



# HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

## ÇOK BOYUTLU MADDE TEPKİ KURAMI TEST EŞİTLEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Burcu DEMİRÖZ

Doktora Tezi

Ankara, 2024

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

*Daha ileriye... En İyiyeye...*



# HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

## ÇOK BOYUTLU MADDE TEPKİ KURAMI TEST EŞİTLEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

## COMPRASION OF MULTIDIMENSIONAL ITEM RESPONSE THEORY TEST EQUATING METHODS

Burcu DEMİRÖZ

Doktora Tezi

Ankara, 2024

## Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Burcu DEMİRÖZ'¼n hazırladıđı “Çok Boyutlu Madde Tepki Kuramı Test Eşitleme Yöntemlerinin Karşılaştırılması” başlıklı bu çalıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalında Doktora** Tezi olarak kabul edilmiřtir.

J¼ri Başkanı Prof. Dr. Selahattin GELBAL İmza

J¼ri Üyesi (Danıřman) Prof. Dr. Nuri DOĐAN İmza

J¼ri Üyesi Doç. Dr. Kübra Atalay Kabasakal İmza

J¼ri Üyesi Doç. Dr. Celal Deha Dođan İmza

J¼ri Üyesi Doç. Dr. S¼meyra Soysal İmza

Enstit¼ Y¼netim Kurulunun  
..../.../.... Tarihli ve .....  
sayılı kararı.

Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, Öğretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 25 / 06 / 2024 tarihinde uygun gör¼lm¼ř ve Enstit¼ Y¼netim Kurulunca ..... / ..... / ..... tarihi itibarıyla kabul edilmiřtir.

Prof. Dr. İsmail Hakkı MİRİCİ  
Eđitim Bilimleri Enstitüsü M¼d¼r¼

## Öz

Bu arařtırmada, çok boyutlu testlerden elde edilen puanların eřitlenmesinde kullanılan bi-faktör ÇB-MTK gözlenen puan eřitleme (BF-MIRT), tam ÇB-MTK gözlenen puan (MOSE) ve ÇB-MTK gözlenen puan eřitleme tek boyutlu yaklařım (AOSE) yöntemlerinden denk olmayan gruplarda ortak madde deseni altında elde edilen eřitlenmiř puanlar ve bu puanlara ait eřitlemenin standart hatası (SEE), yanlılık (BIAS) ve hata kareler ortalamasının karekökü (RMSE) deęerlerinin çeřitli faktörlere göre karřılařtırılması amaçlanmıřtır. Sonuçların, ÇB-MTK test eřitleme yöntemlerinden hangi kořullarda daha doęru sonuçlar elde edilebileceđine dair bazı yönergeler saęlaması hedeflenmiřtir. Arařtırmada simülasyon verileri kullanılmıřtır. Simülasyon veri setleri için örneklem büyüklüğü, ortak madde oranı, boyutlar arasındaki iliřki düzeyi, kalibrasyon ve test eřitleme yöntemleri manipüle edilen faktörler olarak ele alınmıřtır. 162 (3x3x3x2x3) kořul incelenmiřtir. Veri setlerinin üretilmesi ve eřitleme çalıřmaları R programlama dili kullanılarak gerçekeřtirilmiřtir. Eř zamanlı kalibrasyon yöntemi kullanılarak gerçekeřtirilen test eřitleme uygulamaları sonucunda elde edilen eřitlenmiř puanların benzer olduđu gözlenmiřtir. İki kalibrasyon yöntemi içinde örneklem büyüklüğü arttıkça hata deęerlerinin azaldıđı gözlenmiřtir. Örneklem büyüklüğünün en az 3000 olması önerilmektedir. Eř zamanlı kalibrasyon yöntemi kullanılarak gerçekeřtirilen test eřitleme uygulamalarında ortak madde oranı %20 olduęunda en küçük yanlılık deęerleri gözlenmiřtir. Ayrı zamanlı kalibrasyon yöntemi kullanılarak gerçekeřtirilen test eřitleme uygulamalarında ortak madde oranı %50 olduęunda en hata deęerleri gözlenmiřtir. Ortak madde oranı eř zamanlı kalibrasyonda en çok %20; ayrı kalibrasyonda en az %50 olmalıdır. Eř zamanlı kalibrasyonda ayrı kalibrasyon yönteminden daha küçük hata deęerleri gözlenmiřtir.

**Anahtar sözcükler:** test eřitleme, çok boyutlu test eřitleme, madde tepki kuramı, bi-faktör model, ölçek kalibrasyonu, eřitleme hatası

## Abstract

The purpose of this study is to compare the equated scores obtained from the bi-factor MIRT observed score equating (BF-MIRT), the full MIRT observed score equating (MOSE), and the unidimensional approximation observed score equating (AOSE) methods under common item non-equivalent groups design in terms of standard error of equating (SEE), BIAS, and root mean square error (RMSE) of these equated scores across various conditions. The results aim to provide guidelines on which MIRT test equating methods yield more accurate results under specific conditions. In the study, simulation data were used. Sample size, ratio of common items, level of relationship between dimensions, calibration methods, and test equating methods are the factors that were manipulated. 162 conditions were examined. The data generation and equating studies were conducted using the R programming language. It was observed that the equated scores obtained using concurrent calibration were similar. For both calibration methods, it was found that as the sample size increased, the error values decreased. A sample size of at least 3000 is recommended. In test equating applications performed using concurrent calibration, the smallest bias values were observed when the proportion of common items was 20%. In test equating applications performed using separate calibration, the lowest error values were observed when the proportion of common items was 50%. The proportion of common items should be at most 20% in concurrent calibration and at least 50% in separate calibration. Smaller error values were observed in concurrent calibration compared to separate calibration.

**Keywords:** test equating, multidimensional test equating, item response theory, bi-factor model, linking, equating error

## Teşekkür

Doktora sürecimin her döneminde desteğini hiç esirgemeyen, tez sürecinin tüm aşamalarında yanımda olan ve tüm sorularıma ayrıntılı olarak yanıt veren, yol gösteren ve rehberlik eden danışmanım Prof. Dr. Nuri DOĞAN'a,

Ölçme ve değerlendirme bilim dalında yüksek lisans eğitimime başladığım ilk günden bu yana bilgisini ve emeğini esirgmeden önerileri ve fikirleriyle beni yönlendiren, doktora tez izleme komitemde yer alarak kıymetli öneriler sunan Prof. Dr. Zekeriya NARTGÜN'e,

Doktora eğitimimde gelişmem ve ilerlememde büyük katkılar sunan, aldığım dersler aracılığıyla bilgisinden ve deneyiminden faydalandığım değerli hocalarım Prof. Dr. Selahattin GELBAL, Prof. Dr. Hülya KELECİOĞLU ve Prof. Dr. Burcu ATAR'a,

Tez izleme komitem ve jürimde bulunan, değerli önerileri ile tezimin gelişmesine katkıda bulunan Doç. Dr. Kübra ATALAY KABASAKAL'a

Lisansüstü eğitimim süresince üzerimde emeği olan tüm hocalarıma,

Bu zorlu süreçte beni yalnız bırakmayan, eğitim hayatımın her aşamasında bana destek olan, bugünlere gelmeme katkı sağlayan, sevgisini ve desteğini her zaman hissettiğim değerli annem Sevgi DEMİRÖZ'e çok teşekkür ederim..

**İçindekiler**

Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
Tablolar Dizini.....	viii
Şekiller Dizini.....	ix
Simgeler ve Kısaltmalar Dizini.....	x
Bölüm 1 Giriş.....	1
Problem Durumu.....	2
Araştırmanın Amacı ve Önemi.....	4
Araştırma Problemi.....	6
Sınırlılıklar.....	7
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	8
Boyutluluk.....	8
Test Eşitleme.....	11
İlgili Araştırmalar.....	32
Bölüm 3 Yöntem.....	44
Araştırmanın Türü.....	44
Test Eşitleme Deseni.....	44
Simülasyon Koşulları.....	45
Verilerin Üretilmesi.....	48
Verilerin Analizi.....	52
Değerlendirme Ölçütleri.....	53
Bölüm 4 Bulgular, Yorumlar ve Tartışma.....	55
Birinci Alt Probleme Ait Bulgular.....	55
İkinci Alt Probleme Ait Bulgular.....	57
Üçüncü Alt Probleme Ait Bulgular.....	90



Yorumlar ve Tartışma.....	93
Bölüm 5 Sonuç ve Öneriler.....	100
Sonuçlar .....	100
Öneriler .....	102
Kaynaklar .....	106
EK-A: Veri Setleri (Koşullar) .....	115
EK-B: Tek Yönlü ANOVA Sonuçları .....	116
EK-C: Post-Hoc Tukey Testi Sonuçları .....	117
EK-Ç: Araştırma Etik Komisyon İzin Muafiyeti Formu .....	119
EK-D: Etik Beyanı.....	120
EK-E: Doktora Tez Çalışması Orijinallik Raporu .....	121
EK-F: Thesis/Dissertation Originality Report .....	122
EK-G: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı .....	123

**Tablolar Dizini**

<b>Tablo 1</b> <i>Simülasyon Koşulları</i> .....	48
<b>Tablo 2</b> <i>İki yönlü ANOVA Sonuçları</i> .....	55
<b>Tablo 3</b> <i>Testin Kendisi ile Eşitlenmesi İki yönlü ANOVA Sonuçları</i> .....	90

**Şekiller Dizini**

<b>Şekil 1 Çok Boyutlu Ölçek Kalibrasyonu</b> .....	21
<b>Şekil 2 Eşitlenmiş Puanlar</b> .....	56
<b>Şekil 3 Eş Zamanlı Kalibrasyon SEE Değerleri</b> .....	58
<b>Şekil 4 Ayrı Kalibrasyon SEE Değerleri</b> .....	62
<b>Şekil 5 SEE ve Yöntem Etkileşimi</b> .....	67
<b>Şekil 6 Eş Zamanlı Kalibrasyon Yanlılık Değerleri</b> .....	68
<b>Şekil 7 Ayrı Kalibrasyon Yanlılık Değerleri</b> .....	73
<b>Şekil 8 Yanlılık ve Yöntem Etkileşimi</b> .....	79
<b>Şekil 9 Eş Zamanlı Kalibrasyon RMSE Değerleri</b> .....	80
<b>Şekil 10 Ayrı Kalibrasyon RMSE Değerleri</b> .....	84
<b>Şekil 11 RMSE ve Yöntem Etkileşimi</b> .....	89
<b>Şekil 12 Kendi ile Eşitlenmiş Puanlar</b> .....	92

## Simgeler ve Kısaltmalar Dizini

**AFA:** Açımlayıcı Faktör Analizi

**AOSE:** Gözlenen Puan Eşitleme Tek Boyutlu Yaklaşım

**ATSE:** Gerçek Puan Eşitleme Tek Boyutlu Yaklaşım

**BF-MIRT:** Bi-Faktör Çok Boyutlu Madde Tepki Kuramı

**BY-ÇBMTK:** Basit Yapılı Çok Boyutlu Madde Tepki Kuramı

**ÇBE:** Çok Boyutlu Test Eşitleme

**ÇB-MTK:** Çok Boyutlu Madde Tepki Kuramı

**DFA:** Doğrulayıcı Faktör Analizi

**BDS:** Beklenen Doğru Sayısı

**KTK:** Klasik Test Kuramı

**MOSE:** Tam Çok Boyutlu Madde Tepki Kuramı Gözlenen Puan Eşitleme

**MTK:** Madde Tepki Kuramı

**MTRS:** Ölçekleme Katsayılarının Eşzamanlı Kestirimi İçin Test Karakteristik Yüzeyini Eşleştirme

**RMSE:** Hata Kareler Ortalamasının Karekökü

**SEE:** Eşitlemenin Standart Hatası

**SMO:** Basit Yapılı Çok Boyutlu Madde Tepki Kuramı Gözlenen Puan

**SMT:** Basit Yapılı Çok Boyutlu Madde Tepki Kuramı Gerçek Puan Eşitleme

**TB-MTK:** Tek Boyutlu Madde Tepki Kuramı

**TBE:** Tek Boyutlu Test Eşitleme

**TRY:** Test Karakteristik Yüzeyi

## Bölüm 1

### Giriş

Eğitimde bireyler hakkında karar vermek için büyük ölçekli standart testlerden faydalanılmaktadır. Testlerden elde edilen puanlar genellikle sertifikasyon, eğitim programına kabul veya mesleğe yerleşme gibi yüksek riskli kararlar vermek için kullanılır. Standartlaştırılmış testlerin sonuçlarının son derece önemli kararlar almada kullanılması nedeniyle test puanları, yetenek veya becerilerin en doğru kestirimlerini yansıtmalıdır. Adil bir değerlendirmenin oluşturulması da çok önemlidir. Adaleti sağlamak için aynı test tüm popülasyona birden fazla kez uygulanabilir. Fakat bu durumda test güvenliğinin sağlanması imkânsız hale gelecektir. Test güvenliğini sağlamak için büyük ölçekli standart testler uygulanırken farklı zamanlarda aynı testin farklı formları kullanılarak uygulamalar gerçekleştirilir. Örneğin ALES, YDS ve SAT gibi büyük ölçekli ve sonuçları birkaç yıl geçerli olan testlerin büyük bir kısmı yılda birkaç defa tekrarlanmaktadır. Bu uygulamalarda test güvenliğinin sağlanması amacıyla alternatif test formları kullanılmaktadır. Bireyler farklı dönemlerde uygulanan bu testlerden aldıkları puanlar ile çeşitli kurumlara başvurmaktadır. Bu nedenle farklı test formlarından elde edilen puanların karşılaştırılabilir olması gerekmektedir. Puanların karşılaştırılabilir olması için paralel/eşdeğer test formlarından faydalanılır. Paralel test formları aynı içeriğe ve istatistiksel özelliklere uyacak şekilde geliştirilir. Fakat madde güçlüğü gibi istatistiksel özellikler açısından tamamen eşdeğer olan test formları oluşturmak neredeyse imkansızdır. Test formları arasındaki güçlük farklılıklarından dolayı bazı bireyler daha kolay olan test formunu alır. Daha kolay test formunu alan bireyler avantaj elde etmiş olur. Sınava girenler genellikle rekabet içinde olduklarından veya puanları birlikte karşılaştırıldıklarından daha zor test formunu alan bir aday için bu durum haksızlık olacaktır. Haksızlığı engellemek için test formlarındaki güçlük farklılıklarını düzelterek bir düzenleme yapılması gerekmektedir. Test eşitleme, alternatif test formlarındaki puanların birbirinin yerine kullanılabilmesi amacıyla, güçlüklerdeki bu farklılıkları düzenlemek için kullanılan istatistiksel bir yöntemdir (Kolen & Brennan, 2004).

Eşitleme süreci test ve madde özelliklerini düzenler. Böylece hangi test formunun uygulandığına bakılmaksızın dağılımlar aynı olur. Test eşitleme sonucunda testlerin güçlük farklarından kaynaklanan etkiler engellenmiş ve puanlar aynı ölçek üzerine yerleştirilmiş olacaktır.

Güvenlik nedenleriyle farklı formları uygulanan testler için test eşitleme yöntemleri geliştirilmiştir. Eşitleme yöntemleri dayandıkları kurama göre farklılık göstermektedir. Klasik test kuramına dayanan yöntemler: Ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelli eşitleme. Madde tepki kuramına (MTK) dayalı yöntemler: Gerçek puan eşitleme ve gözlenen puan eşitleme. Madde tepki kuramına dayalı yöntemler ile eşitleme yapılmadan önce ölçek kalibrasyonu yapılması gerekmektedir. Test formlarından elde edilen parametrelerinin aynı ölçekte yer alması için yapılan işlemlere ölçek kalibrasyonu denir. Ölçek kalibrasyon yöntemleri eş zamanlı ve ayrı kalibrasyon olmak üzere iki başlık altında toplanır. Kalibrasyonlar tamamlandıktan sonra eşitleme yapılabilir.

### **Problem Durumu**

Kullanılacak istatistik yöntemler verilerin yapısına ve araştırma problemine göre belirlenmektedir. Test formları arasında eşitleme yaparken de veri yapısına göre model seçilmelidir. Test eşitleme yönteminin başarısı kısmen, belirli bir modelle ilgili varsayımların ne ölçüde sağlandığı ile ölçülür. MTK modelleri, özellikle madde yanıtlarının tek boyutlu olduğu şeklindeki güçlü varsayımlara dayanmaktadır. Tek boyutluluk, bir testte veya ölçme aracında ölçülmek istenen baskın tek bir özelliğin, yeteneğin olmasıdır. Birçok test durumunda tek boyutluluk varsayımı çoğunlukla ihlal edilmektedir (Li & Lissitz, 2000). Bu durumda varsayım ihlali hatalı sonuçlara neden olabilir. Örneğin tek boyutlu madde tepki kuramında tek boyutluluk varsayımının ihlal edilmesi durumunda madde parametreleri ve testi alan kişinin yetenek değerlendirmelerine ait parametrelerde ciddi sapmalar meydana gelebilir (Ackerman, 1989). Güvenilir sonuçlar elde edebilmek için verilerin tek

boyutluluğunu gerektiren psikometrik yöntemlerin uygulanmasından önce tek boyutluluk varsayımının doğrulanması gereklidir (Gao, 1997).

Pek çok test doğası gereği çok boyutludur ve bazen belirlenemeyen çok sayıda boyutluluk kaynağı vardır. Reckase (2009), çok boyutluluğun hem incelenen örneklemin hem de testin özelliklerine bağlı olduğunu belirtmiştir. Örneğin çoktan seçmeli ve açık uçlu soruların yer aldığı testlerde madde türü boyutluluğa neden olabilmektedir. Bazı büyük ölçekli testler ise farklı özellikleri ölçen alt testlere sahiptirler. Örneğin PISA her bir döngüsünde bir alan ağırlıklı alan olarak seçilmekte ve o alanda derinlemesine analizler gerçekleştirilmektedir. Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) sözel ve sayısal olmak üzere iki alt testten oluşmaktadır. Bu sınavlarda içerik alanı boyutluluk kaynağıdır. Bir testte birden fazla boyut ölçülüyorsa, karmaşık çok boyutlu alanın tanımlanması şartıyla, çok boyutlu madde tepki kuramı (ÇB-MTK) çerçevesi, yetenekleri ve maddeleri tek boyutlu bir madde tepki kuramı (TB-MTK) çerçevesinden daha doğru bir şekilde modelleyebilir. Testin tek boyutlu olmadığı gösterildiğinde ÇB-MTK modellerinin TB-MTK modellerine yerine kullanılması, varsayım ihlallerinin etkilerini azaltmak için bir yoldur.

ÇB-MTK TB-MTK'nın bir uzantısı olarak ortaya çıkmıştır ve daha karmaşık madde tepki modellerine olan ihtiyaçtan geliştirilmiştir. ÇB-MTK, veri yapısı iki veya daha fazla boyuttan oluştuğunda, madde ile birey arasındaki etkileşimi inceleyen bir modeldir. ÇB-MTK modelleri, bir testteki maddelerin birkaç gizil özelliği/yeteneği ölçmesine izin vererek daha esneklerdir. Veri yapısı bi-faktör model ve madde takımı yanıt modeli tarafından daha doğru açıklandığında ise bu modeller tercih edilir. Bi-faktör modelde, maddeler bir genel faktörü ve bir özel faktörü ölçer. Genel ve özel faktörler aynı düzeydedir. Madde takımı yanıt modeli, bi-faktör modelin kısıtlanmış bir versiyonudur. Belirli boyuttaki yüklerin, her madde parçasındaki genel boyuttaki yüklerle orantılı olacak şekilde sınırlandırılmasıyla elde edilir (Rijmen, 2010). Bi-faktör ve madde takımı yanıt modeli gibi modeller, ÇB-MTK modellerinin özel ve kısıtlanmış durumları olarak kabul edilmektedir. Çünkü maddeler birden fazla boyutu ölçerken iki boyuttan fazlası ölçülmez. ÇB-MTK modelleri ayrıca, maddelerin yalnızca bir

boyutta veya bir boyutta çok yüksek ve diğer boyutta/boyutlarda ihmal edilebilir düzeyde ilişkili olduğu basit bir yapı veya yaklaşık basit yapı ilişkisi belirtebilir. Basit yapıli modellerde maddelerin tümü aynı boyutlar tarafından ölçülmediği için test çok boyutlu kabul edilir.

