.3827	0.1521	0.6965	0.3994	0.5274
M36	0.2561	0.6128	0.0914	0.7624	0.0226	0.8806
M37	50.1338	0.0000 ***	11.2585	0.0008 ***	47.2600	0.0000 ***
M38	0.6295	0.4275	0.0894	0.7650	0.7314	0.3924
M39	0.1937	0.6598	0.4454	0.5045	0.3538	0.5520
M40	54.1794	0.0000 ***	15.0293	0.0001 ***	52.1473	0.0000 ***
M41	0.0031	0.9555	0.4453	0.5046	0.0630	0.8018
M42	0.0403	0.8409	0.2522	0.6155	0.1424	0.7059
M43	0.2277	0.6332	1.0110	0.3147	0.0609	0.8051
M44	0.2185	0.6402	0.9454	0.3309	0.3870	0.5339
M45	1.0671	0.3016	0.2298	0.6317	0.6424	0.4229
M46	0.2965	0.5861	0.2070	0.6491	0.5346	0.4647
M47	0.1742	0.6764	0.0118	0.9135	0.0877	0.7671
M48	1.1866	0.2760	0.3198	0.5717	0.5748	0.4484
M49	0.3249	0.5687	0.8433	0.3585	0.1107	0.7393
M50	0.0841	0.7718	0.7027	0.4019	0.0003	0.9861
M51	1.9553	0.1620	0.6867	0.4073	0.7647	0.3819
M52	0.5101	0.4751	0.0679	0.7944	0.2211	0.6382
M53	2.9009	0.0885 .	2.9604	0.0853	2.8886	0.0892 .
M54	0.1292	0.7192	0.0076	0.9307	0.0520	0.8196
M55	1.6865	0.1941	2.6864	0.1012	0.8431	0.3585
M56	51.6546	0.0000 ***	28.2418	0.0000 ***	52.6928	0.0000 ***
M57	0.8217	0.3647	0.2193	0.6395	0.3534	0.5522
M58	0.2402	0.6241	0.4221	0.5159	0.0987	0.7534
M59	0.0398	0.8418	0.4212	0.5163	0.0017	0.9673
M60	0.0053	0.9420	0.0839	0.7721	0.0983	0.7538

p: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

BENZERLİK BİLDİRİMİ

"Değişen Madde Fonksiyonunu Kontrol Altına Almada Çok Boyutlu Madde Tepki Kuramı Modellerinin Kullanımı" başlıklı doktora tezimin ana bölümü (ön bölüm, kaynaklar ve ekler hariç) Turnitin İntihali Engelleme Programı aracılığıyla incelenmiş ilgili rapor tez danışmanım tarafından da kontrol edilmiştir. Kontrol sırasında (1) "Kapak, Önsöz, Özet, İçindekiler, Kısaltmalar/Simgeler Sayfası, Tablolar Dizini, Şekiller Dizini, Dipnotlar, Kaynaklar, Ekler, Özgeçmiş" (2) "Tırnak içinde ve blok olarak verilen (doğrudan) alıntılar" (3) Yedi sözcüğe kadar olan benzerlik/örtüşme içeren metin kısımları" (4) "Öğrencinin lisansüstü teziyle ilgili yapmış olduğu yayınlar ve yararlanılan mevzuat metinleri" dışarıda tutulmuştur. Benzerlik kontrolüne ilişkin rapordan elde edilen sonuçlar aşağıdaki tabloda sunulmuştur.

Rapor Tarihi	: 11.11.2023
Gönderim Numarası	: 2224715674
Sayfa Sayısı	: 130
Sözcük Sayısı	: 34346
Karakter Sayısı	: 219437
Benzerlik Oranı	: % 8
Savunma Tarihi	: 06/10/2023

Yukarıda belirtilen sonuçları gösteren Turnitin İntihali Engelleme Programı'na ilişkin orijinal raporu, sonuçlarda herhangi bir değişiklik yapmaksızın bu beyanım ekinde Enstitüye teslim ettiğimi, tezimin %10'dan fazla benzerlik oranı içerdiğinin ve tek bir kaynakla eşleşme oranının %2'yi geçtiğinin belirlenmesi durumunda, bundan doğabilecek tüm yasal sorumluluğu kabul ettiğimi bildirir, saygılarımı sunarım.

Öğrencinin Adı ve Soyadı: Hatice GÜRDİL

Tarih: 16/10/2023

İmza:

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı ve Soyadı : **Hatice GÜRDİL**

E-posta Adresi :

İş Deneyimi

Kurum	**Görev**	**Tarih**
MEB- Orgeneral Bedrettin Demirel İÖO	Öğretmen	2008-2009
MEB- Yukarı Gündeş Şehit Şenay İlan İlköğretim Okulu	Öğretmen	2009- 2015
MEB- Namık Kemal İlkokulu	Öğretmen	2015-2015
MEB- Seyit Burhan Toprak İlkokulu	Öğretmen	2015-2017
MEB- Atam İlkokulu	Öğretmen	2017- Devam ediyor.
MEB- Mesleki ve Teknik Eğitim Genel Müdürlüğü (Görevlendirme- Uzaktan)	Ölçme ve Değerlendirme Uzmanı	Kasım 2020 – Devam ediyor.

Akademik Bilgiler

Eğitim Durumu

Doktora	Ankara Üniversitesi Ölçme ve Değerlendirme Doktora	2018-2023
Yüksek Lisans	Ankara Üniversitesi Ölçme ve Değerlendirme Tezli Yüksek Lisans	2014-2018
Lisans- 2	Anadolu Üniversitesi Sosyoloji	2012-2016
Lisans- 1	Hacettepe Üniversitesi Sınıf Öğretmenliği	2004-2008

Yayınlar

Gürdil Ege, H. (2018). Karma Testlerde İç Tutarlılık Kestirimlerinin Farklı Benzetim Koşullarında İncelenmesi Ankara: Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı. (Danışman: Doç. Dr. Ergül DEMİR)

Gürdil Ege, H. ve Demir, E. (2019).Examining of Internal Consistency Coefficients in Mixed-Format Tests in Different Simulation Conditions. EJER.