ÇB-MTK'nın geliştirilmesiyle birlikte çok boyutlu testler için kalibrasyon ve eşitleme problemine çözümler üretilmeye başlanmıştır. ÇB-MTK çerçevesi içerisinde çok boyutlu test eşitlemeye yönelik ilk çalışmalardan biri, random gruplar deseni altında çok boyutlu iki parametrelili bir lojistik modelle bağlantılı olarak gözlenen ve gerçek puan eşitleme prosedürlerini öneren Brossman (2010) tarafından yapılan çalışmadır. Brossman (2010) karmaşık bir ÇB-MTK modeli kullanmıştır. Tam ÇB-MTK gözlenen puan eşitleme (MOSE), ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım (AOSE), ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşım (ATSE) yöntemlerini geliştirmiştir. Basit yapıli ÇB-MTK gözlenen puan (SMO) (Lee & Brossman, 2012), basit yapıli ÇB-MTK gerçek puan eşitleme (SMT) (Kim vd., 2019), bi-faktör ÇB-MTK gözlenen puan (Lee & Lee, 2016), bi-faktör ÇB-MTK gerçek puan (Lee vd., 2015), madde takımı yanıt modeli ÇB-MTK gözlenen puan ve madde takımı yanıt modeli ÇB-MTK gerçek puan (Tao & Cao, 2016) yöntemleri de ilerleyen süreçlerde geliştirilmiştir. Çok boyutlu testlerin kullanımı yaygınlaşsa da ÇB-MTK test eşitleme yöntemlerinin gelişimi yavaş ilerlemiştir. Alan yazın incelendiğinde var olan yöntemlerin geliştirilmeye çalışıldığı, yeni yöntemler önerilmeye devam edildiği görülmektedir (Kim & Lee, 2022).

### **Araştırmanın Amacı ve Önemi**

Günümüzde eğitsel veya psikolojik testlerde birden çok özellik ölçülmektedir. ÇB-MTK modelleri son yıllarda daha fazla ilgi görmektedir ve testlerin psikometrik özelliklerini analiz etmek için kullanılmaktadır. Bununla birlikte, ÇB-MTK modelleri kullanıldığı zaman ölçek kalibrasyonu ve test eşitleme çok daha karmaşık hale gelmektedir. Alan yazın incelemesinde, tek boyutlu test eşitleme konusunda çok fazla sayıda çalışma olmasına rağmen çok boyutlu test eşitleme için erişilebilen çalışmaların sınırlı sayıda olduğu



görülmüştür. Mevcut çalışmalarda çok boyutlu test eşitleme yöntemleri geliştirilmekte ve önerilen yöntemlerin performansları incelenmeye devam edilmektedir.

Tek boyutluluk varsayımın ihlal derecesine bağlı olarak yani boyutluluk arttıkça, örneklem büyüklüğü, yetenek dağılımı, testteki ortak madde oranı koşullarına bağlı olarak eşitleme yöntemlerinin performansı etkilenecektir. Eşitleme yöntemlerinin performanslarının nasıl etkilendiğini tam olarak anlamak için bağımsız değişkenlerin manipülasyonuna izin veren araştırma çalışmaları yapılmaktadır. Bunun için simülasyon çalışmalarından faydalanılır. Simülasyon çalışmaları, istenen değişkenleri manipüle etmek ve kontrol etmek için bir teknik sunar. Tam ÇB-MTK gözlenen puan, MTK gözlenen puan eşitleme tek boyutlu yaklaşım eşitleme yöntemleri Brossman (2010) ve Bi-faktör ÇB-MTK gözlenen puan eşitleme yöntemi Lee ve Lee (2016) tarafından geliştirilmiş olmasına rağmen bu yöntemlerin çeşitli koşullar altında nasıl performans gösterdiğini belirlemek için yeterince simülasyon çalışması yapılmamıştır. Alan yazın incelendiğinde denk olmayan gruplarda ortak madde deseni altında gerçekleştirilen çok az çalışmaya ulaşılabilmektedir. Yapılan çalışmaların genellikle random grup deseni altında test eşitleme uygulamaları olduğu belirlenmiştir. Bunun yansıra kalibrasyon yöntemlerinin kullanılan test eşitleme yöntemlerine etkisinin de incelenmesi gerektiği belirlenmiştir. Bu nedenle, farklı test eşitleme desenleri altında gözlenen puan eşitleme yöntemlerinin performansının daha fazla araştırılması ve karşılaştırılması gereklidir. Bu durum göz önünde bulundurularak bu çalışmada, çok boyutlu testlerden elde edilen puanların eşitlenmesinde kullanılan bi-faktör ÇB-MTK gözlenen puan eşitleme (BF-MIRT), tam ÇB-MTK gözlenen puan (MOSE) ve ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım (AOSE) yöntemlerinden denk olmayan gruplarda ortak madde deseni altında elde edilen eşitlenmiş puanlar ve bu puanlara ait eşitlemenin standart hatası (SEE), yanlılık (BIAS) ve hata kareler ortalamasının karekökü (RMSE) değerlerinin çeşitli faktörlere göre karşılaştırılması amaçlanmıştır. Karşılaştırma sonuçlarına göre test eşitleme yöntemlerinden hangi koşullarda daha doğru sonuçlar elde edilebileceğine dair bazı yönergeler sağlanması da hedeflenmiştir.

Eşitleme hatası, eşitleme sonuçlarının ve alınacak kararların doğruluğunu etkilemektedir. Eşitleme hatası çeşitli faktörlerden etkilenmektedir. Testlerin çok boyutlu yapısı göz önüne alındığında, farklı çok boyutlu ölçek kalibrasyon ve test eşitleme yöntemlerinin performansını derinlemesine araştırmak ve farklı koşullar altında hangi yöntemlerin daha iyi performans gösterdiğini, hata değerlerinin daha küçük olduğunu gözlemlemek kritik öneme sahiptir. Bu bağlamda bu araştırma denk olmayan gruplarda ortak madde deseninde hangi koşullar altında hangi yöntemlerin daha az hata değerleri gösterdiğine yönelik bilgiler sunması, uygulayıcılarına yönergeler sağlaması ve literatürdeki boşluğu kapatması bakımından önemli görülmektedir.

### **Araştırma Problemi**

Çok boyutlu Madde Tepki Kuramına dayalı test eşitleme yöntemlerinden elde edilen eşitlenmiş puanlar ve bu puanlara ait yanlılık (BIAS), eşitlemenin standart hatası (SEE) ve hata kareler ortalamasının karekökü (RMSE) değerleri farklı koşullara bağlı olarak nasıl değişmektedir?

### **Alt Problemler**

Araştırmanın problem cümlesi doğrultusunda aşağıdaki alt problemlere de yanıt aranmıştır.

1. Denk olmayan gruplarda ortak madde deseni altında tam ÇB-MTK gözlenen puan, ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım ve bi-faktör ÇB-MTK gözlenen puan eşitleme yöntemleri kullanılarak elde edilen eşitlenmiş puanlar farklılaşmaktadır mıdır?

2. Denk olmayan gruplarda ortak madde deseni altında tam ÇB-MTK gözlenen puan eşitleme, ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım ve bi-faktör ÇB-MTK gözlenen puan eşitleme yöntemleri kullanılarak elde edilen eşitlenmiş puanlar için yanlılık (BIAS), eşitlemenin standart hatası (SEE) ve hata kareler ortalamasının karekökü (RMSE) değerleri,

- a) Örneklem büyüklüğüne (500, 1500 ve 3000) göre nasıl değişmektedir?
- b) Ortak madde oranına (%20, %35 ve %50) göre nasıl değişmektedir?
- c) Çok boyutlu eş zamanlı kalibrasyon ve ayrı kalibrasyon (Stocking Lord) yöntemlerine göre nasıl değişmektedir?
- d) Boyutlar arasındaki ilişki düzeyine ( $\rho=0$ ,  $\rho=0,30$  ve  $\rho=0,60$ ) göre nasıl değişmektedir?

3. Denk olmayan gruplarda ortak madde deseni altında tam ÇB-MTK gözlenen puan eşitleme, ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım ve bi-faktör ÇB-MTK gözlenen puan eşitleme yöntemleri kullanılarak testin kendisi ile eşitlenmesinden elde edilen eşitlenmiş puanlar farklılaşmakta mıdır?

#### **Sınırlılıklar**

- 1) Araştırma ikili puanlanan veriler ile sınırlıdır.
- 2) Simülasyon çalışması örneklem büyüklüğü, ortak madde oranı, ölçek kalibrasyon yöntemleri, boyutlar arasındaki ilişki düzeyi ve çok boyutlu test eşitleme yöntemleri koşulları ile sınırlıdır.
- 3) Madde yanıt setleri telafi edici çok boyutlu üç parametrelili lojistik (M-3PLM) madde tepki kuramı modeli kullanılarak üretilmiştir.
- 4) Her boyut 30, tüm test ise 60 madde ile sınırlandırılmıştır.

## Bölüm 2

### Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

#### Boyutluluk

Boyutluluk, belirli bir veri setindeki madde yanıtları arasındaki korelasyonları dikkate alarak belirlenen gizil değişkenler olarak tanımlanabilir (Camilli vd., 1995). Tindal ve Haladyna (2002) boyut kavramını test maddeleri tarafından ölçülen minimum yetenek sayısı olarak tanımlamaktadır. Kline'e (2005) göre ise somut ve gözlenebilir olmayan, doğrudan ölçülemeyen; soyut veya gizil nitelik taşıyan psikolojik özelliklere "boyut" ya da "faktör" adı verilir. Bir başka ifade ile niteliksel olarak farklı alanlara, yani bir alandan diğerine nispeten az genelleme yapılabilen alanlara ayrı faktörler denir (Gorsuch, 2015). Standartlar'a (1999) göre boyutluluğun incelenmesi test geçerliği için önemlidir. Yapı geçerliği kanıtlarının toplanması test boyutlarının bir değerlendirmesini içerir. Testin geçerliğini tehdit edecek testin yapısı ile ilgisi olmayan olası faktörler veya gizil değişkenler kontrol edilmelidir. Yani, yanıt matrisinin boyutsal yapısının, bireylerin performansın altında yattığı varsayılan özellikler ile ne ölçüde tutarlı olduğu incelenmelidir.

Yapı geçerliğinin yanı sıra çoğu analiz yöntemi tek boyutluluk varsayımına sahiptir. Örneğin puanların güvenilirliğini belirlemede kullanılan birçok yöntem testin tek boyutlu olduğunu varsayar. Test geliştiriciler tek boyutluluk varsayımının ihlali halinde test edilen güvenilirliğin gerçek güvenilirliği doğru bir şekilde temsil edip etmediğini sorgulamalıdır. Ayrıca, bir testin boyutsal yapısını anlamak, eşitleme ve kalibrasyon gibi diğer psikometrik çalışmalar için önemlidir. Alt puanların doğru yorumlanması ve alt puanların öğretim faaliyetlerinde kullanılması için boyutluluğun değerlendirilmesi önemlidir.

Boyutluluk test ve madde boyutluluğu olmak üzere ikiye ayrılır. Testin boyutluluğu, geniş içerik alanlarını kapsayan yetenekleri modellemek için kullanılırken, madde boyutluluğu, sadece bir madde veya yalnızca birkaç madde ile ilişkilendirilen yetenekleri modellemek için kullanılır. Bu araştırmada boyutluluk ifadesi testin boyutluluğunu

anlamında kullanılmıştır. Testin boyutluluğunu belirlemede faktör analizi, paralel analiz, yamaç birikinti grafiği, Stout DIMTEST ve POLYTEST vb. yöntemler kullanılmaktadır.

### ***Faktör Analizi***

Faktör analizi yapı geçerliği için kanıt toplamada büyük rol oynamaktadır. Teorik olarak temel bir yapıyı temsil eden bir dizi test maddesi arasındaki karşılıklı ilişkileri hesaba katmak için gereken minimum faktör veya gizil değişken sayısını belirleme amacı ile bir değişken indirgeme tekniğidir (Price, 2017). Faktör analizinin genel amacı, değişkenler arasındaki ilişkileri, kavramsallaştırmaya yardımcı olmak için kısa ama doğru bir şekilde özetlemektir. Bir başka amacı ise çok sayıdaki gözlenen değişkenleri, daha az sayıdaki boyutlar altında bir araya getirmektir (Tabachnick & Fidell, 2006). Faktör analizi açımlayıcı (AFA) ve doğrulayıcı faktör analizi (DFA) olmak üzere iki gruba ayrılır. AFA'da amaç, testi yanıtlayan bireylerin özellikleri olarak ortaya çıkan belirli yapıları ölçmek için tasarlanmış bir dizi test maddesinden (veya alt test toplam puanlarından) bir dizi faktör belirlemektir. AFA'da önceden herhangi bir teorik yapı öne sürülmez. Bunun yerine araştırmacı, bir dizi temel yapıyı ölçmek için tasarlanmış geniş bir test maddesi grubuna verilen yanıtları kullanarak bir faktör analizi gerçekleştirir. AFA maddeler arasındaki korelasyona dayalı bir analizdir. Test maddelerinin yüksek veya düşük korelasyon göstermelerine göre gruplaşmalarını inceleyerek hangi maddelerin hangi faktörden etkilendiğini inceler. Birbirleri ile yüksek korelasyon gösteren maddelerin aynı gizil değişken yani aynı faktörden etkilendiğini varsayarak bu maddeleri aynı faktör altında bir araya getirir. Testte yer alan maddelerin faktörler ile olan korelasyonlarına faktör yükü denir (Crocker & Algina, 2008). Bir maddenin faktör ile ilişkili olduğunun söylenebilmesi için faktör yük değerini en az 0,30 düzeyinde olması gerekmektedir. Maddelerin o faktöre ait yüklerinin kareleri toplamı faktörün öz değeri olarak adlandırılır. Faktör sayısı belirlemede öz değerlerden yararlanılmaktadır. Faktör sayısını belirleme faktör analizindeki önemli aşamalardan biridir. Kaiser (1960) K1 kuralını önermiştir. Bu kurala göre öz değeri 1'den büyük olan değişken sayısı faktör sayısı olarak değerlendirilmektedir. AFA ile maddelerin faktör ile olan ilişkileri

belirlendikten sonra DFA yapılabilir. DFA, bir testin neyi ölçtüğüne özgü belirlenen faktör yapısını doğrulamak veya onaylamamada güçlü bir çerçeve sağladığı için yararlı bir tekniktir. DFA tanımlanan yapı ile gözlenen verinin uyumunu inceler (Tinsley & Brown, 2000). Model veri uyumu sağlanıyorsa faktör yapısı doğrulanmış olarak kabul edilir. Fakat alan yazındaki çalışmalar incelendiğinde DFA'nın yanlış faktör yapılarını da doğrulama eğilimi olduğu gözlenmiştir (Orçan, 2018). Bu nedenle faktör yapısının alan yazın veya farklı yöntemlerden elde edilen bulgular ile desteklenmesi önerilmektedir.

### ***Paralel Analiz***

Paralel analiz yöntemi Horn (1965) tarafından önerilen ve random veri üretilmesine dayanan bir yöntemdir. Paralel analizde gerçek veri setine benzer olacak şekilde Monte Carlo simülasyon çalışması yürütülür. Simülasyon sonucunda üretilen veri setlerindeki bileşenlerin öz değerleri hesaplanır. Karşılaştırmada eşik değeri için Horn (1965) rassal olarak üretilen matrislerden elde edilen öz değerlerin ortalamasını almayı önermiştir. Fakat ortalama yöntemi olduğundan daha fazla faktör bulunmasına neden olmaktadır (Kılıç & Uysal, 2021). Bu nedenle üretilen veriye ait öz değerlerin gerçek veri setindeki öz değerden büyük olduğu nokta anlamlı faktör sayısı olarak belirlenir (Ledesma & Mora, 2007). Faktör analizi için faktör sayısını belirlemek amacıyla kullanılabilir.

### ***Yamaç Birikinti Grafiği***

Cattell (1966) tarafından önerilen yöntem faktör öz değerlerini temel almaktadır. Grafiğin x eksenini bileşen sayısını ve y eksenini bu bileşenlerin öz değerlerini göstermektedir. Bir noktadan sonra grafiğin eğimi sifıra yaklaşır. Eğri üzerindeki bu nokta, tutulacak maksimum bileşen sayısını gösterir (Ledesma vd., 2015). Yaklaşım basit ve genel olarak kullanışlı olsa da öznel yorumlamalara açık olması nedeniyle eleştirilmektedir.

### ***Dimtest***

Stout (1987) tarafından geliştirilen yöntem testin tek boyutlu olup olmadığını belirlemede kullanılır. Testin tek boyutluluğunu sınamak için yokluk hipotezi kurulur.

$H_0$  : Test tek boyutludur.

Yokluk hipotezi t istatistiği ile sınırlanır. Kovaryans matrisleri ile t istatistiği hesaplanır. DIMTEST t istatistiği sıfıra eşit veya yakınsa yokluk hipotezi kabul edilir (Güler, 2017).  $\alpha$  belirlenen anlamlılık düzeyi olmak üzere t'nin istatistiği standart normal dağılımın  $100*(1-\alpha)$ 'inci yüzdeleri diliminden büyükse, sıfır hipotezi  $\alpha$  anlamlılık düzeyinde reddedilir (Kaplan, 2004).

DIMTEST t istatistiği ikili puanlanan veri setlerinde tek boyutluluğun değerlendirilmesinde kullanılırken, yaklaşık basit yapıya çok boyutlu doğrulayıcı modellerde de boyutluluk değerlendirmesi yapmak amacıyla kullanılabilir (Svetina, 2011). Orijinal DIMTEST yöntemine dayalı olarak, Poly-DIMTEST, Li (1995) tarafından hem ikili hem de çoklu puanlanan maddelerin yer aldığı testlerin tek boyutluluğunu hipotez testi ile sınamak amacıyla geliştirilmiştir.

### **Test Eşitleme**

Test puanlarının diğer test formundan elde edilen puanlarla ilişkilendirilmesine veya dönüştürülmesine test eşitleme denir (Hambleton & Swaminathan, 1985). Angoff'a (1984) göre test eşitleme, bir test formunun birim sistemini diğer bir test formunun birim sistemine dönüştürmektir. Peterson vd. (1993) ise test eşitlemeyi, bir test formundaki ham puanların başka bir test formundaki ham puanlarla ifade edilerek, bu iki test formu arasında ilişki kurulan deneysel bir süreç olarak tanımlamıştır. Kolen ve Brennan'a (2004) göre eşitleme, sınav formları arasındaki puanları doğru ve adil bir şekilde karşılaştırabilmek için istatistiksel yöntemler kullanılarak sınava girenlerin puanlarının test formları arasında dönüştürülmesidir. Bir başka ifadeyle test eşitleme, benzer içerik ve benzer güçlük düzeyindeki test formları arasındaki farklılıkları düzenleyerek bu test formlarından elde edilen puanların birbiri yerine kullanılmasını sağlayan istatistiksel bir süreçtir (Kolen & Brennan, 2014). Eşitleme test formlarının zorluklarına göre ayarlama yapmaktadır. Test formlarının içeriklerine göre ayarlama yapılmaz. Test eşitlemenin genel amacı aynı veya

farklı zamanlarda bir testin farklı formlarını alan bireylerin beraber değerlendirilebilmesini sağlamaktır. Farklı test formlarından elde edilen puanları aynı ölçekte sıralanmasını, karşılaştırılabilmesini ve birbiri yerine kullanılabilmesini sağlar.

İki test formundan elde edilen sonuçların eşitlenebilmesi için formların belirli koşulları karşılaması gerekmektedir. Bu koşullar aynı yapıyı ölçme, benzer güvenirlik, simetriklik, eşitlik ve grup değişmezliğidir (Dorans & Holland, 2000). Test formların aynı yapıyı ölçmesi aynı özelliği ya da yeteneği ölçmesi anlamına gelmektedir. Farklı içerik ve özelliği ölçen test formları eşitlenemez. Örneğin matematik testi ile fen bilimleri testi eşitlenemez. İstatistiksel olarak puanlar birbirine dönüştürülse bile puanlar aynı anlama gelmeyecektir. Benzer güvenirlik koşulu test formlarının güvenirliklerinin eşit veya yakın olmasıdır. Test formlarının güvenirlikleri yükseldikçe eşitleme çalışmaları daha iyi sonuçlar verecektir. Simetriklik koşulu test formlarından hangisinin referans form olduğu fark etmeksizin eşitlemenin yapılabilmesidir (Hambleton vd., 1991). X formunun Y formuna, Y formunun da X formuna dönüştürülebiliyor olmasına denir. X formundan Y formuna ya da Y formundan X formuna eşitlemesinden elde edilen eşitlenmiş puanlar benzer olur. Eşitlik koşulu sınava giren bireyin herhangi bir test formunu almasının farklılık oluşturmadığını anlamına gelmektedir. Her bir yetenek düzeyinde, X puanlarının koşullu frekans dağılımı, Y formunun dönüştürülmüş puanlarının koşullu frekans dağılımı ile aynı olmalıdır (Kolen & Brennan, 2014). Grup değişmezliği eşitleme yapmak için kurulacak olan eşitleme bağıntısının, test formlarının uygulandığı gruplardan bağımsız olmasıdır.

Eşitleme koşullarının bazıları test formları birbirine eşitlendikten sonra kontrol edilebilmektedir. Bu nedenle koşulların sağlanması eşitlemenin doğru bir şekilde yapılmasına da bağlıdır. Test eşitleme süreci aşağıdaki adımlardan oluşmaktadır:

- 1.Eşitleme deseninin seçilmesi
- 2.Eşitleme desenine göre test formlarının uygulanması
- 3.Test eşitleme yönteminin seçilmesi ve uygulanması



#### 4.Eşitleme sonuçlarının değerlendirilmesi

##### **Test Eşitleme Desenleri**

**Tek Grup Deseni.** Tek grup deseninde iki test formu da tüm bireylere uygulanır. İki test formu tüm bireylere uygulandığı için test formları arasındaki güçlük farklılığının etkisi ortadan kalkmaktadır. Bireylerin yeteneklerindeki değişim kontrol edilebilmektedir. Küçük örneklemelerde dahi küçük eşitleme hatası elde edilir. Desenin sınırlılığı ise iki uygulama yapılmasından dolayı daha çok süre gerektirmesi ve bu süre içerisinde bireylerin performanslarının yorgunluk gibi durumlardan etkilenbilmesidir. İlk test formu öğrenme gibi nedenler ile ikinci test formunun sonuçlarını etkileyebilmektedir. Bu etkiye sıra etkisi denilmektedir.

**Random Grup Deseni.** Random grup deseninde ortak bir evrenden gelen bireyler seçkisiz olarak iki gruba ayrılır. Gruplardan birisine X formu uygulanırken diğerine Y formu uygulanır. Böylece ortak bir evrenden gelen bireyler, benzer X ve Y formlarına seçkisiz olarak atanmış olur (Kolen & Brennan, 2014). Grupların yetenek düzeylerinin birbirine eşit olduğu söylenemez. Her bireye tek test formu uygulandığı için uygulama süresi kısalmış olur. Fakat bu özellik daha büyük örneklemelere ihtiyaç duyulmasına neden olmaktadır. Random grup deseninde, büyük bir heterojen grup seçilmelidir. Sıra etkisi gözlenmemesi random grup deseninin en önemli avantajıdır. Random grup deseni kullanıldığında iki test formunun uygulandığı gruplar arasındaki performans farklılıkları, test formları arasındaki güçlük farklılığının doğrudan bir göstergesidir.

**Dengelenmiş Tek Grup Deseni.** Tek grup deseninin sıra etkisi sınırlılığını önlemek amacıyla geliştirilmiştir. Aynı evrenden gelen bireyler seçkisiz olarak iki gruba ayrılır. İlk gruba önce X formu sonra Y formu uygulanırken ikinci gruba önce Y sonra X formu uygulanır. Her iki test formu için eşit süre verilmektedir. Dengelenmiş tek grup deseni, tek ve random grup desenlerinin bir araya getirilmiş şeklidir. Random grup desenine nazaran daha küçük örneklemelerde kullanılabilir.

**Denk Olmayan Gruplarda Ortak Madde Deseni.** Denk olmayan gruplarda ortak madde deseni hem esnek hem de en karmaşık test eşitleme desenlerinden biridir. İki gruba uygulanan farklı test formları ortak maddeler aracılığı ile eşitlenir. Her bireye tek test formu uygulanır. Böylece zaman tasarrufu sağlar. Sıra etkisi oluşmaz. Ortak (ankor) maddeler, her bir test formunu alan gruplar arasındaki yetenek farklılıklarını ortadan kaldırmak amacıyla kullanılır. Ortak maddeler, test formlarıyla aynı özelliklere sahip olmalıdır (Choi, 2019). Ortak maddeler aynı yapıyı ölçecek şekilde hazırlanmalıdır. Toplam testle aynı sayıda boyuta sahip olmalıdır. Boyutluluğun kaynağı madde türü ise, ortak maddeler tüm madde türlerini içermelidir. Boyutluluğun kaynağı içerik alanı ise ortak madde seti her bir içerik alanını kapsamalıdır. Bu şekilde, aynı boyut yapısına sahip ortak maddeler, sadece form zorluğundan kaynaklanan puan farklılıklarını düzeltebilir. Ortak maddeler aynı yapıyı ölçer ise yanlılık engellenir. Ortak maddeler testin tüm içeriğini yeterli düzeyde kapsar ise eşitlemelerin doğruluğunun artmasına yardımcı olur. Maddelerin test içerisindeki yerleri de madde parametrelerini etkileyebileceği için ortak maddelerin test formlarındaki yerleri aynı olmalıdır (Kolen & Brennan, 2004). Gruplar için orijin ve ölçü birimi (ortalama ve standart sapma) ile boyutlar arasındaki korelasyon farklı olabilir.

Test eşitleme deseni belirlendikten ve uygulama yapılarak veriler toplandıktan sonra test eşitleme aşamasına geçilir. Hangi yöntemin tercih edileceğine karar verilirken testin boyutluluğu göz önünde bulundurulmalıdır. Çok boyutlu testlerde tek boyutlu eşitleme yöntemlerinin kullanılması hatalı sonuçlara neden olabilir. Çok boyutluluğun derecesi kalibrasyon yöntemlerini etkilemektedir (Güler, 2019). Çok boyutluluğun derecesinin yüksek olduğu koşullarda eşitlenmiş puanların çok büyük ya da çok küçük kestirildiği belirlenmiştir (Livingston vd., 1990). Goldstein ve Wood (1989), ÇB-MTK'ya dayalı eşitleme yöntemlerinin kalitesi üzerindeki etkisinin, her iki test için de madde yanıtlarının altında yatan gizil özelliklerin aynı lineer bileşimi veya referans bileşimi olduğu sürece ihmal edilebilir olduğunu belirtmiştir. Fakat bu aynı lineer bileşim her zaman gözlenemeyebilir.

### ***Test Eşitleme Yöntemleri***

**Tek Boyutlu Test Eşitleme (TBE).** Tek boyutlu testlerde (TBE) eşitleme yöntemleri klasik test kuramına ve madde tepki kuramına dayalı yöntemler olmak üzere ikiye ayrılır.

**Klasik Test Kuramına Dayalı Eşitleme Yöntemleri.** Klasik test kuramına göre eşitleme yöntemleri ortalama eşitleme, doğrusal eşitleme ve eşit yüzdellikli eşitlemedir. Her yöntemin kendi varsayımları bulunmaktadır. Kullanılan eşitleme desenine göre eşitleme formülleri değişiklik göstermektedir.

**Ortalama eşitleme yöntemi.** Ortalama eşitleme yönteminde test formlarının madde güçlükleri arasındaki farkın tüm ölçek boyunca aynı olduğu varsayılır. Başka bir ifade ile X formunun uygulanmasından elde edilen puanların Y formunun uygulanmasından elde edilen puanlardan sabit bir güçlük düzeyi kadar farklı olduğu varsayılır. X formunun Y formuna eşitlenmesi ile elde edilen eşitlenmiş puan, X formundan elde edilen ham puandan X formunun ortalamasının çıkarılıp Y formunun ortalamasının eklenmesi ile bulunur. Eşitleme sonucunda elde edilen puanlar ile Y formuna ait puanların ortalamaları eşitlenirken merkezi dağılım ölçülerinde değişiklik olmamaktadır.

**Doğrusal eşitleme yöntemi.** Doğrusal eşitleme yöntemi grup dağılımlarının aynı olduğunu varsayar. X ve Y formlarından elde edilen puanların yani grupların standart sapma ve aritmetik ortalamalarının eşit olduğu varsayılır. Bu iki şart sağlanmaz ise ham puanlar arasında bir dönüşüm elde edilemez (Kelecioğlu, 1994). Standart puanlara dayalı bir yöntemdir. Aynı z puanına sahip puanlar birbirine denk kabul edilir. Doğrusal eşitleme yöntemden elde edilen eşitlenmiş puanlar ölçek puanlarının aralığının dışına taşabilmektedir. Böyle bir durumla karşılaşılması halinde eşitlenmiş puanların ölçeğin dışına çıkmasına izin verilebileceği gibi aşırı puanları ölçek sınırları içinde tutmak için kırpma yöntemi uygulanabilir. Bu değerler maksimum veya minimum puanlara eşitlenebilir.

**Eşit yüzdellikli eşitleme yöntemi.** Puan dağılımlarının farklılaştığı durumlarda kullanılmaktadır. X formunun uygulanmasından elde edilen puanların Y formunun uygulanmasından elde edilen puanlara dönüştürülmesi ile elde edilen eşitlenmiş puanların yığılmalı dağılımının Y formunun yığılmalı dağılımına eşit olduğunu varsayar. Eşit yüzdellikli

eşitleme iki aşamada gerçekleştirilmektedir. İlk aşamada her bir test formunun yığılmalı frekans dağılımları hesaplanır. İkinci aşamada yığılmalı frekans dağılımlarına göre aynı yüzdelik puanlara karşılık gelen puanlar eşit kabul edilir. Aynı yüzdeliklere denk gelen puanları belirlemek zordur. Bu nedenle ham puanlar sürekli hale getirilir. Ham puanları sürekli hale getirmek için düzgünleştirme veya öteleme yöntemleri kullanılır.

***Madde Tepki Kuramına Dayalı Eşitleme Yöntemleri.*** Madde tepki kuramı bireylerin yetenekleri ile maddeyi doğru cevaplama olasılıkları arasında ilişki olduğu varsayımı üzerine kurulu bir modeldir. Bu nedenle yetenek ve parametre kestirimleri önem taşımaktadır. Eşitlenecek test formlarının uygulandığı grupların yeteneklerinin ölçüklerinin aynı olması gerekmektedir. Tek grup, random grup ve dengelenmiş tek grup desenleri kullanıldığında bireylerin yeteneklerinin aynı ölçekte olduğu kabul edilir. Farklı test formlarından gelen parametre kestirimlerinin orijin ve ölçü birimini aynı değerlerde olduğu varsaymak, parametre kestirimlerini otomatik olarak aynı ölçekte yerleştirir (Lee & Lee, 2016). Denk olmayan gruplarda ortak madde deseni kullanılarak yapılan eşitleme çalışmalarında, farklı test formlarından kestirilen parametreler grupların farklı olmasından dolayı aynı ölçek üzerinde yer almaz. İki test formundan kestirilen parametrelerin aynı ölçekte yer alması için doğrusal bir dönüştürmeye ihtiyaç duyulur (Han, 2008).

Parametrelerin aynı ölçeğe dönüştürülmesinde ayrı, eş zamanlı ve sabit kalibrasyon yöntemleri kullanılmaktadır. Ayrı kalibrasyon yönteminde parametreler ayrı ayrı kestirilerek ortak maddeler üzerinden kalibrasyonlar gerçekleştirilir. Ortalama-ortalama, ortalama-sigma, Haebara ve Stocking Lord ayrı kalibrasyon yöntemleridir. Eş zamanlı kalibrasyon yönteminde ise iki test formundaki ortak maddelerin aynı madde parametrelerine sahip olduğu varsayılarak iki test formuna ilişkin madde parametrelerini birlikte kestirir. Testlerin uygulandığı grupların yetenek dağılımlarındaki farklılıklara göre madde parametreleri kestirildiği için, bu parametreler aynı ölçek üzerinde yer almaktadır. Sabit kalibrasyonda, ortak maddeler için madde parametreleri, yeni form maddeleri kalibre edildiğinde eski formdan kestirilen değerlere sabitlenir (Kim, 2018). Yeni form için gizil değişkenlerin

dağılımını, sabit parametre tahminleri ve yeni formun ortak maddelerinden elde edilen yanıtlarını kullanarak tahmin eder. Böylece, başka bir kalibrasyon adımına gerek kalmadan tahmini dağılımı kullanarak madde parametreleri istenilen ölçekte elde edilir. Kalibrasyonlar gerçekleştirildikten sonra testler eşitlenir. Madde tepki kuramına dayalı yöntemler gerçek ve gözlenen puan eşitleme olmak üzere ikiye ayrılır.

*Gerçek puan eşitleme yöntemi.* Bireylerin gerçek puanları, her maddenin doğru yanıtlanma olasılığının belirlenmesi ve bu olasılıkların toplanması ile elde edilir. Genellikle gerçek puan, alt asimptot parametrelerinin toplamı ile yeni formdaki toplam madde sayısı arasında değerler almaktadır. Gerçek puanlar farklı yetenek düzeylerine göre kestirilerek X formu üzerinde seçilen gerçek puana karşılık gelen yetenek düzeyi belirlenir. Belirlenen bu yetenek düzeyi için Y formunda karşılık gelen gerçek puan hesaplanır (Kolen & Brennan, 2004). Gerçek puan eşitlemede puanlar gruptan bağımsız kestirilmiş olur.

*Gözlenen puan eşitleme yöntemi.* Gözlenen puan eşitleme için, koşullu gözlenen puan dağılımları,  $f(x|\theta_i)$ , özyinelemeli bir formül kullanılarak her yetenek düzeyinde ( $\theta_i$ ) belirlenir (Lord ve Wingersky, 1984). Elde edilen dağılımlar daha sonra yetenek yoğunluğu ( $\psi(\theta)$ ) ile çarpılır ve her test formu için marjinal bir gözlenen puan dağılımını belirlemek için tüm yetenek seviyeleri üzerinden toplanır. Her test formu için marjinal dağılımlar ( $f(x)$  ve  $f(y)$ ) belirlendikten sonra, iki test formu geleneksel eşit yüzdelikli yöntem kullanılarak eşitlenebilir.

**Çok Boyutlu Test Eşitleme (ÇBE).** Çok boyutlu test eşitlemenin (ÇBE) amacı TBE ile aynıdır. Fakat birden fazla gizil değişkeni, yeteneği veya özelliği ölçen testler arasında eşitleme gerçekleştirilir. Özellikle boyut sayısı arttıkça, aynı koordinat sistemi üzerindeki alternatif test formlarının sonuçlarını bağlamak veya aynı koordinat sistemine yerleştirmek için gerçekleştirilir. Tek boyutlu test eşitlemede olduğu gibi ÇBE süreci de eşitleme deseninin seçilmesi ile başlar. Seçilen desene göre test formları uygulanır. Test eşitleme yöntemi seçilir, testler eşitlenir ve sonuçlar değerlendirilir. TBE'ye benzer şekilde, ÇBE için

eşitleme desenleri, dengelenmiş tek grup, random grup ve denk olmayan gruplarda ortak madde desenlerini içerir (Peterson, 2014).

**Çok Boyutlu Kalibrasyon Yöntemleri.** ÇBE'nin yapılabilmesi için ön koşul, madde ve yetenek kestirimlerinin aynı ölçekte olmasıdır. Bunun için ilk olarak ölçek kalibrasyonları yapılmalıdır. ÇBE için ölçek kalibrasyon yöntemleri, madde ve kişi parametre kestirimlerini aynı ölçeğe yerleştirmek için geliştirilmiştir. Kalibrasyon yöntemleri tek boyutlu yöntemlere göre daha karmaşıktır. ÇBE için kalibrasyon doğrusal bir dönüşümdür. Ancak, birden fazla parametre seti olduğundan, matrislerin doğrusal bir dönüşümüdür (Simon, 2008). Kullanılan eşitleme tasarımına bağlı olarak bazı kalibrasyon stratejileri gereklidir.

*Denk olmayan gruplarda ortak madde deseni için yöntemler.* Denk olmayan gruplarda ortak madde deseni için tek boyutlu kalibrasyon yöntemlerine benzerdir. Bu çalışma, iki farklı ÇB-MTK kalibrasyon yöntemini dikkate almıştır: a) Ayrı Kalibrasyon, b) eş zamanlı kalibrasyon. ÇBE çerçevesinde, denk olmayan gruplarda, kalibrasyon her iki test formu için ortak maddeler veya bireyleri içerebilir.

Eş zamanlı kalibrasyonda, X ve Y formları için ayrı test formları tek bir testmiş gibi ele alarak tüm parametreleri aynı anda kestirir. Bunun için parametreler aynı ölçekte, ancak bir simülasyondaki üretici parametrelerle aynı ölçekte olmayabilirler (Simon, 2008). Eşzamanlı kalibrasyon için, iki test formundan oluşan denk olmayan gruplarda ortak madde deseni üç alt testten oluşur: 1) Y (referans/base grup) formundaki benzersiz test maddeleri, 2) X (hedef/odak/target grup) formundaki benzersiz test maddeleri ve 3) hem Y hem de X formunda bulunan ortak maddeler. Her test formunun benzersiz maddeleri diğer test formu için eksik veya ulaşılamamış veri olarak tanımlanır. İki test formundan elde edilen yanıt verileri birleştirilir. İki test formundan gelen yanıt verileri birleştirildikten sonra, ortak maddeler dışındaki maddelerde eksik yanıtlar gözlenir. Eş zamanlı kalibrasyon eksik verilerle gerçekleştirilir. Eksik veri mekanizmanın rastlantısal (MAR) olduğu kabul edilir. Eş zamanlı kalibrasyonda kullanılan prosedürler joint maximum likelihood estimation (JMLE) ve marginal maximum likelihood estimation tahminidir (MMLE). Bu iki yöntemde beklenti

maksimizasyonuna (EM) dayanır. Bu yöntemde, Beklenti ve Maksimizasyon (EM) adımları, ortalama yapı ve varyans/kovaryans yapılarının farkı yaklaşık bir yakınsama eşiğini karşılayana kadar tekrar tekrar hesaplanır. Tek boyutlu Madde tepki Kuramı (TB-MTK) modelleri için kullanılan EM algoritması bi-faktör modeline uygularken, EM algoritmasının E-adımı, her bir grup için önceki EM döngüsünden elde edilen madde yanıtları ve geçici madde parametre kestirimleri kullanılarak üç tür 'pseudo data' oluşturmayı içerir. M-adımı,  $j$  maddesi için parametre kestirimlerini maksimize eden değerleri bulunarak güncellenir. M-adımında tüm  $G$  grupları için gizil değişkenlerin olasılık dağılımları da güncellenir. Bayes tahmini, JMLE veya MMLE yöntemlerinde de kullanılabilir. Örneğin M-3PLM ÇB-MTK eşzamanlı kalibrasyonda, referans ve hedef grupların madde ( $a$ ,  $d$  ve  $c$ ) ve kişi ( $\theta$ ) parametreleri aynı anda kestirilir ve temel grubun yeteneklerin ortalaması sıfır, standart sapması bir olarak sabitlenir (Kim, 2015). Bu nedenle, her iki grubun kişi parametreleri ortak bir ölçekte yer alır ve iki ayrı kestirim süreci gerekli değildir. Hedef grubun yetenek parametreleri, referans gruba göre kestirilir. Ayrıca, hedef grubun ortalama ve varyansının farklı gruplar arasında serbestçe kestirilmesi gerekir. Referans gruptan ortak maddelerin parametre kestirimlerinin sabitlendiği ve yaygın olmayan madde parametrelerinin hedef grup verileri kullanılarak kestirildiği bu prosedürün varyasyonları vardır.

Denk olmayan gruplara uygulanan birden fazla test formu için madde parametrelerini kestirilirken, ayrı kestirimler sonucunda elde edilen parametreleri aynı koordinat sistemine sahip değildir. Çünkü gizil değişkenlerin dağılımı her bir kalibrasyon için çok değişkenli standart normal bir dağılıma keyfi olarak ayarlanır. Ayrı kalibrasyon yöntemleri, rotasyonel belirsizlik (rotation indeterminacy), korelasyon belirsizliği (correlation indeterminacy), orijin (translation indeterminacy) ve ölçü birimindeki belirsizliği (dilation/axis indeterminacy) gidererek yetenek kestirimlerini ve madde parametre kestirimlerini aynı ölçeklere yerleştirebilir. Rotasyonel belirsizlik yetenek ve madde ayırıcılık parametrelerinin dönme belirsizliğidir. Bu belirsizliklerden dolayı, bir test formunun madde parametrelerini diğer test formunun madde parametrelerine dönüştürmek gereklidir. Bu, iki kalibrasyondan

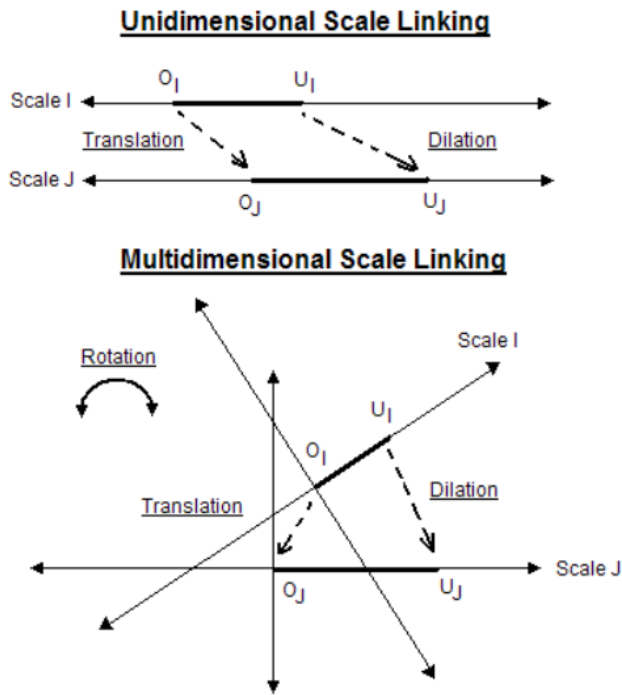
gelen aynı adayların yetenek parametre kestirimlerini mümkün olduğunca benzer hale getirerek gerçekleştirilebilir. Korelasyon belirsizliği ayrı kestirimler sonucunda elde edilen boyutlar arasındaki korelasyonun, aynı aday tasarımı için bazı paralel olmayan maddelerden dolayı farklı kestirilebilmesidir. Korelasyonların farklı olması iki ayrı madde parametresi kestiriminden elde edilen parametre kestirimlerinin, farklı temel vektörlerin bileşimleri veya lineer kombinasyonları olduğunu gösterir. Orijin belirsizliği yetenek parametrelerinin konumları ile ilgilidir. Parametre kestirimlerinde ölçek (orijin ve ölçü birimindeki belirsizlik) belirsizliğinin giderilmesi için, her boyut için yetenek parametresi kestirimleri, her parametre kestirim adımından sonra sıfır ortalama ve birim varyansa standartlaştırılır (Li & Lissitz, 2000). Orijin belirsizliği giderildiğinde a parametreleri değişmez. Fakat yetenek ve d parametreleri değişebilir.

Tek boyutlu test eşitlemeden farklı olarak orijin ve ölçü birimindeki farklılıkları hesaba katmak için iki ölçekleme sabiti A ve B yerine dönme/rotasyon, korelasyon, orijin ve ölçü birimindeki farklılıkları hesaba katmak için skalerler ve matrisler gerekir (Peterson, 2014). Olabilirlik hesaplamalarında boyutun azaltılmasını sağlayan, tüm boyutların tam bağımsızlığına ilişkin standart bi-faktör analiz varsayımı, bi-faktör model için rotasyon belirsizliğini çözer (Kim, 2018). Tam bağımsızlık varsayımını genel faktöre göre özel faktörlere koşullu bağımsızlıkla gevşeterek de elde edilebilir. Bu durumda, rotasyon belirsizliği, genel faktör ile her bir özel faktör arasındaki korelasyonun sıfır olacak şekilde kısıtlanmasıyla düzeltilebilir.



## Şekil 1

### Çok Boyutlu Ölçek Kalibrasyonu



ÇBE için kullanılan yöntemle göre kalibrasyon formülleri değişse de genel formüller  $m$  testin boyut sayısı olmak üzere  $A$  rotasyon ve ölçü birimindeki belirsizliği hesaba katan ve boyutlar ( $Q$ ) için varyans ve kovaryansları düzenleyen  $m \times m$  boyutunda bir eğik veya ortogonal matris (döndürme matrisi/rotation matrix),  $\beta$  orijin belirsizliğini hesaba katan  $m \times 1$  boyutunda bir vektör (öteleme vektörü/translation vector),  $X$  hedef ölçeği ve  $Y$  referans ölçek olmak üzere aşağıdaki gibidir (Davey vd., 2000);

$$Q_{Yi} = (AQ_{Xi}) + \beta$$

$$\alpha_{Yi}^T = (A^{-1})^T a_{Xi}$$

$$d_{Yi} = d_{Xi} - \alpha_{Xi}^T A^{-1}\beta$$

$$c_{Yi} = c_{Xi}$$

Ayrı kalibrasyon yöntemleri için Hirsch (1989) iki test formunu ortak alan bireyler üzerinden ölçek kalibrasyonu yapılmasını önermiştir. Yöntem dört adımda gerçekleştirilir. İlk adımda iki test formu içinde parametre kestirimleri gerçekleştirilir. İkinci adımda yetenek

parametrelerinin koordinatları arasındaki korelasyonları eşleştirilir. Üçüncü adımda ortogonal Procrustes rotasyon matrisi bulunur. Dördüncü adımda iki test formundaki ortak sınava giren bireyler için yetenek parametrelerinin koordinatlarının ortalamaları ve standart sapmaları belirlenir. Ardından ortalama/sigma yöntemi her boyuta uygulanır. Bu yöntemde rotasyon belirsizliğini gidermek için ortogonal rotasyon matrisi kullanılmaktadır. Her boyutta ortak sınava girenler için ortalamalar ve standart sapmalar orijin ve ölçü birimi belirsizliklerini gidermek için katsayıları hesaplamada kullanılır. Elde edilen katsayılar, her boyuttaki test formları arasında ortak maddeler için ortalamaları ve standart sapmaları eşitlemede kullanılır.

Davey vd. (2000) ise TBE için kullanılan kalibrasyon yöntemlerinin uzantısı olan dört yöntem önermektedir. İlk yöntem formlar arasında ortak madde parametre kestirimleri arasındaki kare farkını en aza indiren Divgi (1985) yönteminin bir uzantısıdır (The Direct Method). İkinci yöntem "Eşitlenmiş Fonksiyon Yöntemi" olarak isimlendirilen TBE için kullanılan ortalama-ortalama yönteminin bir uzantısıdır (The Equated/Matching Scale Function Method). Üçüncü yöntem, Stocking ve Lord'un (1983) test karakteristik eğri yönteminin bir uzantısı olup "Eşleşen Test Yanıt Fonksiyonları veya Yüzeyleri" olarak isimlendirilmektedir (Matching Test Responce Function/Surface). Çok boyutlu test karakteristik yüzeyleri arasındaki farkı en aza indirecek dönüşüm matrisleri elde edilir. Dördüncü yöntem ise Haebara madde karakteristik eğri yönteminin çok boyutlu uzantısıdır (The Item Characteristic Function Method). Çok boyutlu madde karakteristik yüzeyleri arasındaki farkı en aza indirecek dönüşüm matrisleri elde edilir. Yöntemlerin hepsi rotasyon (A) ve orijin ( $\beta$ ) matrislerini kullanmaktadır. Boyutlar arasındaki korelasyon iki form içinde aynı ise rotasyon matrisi ortogondur; aksi takdirde eğik rotasyon matrisi kullanılır (Simon, 2008). Yöntemlerin performansları karşılaştırıldığında Stocking-Lord ve ortalama/ortalama yöntemlerinin madde ayırıcılık parametreleri için daha iyi kestirimler sağladığı, Stocking-Lord ve ortalama/sigma yöntemlerinin ise madde güçlük parametresi için daha iyi kestirimler sağladığı sonucuna varılmıştır (Atar & Yeşiltaş, 2017).

Li ve Lissitz (2000), madde parametrelerinin diğ er bir öl çeğ e dönü ştürölmesi, ortogonal Procrustes dönü şümü, bir çeviri vektörü ve tek bir genişleme parametresini içerir. Li ve Lissitz (2000), Oshima vd. (2000) rotasyon matrisinin ortogonal olmadığını, dolayısıyla döndürme matrisinin birden fazla formu olabileceğini (rotasyon belirsizliğı) öne sürmektedir. k genişletme parametresi (dilation parameter), T ortogonal matris (döndürme matrisi/rotation matrix),  $\beta$  vektörü (öteleme vektörü/translation vector) olmak üzere kalibrasyon formülleri aşağıdaki gibidir (Li & Lissitz,2000);

$$a_{jY} = k a_{jX}^T T$$

$$Q_{jY} = \frac{1}{k} (T^{-1} Q_{jX} - \beta)$$

$$d_{jY} = d_{jX} + (a_{jX}^T T)\beta$$

T ortogonal matrisi Oshima vd.'nin (2000) önerdiği kalibrasyon formüllerinde yer alan A matrisi ile benzerdir.  $\beta$  vektörleri iki yöntem içinde aynıdır. Oshima vd. 'nin (2000) önerdiği rotasyon matrisi A, yetenek ve madde ayırıcılık parametrelerini aynı anda döndürür ve genişletir. Li ve Lissitz'in (2000) önerdiği T ortogonal matrisi ise sadece döndürme yapar. Bu nedenle genişletme parametresi k formüle eklenmiştir. T ortogonal matrisi ve ölçekleme katsayıları (k genişletme parametresi ve  $\beta$  vektörü) ayrı ayrı kestirilir. T ortogonal matrisi formların ayırıcılık parametrelerinin kareleri arasındaki farkın toplamını minimize edecek şekilde ortogonal Procrustes rotasyonu (Schönemann, 1966) kullanılarak kestirilir.  $E_1$  artıklar matrisi,  $A_Y$  Y formu ve  $A_X$  X formu ayırıcılık parametreleri matrisi olmak üzere T ortogonal matrisi aşağıdaki formül ile belirlenir (Li & Lissitz, 2000);

$$E_1 = A_X T - A_Y$$

Ölçekleme katsayılarının kestirimi için Li ve Lissitz (2000) üç farklı yöntem önermişlerdir. 1) Ölçekleme katsayılarının eşzamanlı kestirimi için test karakteristik yüzeyini eşleştirme (Matching the Test Response Surface); 2)  $\beta$  vektörü için k parametresini kestirmek için özdeğerlerin oranı ile birleştirilmiş en küçük kareler tahmini ve 3)  $\beta$  vektörü için k parametresini kestirmek için izin oranı ile birleştirilmiş en küçük kareler tahmini.

Ölçekleme katsayılarının eşzamanlı kestirimi için test karakteristik yüzeyini eşleştirme (MTRS) yöntemi ortak maddelerin doğru cevap olasılıklarının toplamından oluşan test karakteristik yüzeyidir (TRY), bu da ortak maddelerin beklenen doğru sayısı olarak adlandırılır (BDS). Ölçekleme katsayıları, referans ve odak formları için beklenen doğru sayısındaki farkı minimize edecek şekilde elde edilir. MTRS yöntemi, ÇB-MTK modelleri için Stocking Lord (1983) prosedürünün uzantısıdır. Bu araştırmada MTRS yöntemi kullanılmıştır.

*Dengelenmiş tek grup ve random grup deseni için yöntemler.* Dengelenmiş tek grup ve random grup deseninde gruplar rastgele oluşturuldukları için m boyutlu test uzayında orijinlerinin ve ölçü birimlerinin aynı olduğu varsayılır. Böylece orijin ve ölçü birimlerindeki belirsizlik giderilmiş olur. Fakat rotasyon ve korelasyon belirsizlikleri devam etmektedir. Korelasyon belirsizliği sorunu genellikle, gizil yeteneklerin çok değişkenli bir normal dağılım izlediğini ve karşılıklı olarak birbirine dik olduğunu belirterek çözülür. Her iki form da farklı koordinat sistemlerinde olsa da sonuç olarak ortaya çıkan marjinal gözlenen puan dağılımı, çok boyutlu madde tepki kuramı modellerinin değişmezlik özelliğinden dolayı koordinat sistemi seçiminden etkilenmez. Böylece rotasyon belirsizliği sorunu çözülür. Brossman ve Lee'de (2013) yaptıkları çalışmada, gözlenen puan tam ÇBE yapmak için random grup deseni altında ölçek kalibrasyonunun gerekli olmadığını bulmuştur. Bunun dışında hiçbir koşulda random grup deseni kullanmak, ÇBE için aynı ölçekte parametre kestirimlerini sağlamaz. Thompson vd. (1997) rotasyon belirsizliğini gidermek için ortogonal rotasyon matrisini (T) kestirim yöntemi önermektedirler.

Kalibrasyon yöntemleri arasındaki farklılıklar dönme belirsizliğini gidermek için parametre kestirimlerini döndürmek için bir ortogonal matris veya ortogonal olmayan (eğik) bir matrisin kullanılıp kullanılmadığı, ölçü birimi belirsizliğini hesaba katmak için ölçü birimi parametrelerinin bağlantı prosedürüne dahil edilip edilmediği, ölçü birimi parametrelerini içeren tasarımlar için, tüm boyutlara bir ölçü birimi parametresinin uygulanıp uygulanmadığı veya her boyutu ayrı ayrı hesaba katmak için ayrı bir ölçü birimi parametresinin kestirilip

edilmediği ve dönüşüm matrisleri ve katsayılarının aynı anda mı yoksa ayrı ayrı mı kestirildiğidir. Çok boyutlu ölçek kalibrasyon yöntemlerinin performansının iki şeye bağlı olduğu düşünülmektedir: (a) yeni ve eski test formlarının boyutlarının ne kadar doğru belirlenebildiği; (b) bu boyutların formlar arasında ne kadar doğru bir şekilde eşleştirilebileceği (Peterson, 2014). Kovaryans matrisleri sınava giren gruplarda aynı olduğunda kalibrasyona gerek yoktur (Simon, 2008).

**Çok Boyutlu Test Eşitleme Yöntemleri.** Çok boyutlu testlerin kullanımının artması ile ÇBE yöntemleri geliştirilmiştir. Bu yöntemler oldukça yeni olup hala geliştirilmeye devam edilen yöntemlerde (Kim & Lee, 2022) bulunmaktadır. Alan yazın incelendiğinde basit yapılı ÇB-MTK gözlenen puan (SMO) (Lee & Brossman, 2012), basit yapılı ÇB-MTK gerçek puan eşitleme (SMT) (Kim vd., 2019), tam ÇB-MTK gözlenen puan eşitleme (MOSE), ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım (AOSE), ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşım (ATSE) (Brossman, 2010), bi-faktör ÇB-MTK gözlenen puan (Lee & Lee, 2016), bi-faktör ÇB-MTK gerçek puan (Lee vd., 2015), madde takımı yanıt modeli ÇB-MTK gözlenen puan ve madde takımı yanıt modeli ÇB-MTK gerçek puan (Tao & Cao, 2016) yöntemlerinin çalışmalarda kullanıldığı belirlenmiştir.

*Basit yapılı ÇB-MTK gözlenen puan yöntemi.* Lee ve Brossman (2012) karışık formatlı testler için basit yapılı bir ÇB-MTK (BY-ÇBMTK) çerçevesi altında gözlenen puan eşitleme yöntemini önermişlerdir. Yöntem tek boyutlu gözlenen puan eşitlemeye benzemektedir. İlk olarak madde parametreleri her boyut için ayrı ayrı dönüştürülür. İkinci adımda her formda her boyut için koşullu gözlenen puan dağılımını hesaplanır. Her form için koşullu toplam puan dağılımını hesaplanır. Koşullu toplam puan dağılımı, boyutların bir fonksiyonudur; her boyuttaki yetenek düzeylerinin bir kombinasyonuna bağlı olarak puanlar üretir. Koşullu toplam puan dağılımı  $f_1(x_1|Q_1)$  birinci boyut için koşullu dağılımı ve  $w_1$  birinci boyutun ağırlığı olmak üzere koşullu yerel madde bağımsızlığı varsayımı altında aşağıdaki gibidir;

$$f(x|Q_1, Q_2) = \sum_{X=w_1x_1+w_2x_2} f_1(x_1|Q_1)f_2(x_2|Q_2)$$

Marjinal toplam puan dağılımını elde etmek için koşullu toplam puan dağılımları çok değişkenli bir yetenek dağılımı  $g(Q_1, Q_2)$  üzerinden toplamaktır. Marjinal toplam puan dağılımı aşağıdaki gibidir;

$$f(x) = \sum_{Q_1} \sum_{Q_2} f(x|Q_1, Q_2)g(Q_1, Q_2)$$

Marjinal toplam puan dağılımları elde edildikten sonra iki dağılımı eşitlemek için geleneksel eşit yüzdelikli eşitleme uygulanır.

*Tam ÇB-MTK gözlenen puan eşitleme.* İlk tam çok boyutlu eşitleme yöntemleri Brossman (2010) tarafından geliştirilmiştir. Yöntemler ikili puanlanan testler içindir. Tam ÇB-MTK gözlenen puan eşitleme TB-MTK gözlenen puan eşitleme yöntemine benzer şekilde uygulanmaktadır. Fakat tek bir yetenek seviyesi yerine yetenek seviyeleri vektörü kullanan Lord-Wingersky algoritması ile her yetenek seviyesi kombinasyonu için koşullu gözlenen puan dağılımlarını kullanır. Koşullu gözlenen puan dağılımı, m boyut sayısı,  $\theta$  boyut vektörü ve r madde indeksi olmak üzere aşağıdaki gibidir (Peterson, 2014);

$$\begin{aligned} f_r(x|\theta_j) &= f_{r-1}(x|\theta_j)(1 - P_r), & x = 0 \\ &= f_{r-1}(x|\theta_j)(1 - P_r) + f_{r-1}(x - 1|\theta_j)P_r, & 0 < x < r, \\ &= f_{r-1}(x - 1|\theta_j)P_r, & x = r. \end{aligned}$$

Algoritma, toplam puan olasılıklarını birinci maddeden başlayarak hesaplar. Birinci madde için toplam puanın sıfır veya bir olma olasılığı bulunmaktadır. Daha sonra ikinci madde eklenir. Toplam puan kombinasyonlarının olasılıkları toplanarak toplam puan olasılığı elde edilir. Tüm maddeler eklenene kadar özyineleme tekrarlanır. Bu koşullu dağılımların yetenek yoğunluğu ile çarpılması ve ardından tüm yetenek seviyelerinde toplanması yoluyla her form için marjinal bir gözlenen puan dağılımı elde edilir. TBE'den farklı olarak marjinal gözlenen puan dağılımlarını elde etmek için kullanılan yetenek yoğunluğu, tek değişkenli

değil çok değişkenlidir (Brossman & Lee, 2013). Marjinal gözlenen puan dağılımı çok değişkenli yetenek yoğunluğu ( $\psi(\theta)$ ) olmak üzere aşağıdaki gibidir (Peterson, 2014);

$$f(x) = \int_{\theta_1} \int_{\theta_2} \dots \int_{\theta_m} f(x|\theta)\psi(\theta) d(\theta).$$

Sürekli veriler için integral yerine toplam kullanılır. Sürekli veriler için marjinal gözlenen puan dağılımı aşağıdaki gibidir (Peterson, 2014);

$$f(x) = \sum_{\theta_1} \sum_{\theta_2} \dots \sum_{\theta_m} f(x|\theta)\psi(\theta).$$

Marjinal gözlenen puan dağılımları elde edildikten sonra formlar eşitlenir. İki formu eşitlemek için MTK eşit yüzdelli eşitleme kullanılır.

*ÇB-MTK tek boyutlu yaklaşımlar.* Tek boyutlu yaklaşımlar Zhang'in (1996) çalışmalarının devamı niteliğindedir. Alan yazında birden fazla boyuttan çeşitli yöntemler ile tek boyutlu yapının elde edilebileceği ve tek boyutlu gözlenen ve gerçek puan eşitleme yöntemleri kullanılabileceği belirtilmiştir (Zhang & Stout, 1999). Çok boyutlu yapıyı tek boyuta indirmek için ilk olarak çok boyutlu yeteneklerin lineer bir bileşimi olan tek boyutlu bir bileşik yetenek ve tek boyutlu bileşik yeteneğe karşılık gelen tek boyutlu madde parametreleri elde edilir. Çok boyutlu 2 parametrelili lojistik modele karşılık gelen tek boyutlu 2 parametrelili lojistik model madde parametrelerini kestirim yöntemi aşağıdaki adımlara göre uygulanır;

1.Çok boyutlu madde parametreleri kestirilir

2.İki form içinde en çok bilgi veren yetenek düzeyleri belirlenir. En çok bilgi veren test seviyesi yönüne karşılık gelen ağırlık vektörü k boyut indeksi  $\hat{a}_{ik}$  i maddesi için madde ayırt ediciliği, m toplam boyut sayısı,  $\hat{w}_k$  k. boyutun doğrusal bileşik katsayılar vektöründe standartlaştırılmış katsayı, N testte yer alan toplam madde sayısı olmak üzere aşağıdaki formül yardımıyla kestirilir (Lee, 2013);

$$\hat{w}_k = \frac{\sum_{i=1}^N \hat{a}_{ik}}{\sqrt{\sum_{k=1}^m (\sum_{i=1}^n \hat{a}_{ik})^2}}$$

Ağırlık vektöründe yer alan elamanların karelerinin toplamı bire eşittir.

3.Çok boyutlu madde ve yetenek parametrelerine karşılık gelen tek boyutlu parametreler kestirilir. Madde ve yetenek parametreleri  $\hat{\Sigma}$  çok boyutlu yetenek kovaryans matrisi,  $\delta_{iw}$  koşullu  $\hat{a}_i^T$  madde ayırıcılık matrisi,  $\hat{w}$  ağırlık katsayısı,  $\hat{\sigma}_{iw}^2$  varyans,  $\underline{\theta}$  gizil vektör  $\theta_w$ 'nin bir bileşimi ve  $\hat{d}_{iw}$  çok boyutlu zorluk parametresi ile ilgili bir indeks olmak üzere aşağıdaki gibidir (Lee, 2013; Brossman & Lee, 2013);

$$\hat{a}_{iw} = (1 + \hat{\sigma}_{iw}^2)^{-\frac{1}{2}} \hat{a}_i^T \hat{\Sigma} \hat{w}$$

$$\hat{b}_{iw} = - \frac{\hat{d}_{iw}}{\hat{a}_{iw}}$$

$$\hat{d}_{iw} = (1 + \hat{\sigma}_{iw}^2)^{-\frac{1}{2}} \cdot \hat{d}_i$$

$$\hat{c}_{iw} = \hat{c}_i$$

$$\hat{\sigma}_{iw}^2 = \hat{a}_i^T \hat{\Sigma} \hat{a}_i - (\hat{a}_i^T \hat{\Sigma} \hat{w})^2$$

$$\underline{\theta} = \hat{w}_k^T \theta_w$$

Elde edilen parametreler normal ogive model parametreleri olup lojistik modellerde kullanılması tüm yetenek uzayı boyunca 0.01'den az farklılıklara neden olabilmektedir.

Yukarıda bahsedilen parametre kestirimleri gerçekleştirildikten sonra ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşımı kullanılabilir. Yöntemin adımları TB-MTK gerçek puan eşitleme ile aynı şekilde devam eder. İlk olarak karşılığı bulunmak istenen gerçek puan belirlenir. Form Y'deki belirli gerçek puanla ilişkili karşılık gelen tek boyutlu bileşik yetenek ( $\theta_w$ ) düzeyi aşağıdaki formül ile kestirilir (Lee, 2013);



$$f(\theta_{jw}) = \tau_{Xw} - \sum_{i:X} P_{ij}(\theta_{jw}; a_{iw}, b_{iw})$$

Yetenek düzeyi belirlendikten sonra diğer formda bu yetenek düzeyine karşılık gelen puan bulunur. ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşımı için, bileşik yetenek düzeyiyle ilişkili bileşik gerçek puan, her bileşik yetenek düzeyindeki tüm maddeler üzerinde doğru yanıtlar elde etme olasılıklarının toplamıdır. Bileşik gerçek puan aşağıdaki formül ile bulunur (Lee, 2013);

$$\tau_{Xw}(\theta_{jw}) = \sum_{i:X} P_{ij}(\theta_{jw}; a_{iw}, b_{iw})$$

ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşımı, TB-MTK gerçek puan eşitleme yönteminden daha karmaşıktır. Bunun nedeni, test karakteristik eğrisinin bir test karakteristik yüzeyi ile değiştirilmesi ve gerçek puan ile gizil yetenek arasında bire bir eşleşme yerine bire çok eşleşme kombinasyonunun olmasıdır.

ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşımı parametre kestirimleri gerçekleştirildikten sonra TB-MTK gözlenen puan yönteminin aynısı olarak gerçekleştirilir. Gözlenen puan yöntemi yerine tam ÇB-MTK gözlenen puan yöntemi de tercih edilebilir. Koşullu gözlenen puan dağılımı her bir bileşik yetenek seviyesinde Lord-Wingersky formülü ile belirlenir. Her form için gözlenen marjinal dağılımı belirlemek için, koşullu dağılım popülasyondan kestirilen tek boyutlu yetenek dağılımı (yetenek yoğunluğu) ile çarpılır ve kestirilen tek boyutlu yetenek uzayı boyunca toplanır. Bu çalışmada yetenek yoğunluğu normal dağılım olarak ele alınmıştır. Marjinal gözlenen puan dağılımı aşağıdaki gibidir (Lee, 2013);

$$f(x | \theta_w) = \sum_{\theta_{jw}} f(x | \theta_{jw}) \psi(\theta_{jw}),$$

$$f(x | \theta_w) = \int_{\theta_w} f(x | \theta_{jw}) \psi(\theta_{jw}) d\theta_{jw}.$$

Her form için gözlenen marjinal dağılımı belirlendikten sonra iki formu eşitlemek için TB-MTK eşit yüzdelikli eşitleme kullanılır.

*Bi-faktör ÇB-MTK gözlenen puan eşitleme.* Bi-faktör ÇB-MTK gözlenen puan eşitleme yöntemi TB-MTK gözlenen puan eşitleme yöntemine benzemektedir. Örneğin koşullu doğru sayısı tahmin edilir ve marjinal puan dağılımları bu puanlar toplanarak elde edilir. Daha sonra eşit yüzdelikli eşitleme yapılır. TB-MTK gözlenen puan eşitlemede yetenekler tek boyutlu iken Bi-faktör gözlenen puan eşitlemede genel ve gruba özel yetenek üzerine koşullanır. Parametre kestirimleri bi-faktör modele göre gerçekleştirilebilir.

Bi-faktör ÇB-MTK gözlenen puan eşitleme üç adımda gerçekleştirilir:

- 1.Koşullu gözlenen puan dağılımlarını oluşturulması
- 2.Koşullu dağılımların marjinal gözlenen puan dağılımlarına toplanması
- 3.Eşit yüzdelikli eşitleme yapılması

Koşullu gözlenen puan dağılımları ikili puanlanan maddeler için Lord-Wingersky ve çoklu puanlanan maddeler için Hanson (1994) formülü ile belirlenir (Lee & Lee, 2014).

Lord-Wingersky algoritması, doğrudan quadrat nokta kuralları kullanılarak standart çok boyutlu MTK modellerine genişletilebilse bile, modele daha fazla faktör eklendikçe hesaplama karmaşıklığı katlanarak artar. Örneği bi-faktör model için koşullu gözlenen puan dağılımları elde edilirken genel faktör eklenir (Choi, 2019). Bu nedenle, analitik olarak boyut azaltma sağlamak için hiyerarşik madde faktör analizi modelinin belirttiği kısıtlamalardan yararlanan farklı bir algoritmaya ihtiyaç duyulmuştur. Cai (2015) Lord-Wingersky algoritması

ile analitik boyut azaltma algoritmasını birleştirerek Lord-Wingersky 2.0 algoritmasını geliştirmiştir. Bu araştırmada koşullu gözlenen puan hesaplanırken Lord-Wingersky 2.0 versiyonu kullanılmıştır. Marjinal gözlenen puan dağılımları  $m$  boyut sayısı olmak üzere  $g$   $m$  boyutlu yetenek ( $\theta$ ) düzeyi matrisi olmak üzere şu şekilde elde edilir (Sansivieri vd. 2018);

$$f(x) = \int \int \int f(x | \theta_0, \theta_1, \theta_2) g(\theta_0, \theta_1, \theta_2) d\theta_0 d\theta_1 d\theta_2,$$

Bu adım Y formu içinde  $g(y)$  marjinal dağılımını bulmak için tekrar edilir. Marjinal dağılımlar bulunduğundan sonra  $e_y(x)$  X formundan elde edilen  $x$  puanın Y formuna eşitlenmiş puanı,  $F$  X formunun yığılmalı dağılım fonksiyonu ve  $G^{-1}$  Y formunun yığılmalı dağılım fonksiyonu  $G$ 'nin tersi olmak üzere aşağıdaki formül ile eşit yüzdelliğe eşitleme gerçekleştirilir (Sansivieri vd. 2018);

$$e_Y(x) = G^{-1}[F(x)]$$

### **Eşitleme Hatası**

Eşitleme sonucunda elde edilen puanlar kestirim hataları içermektedir. Kestirim hatalarının en az olacağı şekilde eşitleme yapılması gerekmektedir. Test eşitlemede hata random ve sistematik hata olmak üzere ikiye ayrılır. Random hata örneklemden kaynaklanan hatalardır. Eşitlemenin standart hatası (SEE) olarak isimlendirilir. Eşitlemenin tekrarlanması sonucunda elde edilen hataların standart sapmasıdır. Eşitlemenin standart hatasının sıfır olması beklenir. Örneklemin evreni temsil etme düzeyi arttıkça standart hata azalır. Eşitlemenin standart hatası (SEE) genellikle eşitleme çalışmasının yapıldığı örneklem büyüklüğünün yetersiz olduğu durumlarda oluşur ve kişi sayısı artırılarak hata oranı düşürülebilir (Felan, 2002) Sistematik hata ise kullanılan yöntemlerden kaynaklanan yanlılıktır. Eşitlemenin yanlılığı (BIAS) olarak isimlendirilir. Eşitleme varsayımlarının ve koşullarının sağlanmamasından kaynaklanır.

## İlgili Araştırmalar

Madde tepki kuramı çerçevesinde test eşitleme yöntemlerini incelemek için yapılan çalışma sayısı her geçen gün artmaktadır. Hirsch (1989) iki boyutlu iki paralel formu alt test düzeyinde ortak birey deseni altında eşitlemiştir. Çalışmada gerçek ve simülasyon veri setleri kullanılmıştır. Gerçek veri seti 1984 yılında uygulanan Florida CLAST testinden elde edilmiştir. 2000 kişi ve 71 maddeden oluşmaktadır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü 2000 ve madde sayısı 40 olmak üzere sabit tutulmuş olup ortak birey sayısı (100, 200, ve 300) ve sınava giren grupların ortalama yeteneklerindeki fark manipüle edilen faktörler olarak ele alınmıştır. Eşitlemenin doğruluğu, ortak sınava girenler için eşitlenmiş gerçek puanların ve yetenek kestirimleri arasındaki farkların karşılaştırılması ile değerlendirmiştir. Çalışmanın sonuçlarına göre gerçek puanların eşitleme için kullanılabileceği belirlenmiştir. Yetenek kestirimlerinin kararlılığı incelendiğinde, tatmin edici olmayan sonuçlar bulunmuştur.

Camilli, Wang ve Fesq (1995) çok boyutluluğun test eşitlemeye etkisini incelemiştir. Çalışmada gerçek veri seti kullanılmıştır. Gerçek veri seti 1989 ve 1990 yılları arasında uygulanan Hukuk Fakültesi kabul testinden elde edilmiştir. İçerik alanları (Akıl Yürütme (AR), Okuduğunu Anlama (RC) ve Mantıksal Akıl Yürütme) boyutluluk kaynağı olarak ele alınmıştır. Veriler incelendiğinde gruplar arasında yetenek dağılımlarında farklılıklar olduğu belirlenmiştir. Çalışmanın sonuçlarına göre tek boyutluluk ihlallerinin gerçek puan eşitleme üzerinde önemli bir etkisi olmadığı belirlenmiştir.

ÇBE yöntemlerinin geliştirilmesi ve performanslarının incelenmesi için yapılan ilk çalışmalarda Zhang (1996) çok boyutlu bir MTK modeli tarafından yeterince modellenebilen herhangi bir madde yanıt seti için kestirilen tek boyutlu yetenek ve madde parametrelerinin tek boyutlu bir MTK modeli ile kestirilebildiğini göstermiştir. Çalışmada kestirilen tek boyutlu yetenek, gizil değişkenlerin standartlaştırılmış doğrusal bir bileşeni olarak tanımlamıştır. ÇB-MTK gerçek ve gözlenen puan eşitleme tek boyutlu yaklaşımının temellerini atmıştır.

Spence (1996) çok boyutluluğun TBE yöntemlerine etkisini random grup ve denk olmayan gruplarda ortak madde deseni altında incelemiştir. Çalışmada simülasyon verileri kullanılmıştır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü 1000 ve madde sayısı 40 olmak üzere sabit tutulmuş olup çok boyutlu madde sayısı (10, 20, 30 ve 40), kullanılan ÇB-MTK modeli (telafi edici ve telafi edici olmayan model), kalibrasyon yöntemi (eş zamanlı, moment ve karakteristik eğri kalibrasyon yöntemleri) ve grupların yetenek dağılımı (ortalama ((0,0), (-0,8,0)), standart sapma ((1,1), (0,6,1)) manipüle edilen faktörler olarak ele alınmıştır. Eşitlemenin doğruluğu korelasyon, standartlaştırılmış ortalama fark (SDM) ve standartlaştırılmış kök ortalama kare farkı (SRMSD) ile değerlendirilmiştir. Çalışmanın sonuçlarına göre random grup deseni altında kullanılan ÇB-MTK modelinin eşitleme sonuçlarına etkisinin az olduğu görülmüştür. Telafi edici ve telafi edici olmayan modellerin kullanıldığı testler için benzer eşitleme sonuçları elde edilmiştir. SDM ve SRMSD değerleri tüm kalibrasyon yöntemleri için benzer bulunmuştur. Moment ve karakteristik eğri kalibrasyon yöntemleri için tüm eşitleme sonuçları hemen hemen aynıdır. Kalibrasyon yöntemi ve çok boyutlu madde sayısının random grup deseni altında eşitleme sonuçları üzerinde etkisi olmadığı; denk olmayan gruplarda ortak madde deseni altında eşitleme sonuçlarına etkisinin olduğu belirlenmiştir. Eş zamanlı kalibrasyon yöntemi en çok çok boyutlu maddelerin ve farklı sınava giren grupların etkileşiminden etkilenmiştir.

Bolt (1999) çok boyutluluğun TB-MTK gerçek puan eşitlemeye etkisini incelemiştir. Çalışmada simülasyon verileri kullanılmıştır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü 2000 ve madde sayısı 50 olmak üzere sabit tutulmuş olup boyutlar arasındaki ilişki düzeyi ( $\rho=0,3, 0,5, 0,7$  ve 1) manipüle edilen faktör olarak ele alınmıştır. Çalışmanın sonuçlarına göre boyutlar arasındaki korelasyon arttıkça TB-MTK gerçek puan eşitleme yöntemi daha iyi sonuçlar vermektedir.

Bèguin, Hanson ve Glas (2000) çok boyutlu verilere uygulanan tek boyutlu üç parametrelili lojistik (3PLM) model parametrelerinin eş zamanlı ve ayrı kalibrasyon performansını denk olmayan gruplarda ortak madde deseni altında karşılaştırmıştır.

Çalışmada 1992 ve 1995 yıllarında Hollanda'da ortaöğretimin sonunda uygulanan İngilizce sınavlardan elde edilen veriler üzerinden elde edilen madde ve yetenek parametreleri ile üretilen simülasyon verileri kullanılmıştır. Veri seti 50 maddeden oluşmaktadır. Her iki formdan random olarak seçilen 10'ar madde ortak madde olarak ele alınmıştır. Yöntemlerin performansları değerlendirilirken iki kriter kullanılmıştır. İlk kriter Form B'deki tahmini ve gerçek puan dağılımı arasındaki farklardır ve ortalama hata karesi (MSE) ile belirlenmiştir. İkinci kriter ise eşitleme sonucundan elde edilen puanların veriler üretilirken kullanılan modeldeki eşdeğerlerinden farkıdır ve ağırlıklı ortalama karesel hata (WMSE) ile belirlenmiştir. Çalışmanın sonuçlarına göre eşdeğer gruplarda Stocking-Lord yöntemi kullanılarak ölçeklendirmenin uygulandığı ayrı kalibrasyon yöntemi için genellikle tek boyutlu eşzamanlı kestirim yönteminden daha iyi kriter değerleri gözlenmiştir. Ölçekleme yapılmadan elde edilen eşitleme sonuçları ayrı kalibrasyon ile benzer sonuçlar vermiştir. En iyi sonuçlar eş kalibrasyon yönteminde gözlenmiştir. Tek boyutlu kestirim yöntemlerinden elde edilen kriterler, çok boyutlu yöntemler kullanılarak elde edilen kriterlere oldukça benzerdir. Eş değer olmayan gruplarda Tek boyutlu kestirim yöntemlerinden elde edilen hata değeri çok boyutlu yöntemlere göre daha fazladır. İkinci boyutunun kovaryans ve varyansının artmasıyla hata değeri artmaktadır. Kovaryans değeri 0,7 ve üstü için ayrı kalibrasyon daha iyi sonuçlar vermiştir. Boyutluluğun kalibrasyon yöntemlerini etkilediği belirlenmiştir.

Zhang (2012) ÇB-MTK eşitleme yöntemlerini performanslarını denk olmayan gruplarda ortak madde deseni altında incelenmiştir. Çalışmada simülasyon verileri kullanılmıştır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü 2000 ve madde sayısı 40 olmak üzere sabit tutulmuş olup yetenek dağılımı (gruplar arasında fark yok, varyansları farklı, ortalamaları farklı ve korelasyonları farklı), test yapısı (yaklaşık basit yapı (YBY) ve karmaşık yapı (KY)), kalibrasyon yöntemi (eş zamanlı, madde ve test karakteristik fonksiyonuna dayalı (MKF ve TKF), Min'in yöntemi (M) ve ortogonal olmayan prokrustes (NOP) kalibrasyon) ve eşitleme yöntemi (tam ÇB-MTK gözlenen puan eşitleme, ÇB-MTK

gözlenen ve gerçek puan eşitleme tek boyutlu yaklaşım yöntemleri) tek boyutlu yaklaşım manipüle edilen faktörler olarak ele alınmıştır. Eşitlemenin doğruluğu ağırlıklı ortalama yanlılık ve ağırlıklı ortalama kök ortalama kare sapma ( $ARMSD_W$ ) ile değerlendirilmiştir. Çalışmanın sonuçlarına göre test yapısı ve test yapısı dahil tüm etkileşimlerin eşitleme sonuçlarında çok küçük bir etkiye sahip olduğu belirlenmiştir. Üç yetenek dağılım faktörü (ortalaması, korelasyon ve standart sapma) arasında, ortalama faktörü eşitleme sonuçlarını en çok etkilemiştir. Yetenek dağılım değişikliğinin etkileşimi ve ÇB-MTK kalibrasyon yöntemlerinin eşitleme sonuçları üzerinde büyük bir etkiye sahiptir. Yetenek dağılım değişikliğinin etkileşimi ve ÇB-MTK eşitleme yöntemlerinin türü de eşitleme sonuçları üzerinde büyük bir etkiye sahiptir. ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşım yöntemi, tüm yetenek dağılım koşullarında en iyi eşitleme performansını göstermiştir. Yetenek dağılımı farklılaştıkça hem ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşım yöntemi hem de ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım yöntemi tam ÇB-MTK gözlenen puan yönteminden daha iyi performans göstermiştir. Hiçbir yetenek dağılımı ortalama farkı olmadığında, üç ÇB-MTK test eşitleme yöntemi de nispeten iyi performans göstermiştir. Fakat yetenek dağılımı farklılaştıkça fark olmayan durumlardaki eşitleme sonuçlarına göre tüm eşitleme yöntemleri tutarsız sonuçlar üretmiştir. Üç ÇB-MTK test eşitleme yönteminin tümü, önemli bir yetenek dağılımı farkı olduğunda TKF ve MKF kalibrasyon yöntemleri altında en iyi performansı göstermiştir. Yetenek dağılımı değişikliği olduğunda, eşitleme sonuçları, OD yöntemi ve M yöntemleri altında NOP yöntemi altında olduğundan daha küçük tutarsızlıklara sahiptir. NOP kalibrasyon yöntemi altındaki eşitleme yöntemleri, yetenek dağılımı farklılıkları olduğunda en düşük sağlamlığa sahiptir. ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşım yöntemi en küçük yanlılık değerlerine sahiptir. Yetenek ortalama farkı olduğunda, yetenek dağılımı değişikçe tüm kalibrasyon yöntemlerinde üç test eşitleme yöntemi için RMSD değerlerinin arttığı görülmüştür. ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşım yöntemi en küçük ve tam ÇB-MTK gözlenen puan eşitleme büyük RMSD değerlerine sahiptir. NOP kalibrasyon yöntemi kullanıldığında

üç test eşitleme yöntemi için de en büyük RMSD değerleri gözlenmiştir. Yetenek dağılımı farklılaştıkça RMSD değerleri artmıştır.

Lee ve Brossman (2012) basit yapılı ÇB-MTK (BY-ÇBMTK) çerçevesi altında bir MTK gözlenen puan eşitleme yöntemi geliştirmiş ve geliştirilen yöntemin doğruluğunu random grup deseni altında incelemiştir. Çalışmada gerçek ve simülasyon veri setleri kullanılmıştır. Gerçek veri seti Advanced Placement testinden elde edilmiştir. 6000 kişi ve 73 maddeden oluşmaktadır. Madde türü boyutluluk kaynağı olarak ele alınmıştır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü 3000 olmak üzere sabit tutulmuş olup boyutlar arasındaki ilişki düzeyi ( $\rho=0,5, 0,8$  ve  $0,95$ ) manipüle edilen faktör olarak ele alınmıştır. Eşitlemenin doğruluğu hata karelerinin ortalaması, yanlılığın karesi ve varyans ile değerlendirilmiştir. Çalışmanın sonuçlarına göre BY-ÇBMTK yöntemi ve eşit yüzdellikli eşitlemenin benzer sonuçlar verdiği belirlenmiştir. BY-ÇBMTK yöntemi için genel eşitleme hatasının TB-MTK gözlenen puan yöntemi için elde edilen hatadan çok daha küçük olduğu belirlenmiştir. Boyutlar arasındaki korelasyon azaldıkça BY-ÇBMTK yönteminin daha az yanlı sonuçlar ürettiği görülmüştür. Madde türünün boyutluluk kaynağı olduğu durumlarda BY-ÇBMTK yönteminin kullanılması önerilmiştir. Boyutlar arasındaki korelasyon .8 ve üstü olduğunda ise TBE yöntemleri tercih edilebilir.

Brossman ve Lee (2013) tam ÇB-MTK gözlenen puan eşitleme, ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım, ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşım yöntemlerini geliştirmiş ve geliştirilen yöntemin doğruluğunu incelemiştir. Çalışmada gerçek veri seti kullanılmıştır. Gerçek veri seti Iowa Eğitimsel Gelişim Testlerinin Seviye 17/18 bataryasının iki formundan elde edilmiştir. Her form için örneklem büyüklüğü 2500'dür. Çalışmanın sonuçlarına göre çok boyutlu yöntemlerin benzer sonuçlar verdiği gözlenmiştir. Çalışmada sunulan çok boyutlu eşitleme yöntemlerinin, veriler tek boyutlu olmadığında TBE yöntemlerinin yeterli bir alternatifi olduğuna karar verilmiştir.

Lee (2013) TB-MTK gözlenen ve gerçek puan, tam ÇB-MTK gözlenen puan, ÇB-MTK gerçek ve gözlenen puan eşitleme tek boyutlu yaklaşımı ve eşit yüzdellikli eşitleme



yöntemlerini random grup deseni altında karşılaştırmıştır. Çalışmada simülasyon verileri kullanılmıştır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü (2000 ve 6000); madde sayısı (30 ve 60) ve boyutlar arasındaki ilişki düzeyi ( $\rho=0,5$  ve  $0,8$ ) manipüle edilen faktörler olarak ele alınmıştır. Eşitlemenin doğruluğu eşitlemenin standart hatası (SEE), işaretli yanlılık, hata kareler ortalamasının karekökü (RMSE) ve DTM (farkın 0,5'ten fazla olması) ile değerlendirilmiştir. Çalışmanın sonuçlarına göre tam ÇB-MTK gözlenen puan yöntemi, özellikle boyutlar arasındaki korelasyon düşük olduğunda, diğer eşitleme yöntemlerine göre daha doğru eşitleme sonuçları (yani daha az hata değeri) sağlamıştır. Eşit yüzdelikli eşitleme, tüm koşullarda daha yüksek miktarda random hata üretmiştir. Çok boyutlu testler için eşitleme yöntemleri, form farklılıkları küçük, örneklem büyük ve test uzun olduğunda daha doğru sonuçlar vermektedir. Çok boyutlu testler kullanıldığında yani tek boyutluluk varsayımı ihlal edildiğinde bile TBE yöntemleri oldukça doğru sonuçlar vermektedir. Test eşitleme yönteminin gözlenen ya da gerçek puan yöntemi olması eşitleme sonuçlarında herhangi bir farklılık yaratmamıştır.

Peterson (2014) karma testlerde ÇB-MTK gözlenen puan eşitleme yöntemlerini random grup deseni altında incelemiştir. Tam ÇB-MTK, bi-faktör ÇB-MTK ve tek boyutlu gözlenen puan, eşit yüzdelikli eşitleme yöntemleri karşılaştırmıştır. Çalışmada gerçek veri seti kullanılmıştır. Gerçek veri seti 2011 yılında uygulanan Advanced Placement testinden elde edilmiştir. Eşitlemenin doğruluğu kök ortalama kare farkı (RMSD) ve DTM ile değerlendirilmiştir. Çalışmanın sonuçlarına göre boyutluluk arttıkça daha iyi performans gösterdiği belirlenmiştir. Bi-faktör ÇB-MTK gözlenen puan eşitleme yönteminin, gruba özgü faktörlerin tanımlanma şekline çok duyarlı olmadığı belirlenmiştir. Eşitleme sonuçlarındaki benzerlikler genellikle tek boyutlu, bi-faktör ve tam ÇB-MTK gözlenen puan yöntemlerinde görülmüştür.

Lee ve Lee (2016) karma formatlı testler için bi-faktör çok boyutlu madde tepki kuramı (BF-MIRT) gözlenen puan eşitleme yöntemini geliştirmiş ve geliştirilen yöntemin doğruluğunu random grup deseni altında incelemiştirlerdir. Çalışmada gerçek ve simülasyon

veri setleri kullanılmıştır. Gerçek veri seti Advanced Placement (AP) testinden elde edilmiştir. 3 tür veri seti oluşturulmuştur: eşleşen gruplar veri seti, aynı formun yarılanmasından elde edilen gruplara ait veri seti ve simülasyon veri seti. Madde türü boyutluluk kaynağı olarak ele alınmıştır. Eşleşen gruplar oluşturulurken aynı sayıda araştırmacıyı rastgele seçerek eşleşen örnekler oluşturmak için üç arka plan değişkeni (düşük ücret, ebeveyn eğitimi ve cinsiyet) ve ortak madde puanları kullanılmıştır. İkinci veri seti için tüm veri setinden 6000 kişi random olarak seçilmiştir. Test tek sayılı ve çift sayılı maddeler olarak iki gruba ayrılmıştır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü 3000 ve madde sayısı 73 (70 çoktan seçmeli ve 3 çoklu puanlanan madde) olmak üzere sabit tutulmuş olup boyutlar (madde türleri) arasındaki ilişki düzeyi ( $\rho=0,5, 0,6, 0,7, 0,8, 0,9$  ve  $0,99$ ) manipüle edilen faktörler olarak ele alınmıştır. Eşitlemenin doğruluğu BF-MIRT ve tek boyutlu madde tepki kuramı (TB-MTK) gözlenen puan eşitleme sonuçlarının karşılaştırılması ile değerlendirilmiştir. Çalışmanın sonuçlarına göre BF-MIRT gözlenen puan eşitleme yöntemi, TB-MTK gözlenen puan eşitleme yöntemine benzer eşitleme sonuçları üretmiştir. Boyutlar arasındaki korelasyon küçük olduğunda yöntemler farklı sonuçlar vermiştir. Korelasyon düşük ( $0,5$  ve altı) olduğunda çok boyutlu yaklaşımların kullanılması önerilmiştir.

Atar ve Yeşiltaş (2017) çok boyutlu ölçek kalibrasyon yöntemlerinden Stocking-Lord, ortalama/ortalama ve ortalama/sigma yöntemlerinin performanslarını denk olmayan gruplarda ortak madde deseni altında incelemiştir. Çalışmada simülasyon verileri kullanılmıştır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü (1000 ve 2000), yetenek dağılımı ortalamaları ((0,0), (-0,5, +0,5)), ortak madde oranı (%15, %30, %60 ve %100) ve boyutlar arasındaki ilişki düzeyi ( $p=0, 0,5$  ve  $0,8$ ) manipüle edilen faktörler olarak ele alınmıştır. Eşitlemenin doğruluğu hata kareler ortalamasının karekökü (RMSE) ve yanlılık (BIAS) ile değerlendirilmiştir. Çalışmanın sonuçlarına göre yetenek dağılımı faktörünün, üç yöntem için de madde ayırıcılık ve madde güçlük parametreleri kestirimlerine önemli bir etkisinin olmadığı bulunmuştur. Örneklem büyüklüğü arttıkça RMSE ve yanlılık

değerleri azalmıştır. Ortak madde oranı en fazla ortalama/sigma yönteminin sonuçlarını etkilemiştir.

Kim, Lee ve Kolen (2019) basit yapılı ÇB-MTK gerçek puan yöntemini geliştirmiş ve geliştirilen yöntemin doğruluğu random grup deseni altında incelemiştir. Çalışmada gerçek ve simülasyon veri setleri kullanılmıştır. Gerçek veri seti Advanced Placement testinden elde edilmiştir. 6000 kişi ve 57 maddeden oluşmaktadır. Madde türü boyutluluk kaynağı olarak ele alınmıştır. Eşitlemenin doğruluğu DTM ile değerlendirilmiştir. Çalışmada simülasyon veri setleri için örneklem büyüklüğü (1000 ve 5000) ve boyutlar arasındaki ilişki düzeyi ( $\rho=0,5, 0,8$  ve  $0,95$ ) manipüle edilen faktör olarak ele alınmıştır. Eşitlemenin doğruluğu eşitlemenin standart hatası (SEE), işaretli yanlılık (İY) ve hata kareler ortalamasının karekökü (RMSE) ile değerlendirilmiştir. BY-ÇBMTK gerçek puan yönteminden elde edilen sonuçlar eşit yüzdelikli, TB-MTK gerçek ve gözlenen puan ve BY-ÇBMTK gözlenen puan eşitleme yöntemlerinden elde edilen sonuçlar ile karşılaştırılmıştır. Çalışmanın sonuçlarına göre tek boyutlu gerçek puan eşitleme yönteminin diğer yöntemlere kıyasla daha küçük eşitlenmiş puanlar üretme eğiliminde olduğu belirlenmiştir. BY-ÇBMTK gerçek puan yöntemi, veriler çok boyutlu olduğunda geleneksel TB-MTK eşitleme ile karşılaştırıldığında daha doğru eşitleme sonuçları gösterdi. BY-ÇBMTK gerçek puan yöntemi, tüm çalışmalarda basit yapılı ÇB-MTK gözlenen puan eşitleme yöntemine benzer şekilde performans göstermiştir.

Kumlu (2019) test ve alt test puanlarının MTK gerçek puan eşitleme yönteminden elde edilen eşitleme sonuçlarına göre madde ve yetenek parametrelerinin yanlılık (BIAS), eşitlemenin standart hatası (SEE), eşitleme hatası (RMSE) değerlerine çeşitli koşulların etkilerini incelemiştir. Çalışmada simülasyon verileri kullanılmıştır. İçerik alanları (Türkçe, Sosyal Bilimler, Matematik ve Fen Bilimleri) boyutluluk kaynağı olarak ele alınmıştır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü (1000 ve 3000), testler ve alt testler arası güçlük düzeyi (0,05, 0,10, 0,15 ve 0,20), ortak madde oranı (%10, %20, ve %30), alt testlerdeki madde sayısı (10,20,30), alt testler arası ilişki düzeyi ( $p=0,2, ,5$  ve  $0,8$ )

ve ölçek kalibrasyon yöntemleri (Ortalama-Ortalama ve Stocking-Lord) manipüle edilen faktörler olarak ele alınmıştır. Eşitlemenin doğruluğu karekök ortalama hata (RMSE), eşitlemenin standart hatası (SEE) ve yanlılık (BIAS) ile değerlendirilmiştir. Çalışmanın sonuçlarına göre test düzeyinde tek boyutlu madde parametrelerine göre test düzeyinde eşitleme yapıldığında alt testler arası ilişki düzeyinin artmasıyla hata miktarlarının azaldığı gözlenmiştir. Test uzunluğunun ve ortak madde oranının artmasıyla hata değerlerinin azaldığı gözlenmiştir. Stocking-Lord yönteminin Ortalama-Ortalama yönteminden daha iyi eşitleme sonuçları ürettiği belirlenmiştir. Alt test düzeyinde tek boyutlu madde parametrelerine göre test düzeyinde eşitleme yapıldığında alt testler arası ilişki düzeyinin artmasının hata değerleri üzerinde çok fazla etkisinin olmadığı gözlenmiştir. Örneklem büyüklüğü ve test uzunluğu arttıkça hata azalmıştır. Ortak madde oranı artmasının hata değerleri üzerinde fazla etkisi olmadığı gözlenmiştir. Test düzeyinde çok boyutlu madde parametrelerine göre test düzeyinde eşitleme yapıldığında örneklem büyüklüğü arttıkça hata azalmıştır. Alt testler arası ilişki düzeyinin azalmasıyla hata miktarları azalmıştır. Alt test düzeyinde tek boyutlu madde parametrelerine göre alt test düzeyinde eşitleme yapıldığında örneklem büyüklüğü ve test uzunluğu arttıkça hata azalmıştır. Hata değerlerinin alt testler arasındaki ortalama güçlük farkına göre farklılık göstermediği gözlenmiştir. Ölçek kalibrasyon yöntemlerinin hata değeri üzerinde etkisinin olmadığı gözlenmiştir.

Panidvadtana vd. (2019) karma testler için çok boyutlu madde tepki kuramı eşitleme yöntemlerinin doğruluğunu denk olmayan gruplarda ortak madde deseni altında karşılaştırmışlardır. Çalışmada simülasyon verileri kullanılmıştır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü 3000 ve madde sayısı 100 olmak üzere sabit tutulmuş olup ölçek kalibrasyon yöntemleri (eş zamanlı ve ayrı kalibrasyon ile test karakteristik eğrisine dayalı yöntemler) ve ortak madde oranı (%10, %20 ve %30) manipüle edilen faktörler olarak ele alınmıştır. Eşitlemenin doğruluğu eşitleme hatasının varyansı ile değerlendirilmiştir. Çalışmanın sonuçlarına göre Tam ÇB-MTK gözlenen puan eşitleme

yöntemi için eş zamanlı kalibrasyon yöntemin kullanıldığında karakteristik eğri yöntemi kullanılan durumlara göre daha küçük hata varyansı gözlenmiştir. Eş zamanlı kalibrasyon yöntemi kullanıldığında Tam ÇB-MTK gözlenen puan eşitleme için ortak madde oranındaki değişim hata varyansını etkilememiştir. Test karakteristik eğri yöntemi kullanıldığında ise ortak madde oranı arttıkça hata varyansı azalmıştır.

Uğurlu (2020) değişen madde fonksiyonu gösteren maddelerde BY-ÇBMTK gözlenen puan, tek boyutlu gözlenen ve gerçek puan ile eşit yüzdellikli eşitleme yöntemlerini denk olmayan gruplarda ortak madde deseni altında karşılaştırmıştır. Çalışmada simülasyon verileri kullanılmıştır. İçerik alanları boyutluluk kaynağı olarak ele alınmıştır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü 4000 (odak grup 1000 ve referans grup 3000), madde sayısı 80 (20 ortak madde) ve DMF maddelerinin sıklığı %20 olmak üzere sabit tutulmuş olup boyutlar arasındaki ilişki düzeyi ( $p=0,5, 0,8$  ve  $0,95$ ) manipüle edilen faktörler olarak ele alınmıştır. Eşitlemenin doğruluğu korelasyon, kök kare farkı (RSD), kök ortalama kare farkı (RMSD) ve DTM ile değerlendirilmiştir. Orta düzeyde DMF içeren maddeler yalnızca ortak maddelerde oluşturulmuştur. Çalışmanın eşit yüzdellikli eşitleme yönteminin düşük frekanslı puanlarda yüksek örneklem değişmezliği verdiği gözlenmiştir. Bu nedenle, eşit yüzdellikli eşitleme yönteminin sadece yüksek frekanslı puan aralıklarında kullanılması önerilmektedir. Diğer tüm yöntemler benzer sonuçlar vermektedir. BY-ÇBMTK gözlenen puan ve eşit yüzdellikli eşitleme yöntemleri, iki form arasındaki grup ortalama yetenek farkı, alt gruplar ve toplam grup için 0,1 veya 0,3 olduğunda kullanılabilir.

Kim ve Lee (2022) basit yapılı çok boyutlu testler için var olan yöntemlerin eksikliklerinden yola çıkarak basit yapılı ayrı gözlenen puan eşitleme yöntemini random grup ve ortak madde deseni altında geliştirmişlerdir. Yöntemde toplam puan için eşitlenmiş puanlar elde edilmesinin yanında boyutlar için eşitlenmiş puanlarında elde edilmesi amaçlanmıştır. Madde türü ve içerik alanları boyutluluk kaynağı olarak ele alınmıştır. Her form için örneklem büyüklüğü 5000'dir. Geliştirilen yöntem TB-MTK gerçek ve gözlenen puan, basit yapılı marjinal gözlenen puan ve basit yapılı gerçek puan yöntemleri ile

karşılaştırılmıştır. Çalışmanın sonuçlarına göre basit yapılı gerçek puan ve basit yapılı ayrı gözlenen puan eşitleme alt ve toplam puan düzeyinde eşitlemeye izin vermektedir. Fakat alt boyutlardaki madde sayısı tüm teste göre az olduğu için alt test puanı bazında eşitlenmiş puanların elde edilmesi istendiğinde alt testlerin güvenilirliğinin yüksek olması önerilmiştir. Geliştirilen yöntemin boyut sayısı arttıkça kullanılmasının zorlaştığı gözlenmiştir. Monte Carlo yaklaşımının kullanılması önerilmiştir.

Kim (2022) bi-faktör model için entegrasyon ve projektif MTK (PIRT) tabanlı tek boyutlu MTK gerçek puan yöntemi denk olmayan gruplarda ortak madde deseni tasarımına genişletmiş ve iki test eşitleme yönteminin performansını karşılaştırmıştır. Çalışmada gerçek ve simülasyon veri setleri kullanılmıştır. Gerçek veri seti 2011 ve 2015 yıllarında uygulanan TIMSS testinden elde edilmiştir. Çalışmada simülasyon veri setleri için örneklem büyüklüğü 3000 ve her biri 15 maddelik üç alandan oluşan toplam 45 madde olmak üzere madde sayısı sabit tutulmuş olup yerel madde bağımlılığı (LID), her bir alan için ortak madde sayısı (2,3,4,5 ve 6) ve sınava giren grupların ortalama yeteneklerindeki fark manipüle edilen faktörler olarak ele alınmıştır. Bi-faktör modeli verilere uydurmak amacıyla, gerçek ve simülasyon verileri için aynı içerik alanına ait olan maddelerin aynı spesifik faktöre yüklendiği varsayılmıştır. Eşitlemenin doğruluğu yanlılık, eşitlemenin standart hatası (SEE), hata kareler ortalamasının karekökü (RMSE) ve DTM (farkın 0.5'ten fazla olması) ile değerlendirilmiştir. Çalışmanın sonuçlarına göre genel boyut için denk gruplarda LID seviyesi düştükçe PIRT yönteminin daha doğru eşitleme sonuçları verdiği belirlenmiştir. Ancak, grup farkının artmasıyla entegrasyon yönteminin performansı iyileşmiştir. Bi-faktör eşitleme yöntemleri tek boyutlu 3 parametrelilik model (3PLM) ile karşılaştırıldığında LID değeri küçük olduğunda 3 PLM daha iyi sonuç vermiştir. Bu nedenle LID değeri küçük olduğunda 3 PLM önerilmiştir. Ancak LID değeri orta olduğunda, 3PLM için genel eşitleme hatası, birçok simülasyon koşulu için DTM kriterini aşmıştır. Ortak madde sayısı arttıkça tüm yöntemler daha iyi sonuçlar vermiştir. Simülasyon sonuçlarına göre, doğru eşitleme sonuçları elde etmek için alan başına üç ila dört ortak maddenin yeterli olduğu belirlenmiştir.

Diferansiyel performanslarına rağmen, entegrasyon ve PIRT tabanlı yöntemleri, çoğu simülasyon koşulu için DTM kriterine göre pratik olarak farklı olmayan eşitleme hataları göstermiştir.

Zor (2023), ortalama-ortalama, ortalama-sigma ve Stocking Lord ölçek kalibrasyon yöntemlerinden elde edilen madde ve yetenek parametrelerinin eşitleme hatalarını karşılaştırmıştır. Çalışmada simülasyon verileri kullanılmıştır. Çalışmada simülasyon veri setleri için örneklem büyüklüğü (1000 ve 2000); ortak madde oranı (%20 ve %40); boyutlar arasındaki ilişki düzeyi ( $p=0,1, 0,5$  ve  $0,9$ ) ve parametre tahmin modeli (2 ve 3 PLM) manipüle edilen faktörler olarak ele alınmıştır. Eşitlemenin doğruluğu hata kareler ortalamasının karekökü (RMSE) ile değerlendirilmiştir. Çalışmanın sonuçlarına göre örneklem büyüklüğü ve ortak madde oranı arttıkça ve boyutlar arası korelasyon azaldıkça azaldığı gözlenmiştir. Ortalama-sigma yöntemi kullanıldığında daha yüksek eşitleme hataları elde edilmiştir. Madde ayırıcılık parametresi için 2PLM'de daha düşük RMSE değerleri elde edilmiştir. Madde güçlük ve yetenek parametrelerinin tahmininde ise, Stocking-Lord yöntemi için 2PLM'de ve ortalama-ortalama ve ortalama-sigma yöntemleri için 3PLM'de daha düşük RMSE değerleri gözlenmiştir.

## Bölüm 3

### Yöntem

Bu bölümde temel alınan araştırma türü, araştırmanın deseni, simülasyon koşulları, verilerin üretilmesi ve gerçekleştirilmiş analizlerle ilgili bilgiler yer almaktadır.

#### Araştırmanın Türü

Bu araştırmada, çok boyutlu testlerden elde edilen puanların eşitlenmesinde kullanılan yöntemlerden denk olmayan gruplarda ortak madde deseni altında elde edilen eşitlenmiş puanların ve bu puanlara ait hata değerlerinin çeşitli faktörlere göre karşılaştırılması amaçlanmıştır. Bu amaç doğrultusunda araştırma, simülasyon araştırması niteliği taşımaktadır. Simülasyon araştırmaları, bilinen olasılık dağılımlarından sözde rastgele örnekleme yoluyla veri oluşturmayı içeren bilgisayar deneyleridir. Simülasyon araştırmaları, birçok senaryoyu/koşulu kapsayabilen daha genel analitik sonuçların aksine, belirli senaryolarda istatistiksel yöntemlerin performansı hakkında ampirik sonuçlar elde etmek için kullanılır (Morris vd., 2019).

Araştırmada çeşitli koşulların hata değerleri ile olan ilişkisi incelenmiştir. Bu yönüyle araştırma ilişkisel araştırma niteliği de taşımaktadır. Değişkenler arasındaki ilişkileri belirlemek ve muhtemel sonuçları tahmin etmek için kullanılan araştırmalara ilişkisel araştırma denir.

#### Test Eşitleme Deseni

Araştırmada denk olmayan gruplarda ortak madde deseni (Nonequivalent Anchor Tests/NEAT) kullanılmıştır. Denk olmayan gruplarda ortak madde deseninde Form X (yeni test formu) ve Form Y (eski test formu) ortak maddeler içerir. X ve Y test formlarını alan bireylerin yetenek düzeyleri farklılaşmaktadır (Kolen & Brennan, 2014). İki test formunda yer alan ortak maddeler ile grupların yetenek düzeyleri ve formların güçlük düzeyleri arasındaki farklar giderilmeye çalışılır. Ortak maddelerden alınan puanlar bireylerin testten



aldıkları puanlara ekleniyor ise iç ortak; testten alınan puanlara eklenmiyor ve ayrı değerlendiriliyorsa dış ortak test olarak isimlendirilir.

### **Simülasyon Koşulları**

Çok boyutlu test eşitleme ile ilgili alan yazındaki araştırmalar incelendiğinde bu araştırmaların bazılarında büyük örnekleme sahip standart testlerden elde edilen gerçek veriler kullanılırken (Brossman, 2010; Peterson, 2014), bazılarında ise araştırma koşullarını sağlayan gerçek verilere ulaşılamadığı için simülasyon verileri kullanılmıştır (Lee, 2013; Lee ve Lee, 2016; Zhang, 2012). Bu araştırmada, incelenen koşulları sağlamak gerçek verilerde mümkün olmayacağı için çeşitli koşulların çok boyutlu test eşitleme yöntemleri üzerine etkilerini belirlemek amacıyla Monte-Carlo simülasyon çalışması yürütülmüştür. Araştırmada simülasyon veri setleri için örneklem büyüklüğü, ortak madde oranı, boyutlar arasındaki ilişki düzeyi, kalibrasyon ve test eşitleme yöntemleri manipüle edilen faktörler olarak ele alınmıştır. Örneklem büyüklüğü 3, ortak madde oranı 3, boyutlar arasındaki ilişki düzeyi 3, kalibrasyon yöntemleri 2 ve çok boyutlu test eşitleme yöntemleri 3 farklı koşul içermek üzere toplamda 162 (3x3x3x2x3) koşul incelenmiştir. Araştırma sonuçlarının tutarlı ve genellenebilir olabilmesi için replikasyon yapılmıştır. Yapılan birçok araştırmada 100 replikasyon yapıldığı görülmüştür. Bu nedenle araştırmada her simülasyon koşulu 100 kere tekrarlanmıştır.

### **Boyut Sayısı**

Alan yazındaki çalışmalar incelendiğinde kullanılan veri setlerinin boyut sayılarının 2 ile 4 arasında değiştiği belirlenmiştir. Brossman ve Lee (2013) Iowa Eğitim Geliştirme Testinin iki formunu eşitledikleri çalışmalarında 3 boyutlu veri seti kullanmıştır. Lee (2013) çalışmasında 2 boyutlu veri setini tercih etmiştir. Peterson (2014) gerçek veriler ile gerçekleştirdiği çalışmasında Advanced Placement (AP) testinin 3 içerik alanını kullanmıştır. Kim (2018) basit yapılı çok boyutlu testlerde eşitleme çalışmasında 2 ve 3 boyutlu veriler kullanmıştır. Kumlu (2019) alt testleri eşitlediği çalışmasında 4 boyutlu veri

seti kullanmıştır. Uğurlu (2020) DMF içeren maddelerde 2 boyutlu veri seti kullanmıştır. Lee ve Brossman (2012), Kim, Lee ve Kolen (2019) ve Lee ve Lee (2016) madde türünü boyutluluk kaynağı olarak ele almış ve 2 boyutlu veri seti kullanmıştır. Zhang (2012), yaklaşık basit ve karmaşık yapıları çok boyutlu testlerde eşitleme çalışmasında 2 boyutlu veriler kullanmıştır. Alan yazındaki çalışmalar ve bi-faktör model yapısı göz önünde bulundurularak verilerin 3 boyutlu olmasına karar verilmiştir.

### **Örneklem Büyüklüğü**

Tek boyutlu ölçek kalibrasyonunda ve test eşitlemede örneklem büyüklüğü arttıkça sonuçların doğruluğunun ve kesinliğinin arttığı bilinmektedir (Çokluk vd., 2022; Gök ve Kelecioğlu, 2014; Lee vd., 2014; Kilmen ve Demirtaşlı, 2012; Kolen ve Brennan, 2014; Wang, 2006). Benzer şekilde ÇB-MTK ölçek kalibrasyon ve test eşitleme yöntemlerinin performansı da örneklem büyüklüğünden etkilenir. ÇB-MTK modellerinde daha fazla parametrenin kestirilmesi gerektiğinden, ÇB-MTK ölçek kalibrasyonu için TB-MTK ölçek kalibrasyonundan daha büyük örneklem büyüklüğü gereklidir. Skaggs ve Lissitz (1986), örneklem büyüklüğünün en az 1000 olması gerektiğini belirtmiştir. Kolen ve Brennan'a (2004) göre MTK gerçek puan eşitleme yöntemlerinde denk olmayan gruplarda ortak madde deseni altında 3 PLM kullanıldığında her form için örneklem büyüklüğünün 1500 olması gerekmektedir. Li ve Lissitz (2000) çok boyutlu madde tepki kuramına dayalı ölçekleme yöntemlerini karşılaştırdıkları çalışmalarında kararlı sonuçlar elde etmek için 2000 örneklem büyüklüğünü önermiştir. Tate (2003), örneklem büyüklüğü 2000 ve üstünde olduğu zaman çok boyutlu bilgisayar programlarının birçoğunun iyi performans gösterdiğini belirlemiştir. Wang ve Liu (2018) karma testler için random grup deseni altında eşit yüzdelli eşitleme için gerekli olan minimum örneklem büyüklüğünü incelemiştir. Eşitleme doğruluğunun 500 ile 3000 örneklem büyüklüğü arasında daha fazla değiştiği belirlenmiştir. 3000 ve 8000 örneklem büyüklükleri arasındaki değişimin daha az olduğu görülmüştür. Alan yazındaki birçok test eşitleme durumu için 3000 örneklem büyüklüğü büyük örneklem olarak ele alınmıştır. Yapılan çalışmalar 5000 ve üstündeki örneklemelerin makul derece doğru

sonuçlar verdiğini göstermektedir (Kim, Lee ve Kolen, 2019). Alan yazın göz önünde bulundurularak bu araştırmada örneklem büyüklükleri küçük örneklemi temsil etmesi amacıyla 500, orta örneklemi temsil etmesi amacıyla 1500 ve büyük örneklemi temsil etmesi amacıyla 3000 olarak belirlenmiştir.

### ***Madde Sayısı ve Ortak Madde Oranı***

Madde sayısı ve ortak madde oranı, ölçek kalibrasyon ve test eşitleme araştırmalarında önemli bir konudur. Madde sayısı ve ortak madde oranı, ölçek kalibrasyon ve test eşitleme sonuçlarını etkilemektedir. Kolen ve Brennan (2004) tarafından önerilen bir kural, birden fazla içerik alanını ölçen testlerle eşitleme yapılırken testlerin en az 30-40 maddeden oluşması gerektiğidir. Bunun yanı sıra bu araştırmada bi-faktör model diklik varsayımı ihmal edilmiştir. Pekmezci (2018) bi-faktör (bi-factor) kuramının, birey parametrelerinin kestiriminde test uzunluğu arttıkça diklik varsayımı ihlalini daha iyi tolere edebildiğini; aynı zamanda madde sayısının 40 ve üstünde olması gerektiğini belirtmiştir. Madde sayısı arttıkça hata miktarı azalmaktadır. Bu nedenle madde sayısı belirlenirken diklik varsayımı ihlalinin daha iyi tolere edilebilmesi ve madde sayısının hata miktarını etkilemesinin önünce geçilmesi amaçlanmıştır. Alan yazındaki önerilerde göz önünde bulundurularak her boyutun 30, tüm testin ise 60 maddeden oluşmasına karar verilmiştir.

Angoff (1971) madde sayısı 40 ve altında olduğunda ortak madde oranının %20; madde sayısı 40'ın üstünde olduğunda ise ortak madde oranının %30 olması gerektiğini belirtmiştir. Spence (1996) eşitleme için ortak madde sayısının testteki tüm madde sayısının %20'si ile %30'u arasında değişmesi gerektiğini söylemiştir. Çok boyutlu testler için ortak madde oranı %50 kullanılmaktadır (Peterson, 2014). Bu araştırmada ortak madde oranı alan yazındaki öneriler göz önünde bulundurularak %20, %35 ve %50 olarak belirlenmiştir.

### ***Boyutlar Arasındaki İlişki Düzeyi***

Alan yazındaki çalışmalar incelendiğinde boyutlar arasındaki ilişki düzeyinin test eşitleme sonuçları üzerinde etkisinin olduğu belirlenmiştir. Lee ve Lee (2016) çalışmalarında TB-MTK ve ÇB-MTK test eşitleme yöntemlerini kullanmışlardır.  $\rho=0.8$  ve üzeri korelasyon değerinde TB-MTK ve ÇB-MTK eşitleme sonuçlarının benzer olabileceği sonucuna varmışlardır. Boyutlar arasındaki yüksek korelasyon boyutluluğun etkisini azaltmakta ve testler tek boyutluluğa yakınsamaktadır. Bu araştırmada çok boyutlu test eşitleme yöntemlerinin karşılaştırılacağından verilerin tek boyutluluğa yakınsamasının önüne geçilmesi amacıyla boyutlar arasındaki ilişki düzeyi  $\rho=0, 0,3$  ve  $0,6$  olarak belirlenmiştir.

### ***Kalibrasyon Yöntemleri***

Denk olmayan gruplarda ortak madde deseni altında TKF yöntemi ile yapılan ölçek kalibrasyonlarının iyi sonuçlar gösterdiği alan yazında ortaya konulmuştur (Atar & Yeşiltaş, 2017; Baker & Al Karni, 1991; Yao & Boughton, 2009). Bu nedenle araştırmada, eş zamanlı kalibrasyon yönteminin yanı sıra TKF'ye dayanan Stocking Lord ayrı kalibrasyon yönteminin kullanılması tercih edilmiştir. Simülasyon koşulları Tablo 1'de özetlenmiştir.

**Tablo 1**

### ***Simülasyon Koşulları***

Eşitleme Deseni	Örneklem Büyüklüğü	Ortak Madde Oranı	Boyutlar Arası İlişki Düzeyi	Kalibrasyon Yöntemleri	Dönüşüm Yöntemler	Replikasyon Sayısı
Denk olmayan gruplarda ortak madde deseni (CINEG)	500	%20	0	Ayrı (Stocking Lord)	MOSE AOSE	100
	1500	%35	,30	Eş zamanlı	Bi-faktör MIRT gözlenen puan	
	3000	%50	,60			

Not: MOSE tam ÇB-MTK gözlenen puan eşitleme, AOSE ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım yöntemidir.

### ***Verilerin Üretilmesi***

Araştırmada uygulama ve anlaşılma kolaylığı nedeniyle ikili puanlanan madde cevap verisi üretilmiştir. Her bir koşul altında her bir grup için cevaplar telafi edici çok boyutlu üç parametrelili lojistik (M-3PLM) MTK kullanılarak üretilmiştir. Veri setlerini oluşturmak için, bir genel faktör (g) ve iki spesifik faktörden (f1 ve f2) oluşan bi-faktör model kullanılmıştır.

Birinci boyut genel faktör; ikinci üçüncü boyut yük verdikleri maddeler için ikinci özel faktör olarak ele alınmıştır. Tüm maddeler bir genel ve bir özel faktöre çapraz yüklenecek şekilde model oluşturulmuştur. İlk otuz madde genel ve birinci özel faktöre yüklenmiştir. Son otuz madde genel ve ikinci özel faktöre yüklenmiştir. Madde cevap verisi, her koşul için madde (60 madde) ve yetenek parametrelerinden 100 kez üretilmiştir.

Verilerin istenen koşulları sağlayıp sağlamadığı kontrol edilmiştir. Üretilen verilerin boyutluluk kontrolleri açımlayıcı ve doğrulayıcı faktör analizi, yamaç birikinti grafiği ile kontrol edilmiştir. Bi-faktör modelinin değerlendirilmesinde çeşitli ölçütlerden yararlanılır. Omega katsayıları, ortak varyansı açıklama oranı (ECV) ve grup faktörlerinin etkisi bulaşmamış korelasyonların oranı (PUC) bu ölçütlerden bazılarıdır. ECV verilerin tek boyutluluğunu inceler ve genel faktörün açıkladığı varyans yüzdesidir. ECV, 0,70'in altında ise verilerin çok boyutlu; 0,90'ın üzerinde ise verilerin tek boyutlu olarak olduğunu göstermektedir (Quinn, 2014). PUC, genel faktörün özel faktörlerden bağımsız oluşturduğu korelasyon sayısıdır. Yüksek PUC değerleri test edilen modelin tek boyutlu olarak ele alınabileceğini göstermektedir. Çalışmada üretilen verilerin bi-faktör modele uygunluğu ECV ve PUC ölçütleri ile değerlendirilmiştir.

### ***Madde Parametrelerinin Üretilmesi***

Çok boyutlu madde tepki kuramında test yapısı, çok boyutlu uzayın her boyutunda madde parametrelerinin birleşiminden oluşmaktadır. Araştırmada çok boyutlu M-3PLM kullanılmıştır. Bunun için madde ayırıcılık (a), madde güçlük (b) ve şans parametresi (c) üretilmiştir. Her boyut için bir tane olmak üzere her madde için üç tane madde ayırıcılık parametresi üretilmiştir. Genel faktör  $a_0$ , birinci özel faktör  $a_1$  ve ikinci özel faktör  $a_2$  ile örneklendirilmiştir. Madde ayırıcılık parametresi, X ve Y testleri için minimum 0.3 maksimum 2 olacak şekilde sürekli tek biçimli (uniform) dağılımdan üretilmiştir.

Madde güçlük parametresi, her madde için bir tane olmak üzere X ve Y testleri için minimum -3 maksimum +3 olacak şekilde ortalaması 0, standart sapması 1 olan normal dağılımdan üretilmiştir. Fakat madde cevap verisi üretilirken madde güçlük parametresi

yerine kesişim parametresi  $d$  kullanılmıştır.  $d$  parametresi MDISC ve MDIFF formülleri kullanılarak madde ayırıcılık ve güçlük parametrelerinden üretilmiştir (Reckase, 2009).

MDISC çok boyutlu madde ayırıcılık parametresi olup maddeye ait ortak ayırıcılık olarak isimlendirilir. Maksimum ayırıcılık parametresi olarak da bilinmektedir. Yetenek ( $\theta$ ) düzeyi uzayındaki en dik eğime sahip noktanın madde yanıt yüzeyi eğimine karşılık gelir (Choi, 2019). MDICS vektör olarak ifade edilir. Vektörün kuyruğu  $p(\theta) = 0,5$  çizgisinde yer alır. Vektörün uzunluğu ile MDISC belirlenir. Tek boyutlu madde tepki kuramındaki madde ayırıcılık indeksi gibi yorumlanmaktadır. MDISC değeri arttıkça maddenin ayırıcılığı artar. MDISC,  $a_{ik}$  her boyut için ayırıcılık olmak üzere aşağıdaki gibidir;

$$MDISC(x_i) = \sqrt{\sum_{i=1}^k a_{ik}}$$

MDIFF çok boyutlu madde güçlüğüdür. Maddenin genel zorluğunun bağlı bir göstergesidir. Bir diğer ifadeyle maddenin  $m$  boyutlu uzaydaki konumunu gösteren parametre, MDIFF olarak da adlandırılır (Peterson, 2014). MDIFF, yetenek ( $\theta$ ) düzeyi uzayındaki orijinden en dik eğim noktasının altındaki noktaya ( $p(\theta) = 0,5$  noktası) olan mesafe ile belirlenir. MDIFF aşağıdaki gibidir;

$$MDIFF(x_i) = \frac{-d_i}{MDISC(x_i)}$$

MDIFF, çok boyutlu madde ayırıcılık vektörünün uzayda hangi bölgede yer aldığını gösterir. MDIFF değeri negatif ise ayırıcılık vektörü uzayda üçüncü bölgede, pozitif ise birinci bölgede yer almaktadır. MDIFF değerinin negatif olması maddenin kolay olduğunu gösterir ve genellikle yetenek ( $\theta$ ) düzeyleri için doğru yanıt verme olasılığı yüksek olan maddelerde gözlenir. MDIFF değerinin pozitif olması ise maddenin zor olduğunu gösterir ve genellikle yetenek ( $\theta$ ) düzeyleri için doğru yanıt verme olasılığı küçük olan maddelerde gözlenir.

Çok boyutlu madde tepki kuramında genellikle MDIFF yerine d parametresi madde güçlüğü indeksi olarak tercih edilir ve dolaylı olarak madde karakteristik yüzeyinin konumu ile ilişkilidir. d parametresi aşağıdaki gibidir;

$$d_i = -(MDIFF(x_i) * MDISC(x_i))$$

Şans parametresi c, X ve Y formları için minimum 0.05 maksimum 0.20 olacak şekilde sürekli tek biçimli (uniform) dağılımdan üretilmiştir. c parametresinin alabileceği maksimum değer, geniş ölçekli sınavlarda tercih edilen 5 seçenekli maddelerin şansla doğru yanıt olma olasılığının %20 olması göz önünde bulundurularak belirlenmiştir. Madde parametrelerini üretmek için R programlama dilinde bulunan 'stats' (R Development Core Team, 2022) paketi kullanılmıştır.

### ***Yetenek Parametrelerinin Üretilmesi***

Grupların yetenek dağılımlarının özelliklerine, ortak maddelerin sayısı gibi test eşitlemedeki diğer faktörleri de etkilediği için dikkat edilmelidir (Cook & Petersen, 1987). Denk olmayan gruplarda ortak madde deseni tasarımına uygun olarak farklı test formlarını alan grupların yetenek düzeyi ortalamaları arasında fark yaratılmıştır. Wang, Lee, Brennan ve Kolen (2008), grupların yetenek dağılımlarının ortalamaları arasındaki farkı 0.05 ve 0.1 aralığında ise geniş ve 0.25 ve daha büyük ise çok geniş olarak adlandırmışlardır. Bu çalışmada grupların yetenek dağılımlarının ortalamaları arasındaki fark 0,05 olarak ele alınmıştır. Yetenek dağılımları her boyut için, eski formu (Y) alan grubun yetenek dağılımı ortalaması 0 standart sapması 1 (N(0,1)); yeni formu (X) alan grubun yetenek dağılımı ortalaması 0.05 standart sapması 1 (N(0.05,1)) olacak şekilde çok değişkenli normal dağılımdan üretilmiştir. Genel faktör  $Q_0$ , birinci özel faktör  $Q_1$  ve ikinci özel faktör  $Q_2$  ile örneklendirilmiştir. Yetenek parametreleri üretmek için R programlama dilinde bulunan 'mvtnorm' (Genz vd., 2021) paketi kullanılmıştır.

İlk üretilen madde ve yetenek parametreleri gerçek/üretici parametreler olarak ele alınmıştır. Bunun için random sabit (seed) değerler atanmış ve diğer koşullarda bu sabit (seed) sayı ile gerçek parametrelerin modeller arası değişmezliği sağlanmıştır.

### **Verilerin Analizi**

Veri setlerinin üretilmesi, madde ve yetenek parametrelerinin kestirilmesi, parametre kalibrasyonları, marjinal puan dağılımlarının belirlenmesi ve eşitleme çalışmaları R programlama dili (R Development Core Team, 2022) kullanılarak gerçekleştirilmiştir.

Araştırmada denk olmayan gruplarda ortak madde deseni altında, eş zamanlı ve ayrı (Stocking Lord) kalibrasyon yöntemleri; tam ÇB-MTK gözlenen puan eşitleme, ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım ve bi-faktör ÇB-MTK gözlenen puan çok boyutlu eşitleme yöntemleri kullanılmıştır. Kullanılan test eşitleme yöntemlerinin son adımlarında gerçekleştirilen eşit yüzdelikli eşitlemede küçük örneklerde diğer düzeltme yöntemlerine göre daha az hata verdiği için (Karagül, 2020; Pak & Lee, 2014; Puhan, 2011) loglinear pre-smoothing yöntemi kullanılmıştır.

Birinci ve üçüncü alt problemlere yanıt aranırken örneklem büyüklüğü 3000, ortak madde oranı %20, boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri seti (koşul 7) kullanılmıştır. Birinci alt problem için eş zamanlı ve ayrı (Stocking Lord) kalibrasyon yöntemlerinden elde edilen madde ve yetenek parametrelerine dayalı eşitleme sonuçları kullanılmıştır. Üçüncü alt problem için testin kendisi ile eşitlenmesinde (kimlik eşitleme/identity equating) koşul 7 için üretilen X ve Y yetenek parametreleri ile iki grup oluşturulmuştur. Bu iki gruba aynı test formunun (X) uygulandığı varsayılarak eşitleme gerçekleştirilmiştir. Eşitleme sonucunda her iki alt problem için tekrarlardan elde edilen puanların her ham puan düzeyinde ortalaması alınmış ve her bir ham puana karşılık gelen eşitlenmiş puan elde edilmiştir.

Üç farklı ÇB-MTK test eşitleme yöntemi uygulandığından, eşitleme yöntemlerinin etkilerini tespit etmek için ANOVA modeli kullanılmıştır. Eşitlenmiş puanların test eşitleme yöntemlerine göre farklılaşıp farklılaşmadığı tek yönlü ANOVA (Varyans Analizi) testi ile



incelenmiştir. Tek yönlü ANOVA testi sonuçlarına göre yöntemlerden elde edilen eşitlenmiş puanlar arasında fark bulunmamıştır. Tek yönlü ANOVA testi sonuçları Ek A'da yer almaktadır. Tek yönlü ANOVA, bir bağımlı değişken için iki veya daha fazla grubun ortalamalarını karşılaştırır (Ross & Wilson, 2017). Tek yönlü ANOVA testi yöntemlerden elde edilen puanların ortalamaları karşılaştırıldığı için yöntemler arasında fark bulunmadığı düşünülmüştür. Bu nedenle puanlar alt, orta ve üst olmak üzere üç gruba ayrılmış ve grup bağımsız değişkeni oluşturulmuştur. Gruplar oluşturulurken puanların normal dağıldığı varsayılmıştır. Normal dağılımda ortalama puandan -1 ve +1 standart sapma uzaklıkta bulunan puanlar orta grubu oluşturmuştur. Puanlar, ortalamanın -1 standart sapma altında kaldığında alt gruba, +1 standart sapma üstünde kaldığında ise üst gruba dahil edilmiştir. Daha sonra gruplara ve yöntemlere göre eşitlenmiş puanların farklılaşp farklılaşmadığı iki yönlü ANOVA testi ile incelenmiştir. İki yönlü ANOVA testi sonuçlarına göre istatistiksel olarak anlamlı bulunan değişkenler için etki büyüklüğü, Cohen eta kare ( $\eta^2$ ) yöntemi ile değerlendirilmiştir. Cohen eta kare yöntemine göre, 0,01'den küçük, 0,06'ya orta ve 0,14'e büyük etki büyüklüğü olarak kabul edilir (Ellis, 2010; Pallant, 2005). Eşitlenmiş puanların hangi gruplar için farklılaştığı Tukey testi ile incelenmiştir. Tukey testi sonuçları Ek B'de yer almaktadır.

### Değerlendirme Ölçütleri

Araştırmada eşitleme sonuçlarının doğruluğunu, her bir test eşitleme yönteminin performansını incelemek amacıyla eşitlenmiş puanlar için yanlılık (BIAS), eşitlemenin standart hatası (SEE) ve hata kareler ortalamasının karekökü (RMSE) değerleri hesaplanmıştır. Tekrarlardan elde edilen eşitlenmiş puanların ortalaması  $\bar{t}(x_i)$ ,  $x_i$  X formundan elde edilen ham puan, R replikasyon sayısı,  $\widehat{t}_{yr}(x_i)$   $x_i$  için r. tekrardan elde edilen eşitlenmiş puan olmak üzere aşağıdaki gibidir;

$$\bar{t}(x_i) = \frac{1}{R} \sum_{r=1}^R \widehat{t}_{yr}(x_i)$$

Test eşitleme değerlendirirken, eşitleme sürecinde üretilen hata iki farklı hata türüne ayrılabilir: Random ve sistematik hata. Eşitlemenin standart hatası sınava giren kişilerin örnekleme kaynaklı eşitlemedeki random hatanın büyüklüğünü belirtmek için kullanılır. Eşitlemenin standart hatası (SEE) aşağıdaki gibidir;

$$SEE(x_i) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\widehat{t}_{yr}(x_i) - \bar{t}(x_i))^2}$$

Yanlılık (BIAS), eşitlemede kullanılan yöntemden kaynaklanan sistematik hatayı belirtmek için kullanılır (Zhang, 2012). BIAS ve RMSE aşağıdaki gibidir (Lee, 2013);

$$BIAS(x_i) = \bar{t}(x_i) - t_y(x_i)$$

$$RMSE(x_i) = \sqrt{(SEE(x_i))^2 + (BIAS(x_i))^2}$$

Her ham puan düzeyi için hata değerleri elde edilmiştir. Bu hata değerlerinin ortalaması alınarak her bir koşul için ortalama hata değerleri hesaplanmıştır. Negatif ve pozitif yanlılık değerlerinin birbirini nötrlemesini önlemek ve aradaki gerçek farkı görebilmek için ortalamaları hesaplanırken yanlılık değerlerinin mutlak değeri alınmıştır (Öztürk Gübeş, 2014; Stone, 1992).

## Bölüm 4

### Bulgular, Yorumlar ve Tartışma

Bu bölümde, çalışmanın her bir alt problemine ilişkin bulgulara, bulgulara yönelik yorumlara ve ulaşılan sonuçların ilgili alan yazın desteğiyle tartışmasına yer verilmiştir.

#### Birinci Alt Probleme Ait Bulgular

**Birinci Alt Problem.** “Denk olmayan gruplarda ortak madde deseni altında tam ÇB-MTK gözlenen puan, ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım ve bi-faktör ÇB-MTK gözlenen puan eşitleme yöntemleri kullanılarak elde edilen eşitlenmiş puanlar farklılaşmaktadır mıdır?”.

Birinci alt problem doğrultusunda yöntemlerden elde edilen eşitlenmiş puanların istatistiksel olarak anlamlı şekilde farklılaşıp farklılaşmadığını incelemek amacıyla iki yönlü ANOVA testi uygulanmıştır. İlgili analizlerin sonuçları Tablo 2’de yer almaktadır.

**Tablo 2**

#### İki yönlü ANOVA Sonuçları

		Ortalama Farkı			F	p	$\eta^2$
		BF-MIRT MOSE	BF-MIRT AOSE	MOSE AOSE			
Eş Zamanlı Kalibrasyon	Yöntem	0,453	0,3207	0,7742	0,01	0,990	0,00003
	Grup	0,1679	0,6543	0,4864	180,45	<2e-16*	0,669
	Yöntem x Grup	0,8571	3,5667	2,7096	0,19	0,943	0,0014
Ayrı Kalibrasyon	Yöntem	4,6792	9,1584	4,4791	1,06	0,348	0,0039
	Grup	1,1430	2,6503	1,5073	175,939	<2e-16*	0,6586
	Yöntem x Grup	2,7470	5,2678	2,5207	1,554	0,189	0,0116

Tablo 2 incelendiğinde eş zamanlı ve ayrı kalibrasyon sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen tam ÇB-MTK gözlenen puan (MOSE), ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım (AOSE) ve bi-faktör ÇB-MTK gözlenen puan eşitleme (BF-MIRT) uygulamalarından elde edilen eşitlenmiş puanlar arasında istatistiksel olarak anlamlı fark bulunmamıştır ( $p>0,05$ ). Gruplardan elde edilen eşitlenmiş puanlar arasında ise istatistiksel olarak anlamlı fark bulunmaktadır ( $p<0,05$ ). Eta

kare ( $\eta^2$ ) değeri incelendiğinde grup değişkeninin eşitlenmiş puanlar üzerinde etkisinin büyük olduğu gözlenmiştir ( $\eta^2=0,669 > 0,14$ ;  $\eta^2=0,658 > 0,14$ ). Puanların hangi gruplar arasında farklılaştığı ile ilgili Tukey testi sonuçları (Ek B) incelendiğinde alt, orta ve üst gruplarda istatistiksel olarak anlamlı fark bulunmadığı belirlenmiştir ( $p > 0,05$ ). Puanlar Şekil 2’de çizgi grafiği ile incelenmiştir.

## Şekil 2

### Eşitlenmiş Puanlar



Şekil 2 incelendiğinde genel olarak, ham puan ve eşitlenmiş puan arasında iki kalibrasyon yöntemi ve her üç yöntem içinde benzer bir desen bulunmaktadır. Ham puanlar ile eşitleme sonucu elde edilen puanlar arasındaki ilişkinin doğrusal olduğu görülmektedir. Eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamaları sonucunda elde edilen eşitlenmiş puanların BF-MIRT ve MOSE eşitleme yöntemleri için benzer olduğu görülmüştür. AOSE yönteminden elde edilen eşitlenmiş puanların ise  $x = 19$

ile  $x = 37$  ham puanları arasında diğer yöntemlerden elde edilen eşitlenmiş puanlardan farklılaştığı görülmektedir. Örneğin  $x = 33$  ham puanında AOSE yönteminden elde edilen eşitlenmiş puan (35,02) ile BF-MIRT yönteminden elde edilen eşitlenmiş puan (33,14) arasındaki fark (1,88) bir puandan fazladır. Benzer durum  $x = 46$  ile  $x = 60$  ham puanları arasında da görülmüştür.  $x = 46$  ham puanından  $x = 60$  ham puanına kadar farkların monoton arttığı belirlenmiştir. Ham puanlar 0 ila 60 aralığında değerler alırken eşitlenmiş puanlar BF-MIRT yönteminde -0,28 ila 60,03; MOSE yönteminde 1,08 ila 59,32 ve AOSE yönteminde -0,5 ila 56,21 arasında değerler almıştır.

Ayrı kalibrasyon (Stocking Lord) yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamaları sonucunda elde edilen eşitlenmiş puanların ise benzer olmadığı görülmüştür.  $x = 38$  ham puanın üst ve altındaki değerler için fark bir puandan fazladır. Farkların monoton bir şekilde uç değerlere doğru arttığı belirlenmiştir.  $x = 22$  ham puanın altındaki değerler için BF-MIRT;  $x = 22$  ham puanın üstündeki değerler için AOSE yönteminin diğer yöntemlerden farklılaştığı belirlenmiştir. Örneğin  $x = 12$  ham puanında MOSE yönteminden elde edilen eşitlenmiş puan (15,11) ile BF-MIRT yönteminden elde edilen eşitlenmiş puan (19,35) arasındaki fark (4,24) bir puandan fazladır.  $x = 58$  ham puanında MOSE yönteminden elde edilen eşitlenmiş puan (53,61) ile AOSE yönteminden elde edilen eşitlenmiş puan (56,38) arasındaki fark (2,76) bir puandan fazladır.  $x = 0$  ile  $x = 37$  ham puanları arasında en büyük değerler BF-MIRT;  $x = 38$  ile  $x = 60$  ham puanları arasında en büyük değerler AOSE yönteminde gözlenmiştir.  $x = 0$  ile  $x = 37$  ham puanları arasında en küçük değerler AOSE;  $x = 38$  ile  $x = 60$  ham puanları arasında en küçük değerler BF-MIRT yönteminde gözlenmiştir. Ham puanlar 0 ila 60 aralığında değerler alırken eşitlenmiş puanlar BF-MIRT yönteminde 8,98 ila 54,65; MOSE yönteminde 4,38 ila 57,63 ve AOSE yönteminde 0,44 ila 59,78 arasında değerler almıştır.

## İkinci Alt Probleme Ait Bulgular

**İkinci Alt Problem.** “Denk olmayan gruplarda ortak madde deseni altında tam ÇB-MTK gözlenen puan eşitleme, ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım ve bifaktör MIRT gözlenen puan eşitleme yöntemleri kullanılarak elde edilen eşitlenmiş puanlar için yanlılık (BIAS), eşitlemenin standart hatası (SEE) ve hata kareler ortalamasının karekökü (RMSE) değerleri,

a) Örneklem büyüklüğüne (500, 1500 ve 3000) göre nasıl değişmektedir?

b) Ortak madde oranına (%20, %35 ve %50) göre nasıl değişmektedir?

c) Çok boyutlu eş zamanlı kalibrasyon ve ayrı kalibrasyon (Stocking Lord) yöntemlerine göre nasıl değişmektedir?

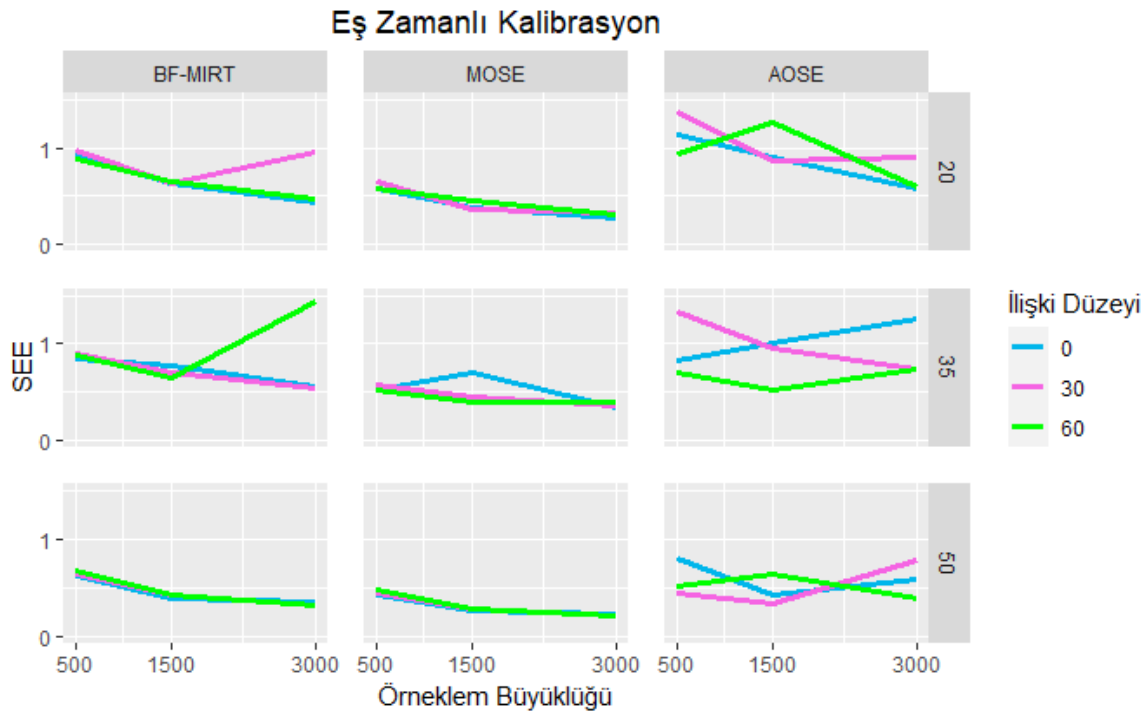
d) Boyutlar arasındaki ilişki düzeyine ( $\rho=0$ ,  $\rho=0,30$  ve  $\rho=0,60$ ) göre nasıl değişmektedir?”.

İkinci alt problem doğrultusunda yöntemlerden elde edilen hata değerlerinin değişimi çizgi grafiği ile incelenmiştir.

Eş zamanlı kalibrasyon sonucunda elde edilen SEE değerlerine ait grafikler Şekil 3'te gösterilmiştir.

### Şekil 3

*Eş Zamanlı Kalibrasyon SEE Değerleri*



Şekil 3 incelendiğinde, eş zamanlı kalibrasyon sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme uygulamalarının sonuçlarına göre, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, BF-MIRT ve MOSE yöntemleri için elde edilen SEE değerlerinin, örneklem büyüklüğü

artıkça azaldığı gözlenmiştir. AOSE yönteminde ise örneklem büyüklüğünün etkisi ortak madde oranına göre değişmektedir. Ortak madde oranı %20 olan veri setlerinde AOSE yönteminden elde edilen SEE değerlerinin örneklem büyüklüğü arttıkça azaldığı gözlenmiştir. Ortak madde oranı %35 ve %50 olan veri setlerinde örneklem büyüklüğü 500'den 1500'e çıkarıldığında SEE değerlerinin azaldığı; 1500'den 3000'e çıkarıldığında arttığı gözlenmiştir. Örneğin, ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den (SEE = 0,88) 1500'e (SEE = 0,64) çıkarıldığında 0,24; örneklem büyüklüğü 1500'den 3000'e (SEE = 0,47) çıkarıldığında 0,17 azalmıştır. Ortak madde oranı %50 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den (SEE = 0,80) 1500'e (SEE = 0,42) çıkarıldığında 0,38 azalırken; örneklem büyüklüğü 1500'den 3000'e (SEE = 0,58) çıkarıldığında 0,16 artmıştır. En küçük SEE değerleri BF-MIRT ve MOSE yöntemleri için örneklem büyüklüğü 3000 olan veri setlerinde gözlenmiştir. AOSE yönteminde ise ortak madde oranının %20 olduğu koşullarda 3000; ortak madde oranının %35 ve %50 olduğu koşullarda 1500 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü arttıkça, SEE değerlerinde en büyük değişimler AOSE; en küçük değişimler MOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, BF-MIRT ve AOSE yöntemleri için elde edilen SEE değerlerinin, ortak madde oranı arttıkça azaldığı gözlenmiştir. MOSE yönteminde ise elde edilen SEE değerlerinin, ortak madde oranı %20'den %35'e çıkarıldığında arttığı; %35'ten %50'ye çıkarıldığında azaldığı gözlenmiştir. Örneğin, örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen SEE değerleri, ortak madde oranı %20'den (SEE = 0,95) %35'e (SEE = 0,53) çıkarıldığında 0,41; ortak madde oranı %35'ten %50'ye (SEE = 0,32) çıkarıldığında 0,21 azalmıştır. Örneklem büyüklüğü 1500 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde MOSE yöntemi kullanılarak elde

edilen SEE değerleri, ortak madde oranı %20'den (SEE = 0,38) %35'e (SEE = 0,70) çıkarıldığında 0,32 artarken; ortak madde oranı %35'ten %50'ye (SEE = 0,26) çıkarıldığında 0,44 azalmıştır. En küçük SEE değerleri tüm yöntemler için ortak madde oranı %50 olan veri setlerinde gözlenmiştir. Ortak madde oranı arttıkça, SEE değerlerinde en büyük değişimler AOSE; en küçük değişimler MOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü ve ortak madde oranı sabit tutulduğunda, BF-MIRT yöntemi için elde edilen SEE değerlerinin, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında arttığı; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında azaldığı gözlenmiştir. MOSE yöntemi için elde edilen SEE değerlerinin, boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. AOSE yönteminde ise elde edilen SEE değerlerinin, boyutlar arasındaki ilişki düzeyi arttıkça azaldığı gözlenmiştir. Örneğin, örneklem büyüklüğü 500 ve ortak madde oranı %50 olan veri setlerinde MOSE yöntemi kullanılarak elde edilen SEE değerleri, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan (SEE = 0,42)  $\rho=0,30$ 'a (SEE = 0,44) çıkarıldığında 0,02; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a (SEE = 0,47) çıkarıldığında 0,03 artmıştır. Örneklem büyüklüğü 1500 ve ortak madde oranı %35 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen SEE değerleri, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan (SEE = 1)  $\rho=0,30$ 'a (SEE = 0,96) çıkarıldığında 0,04; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a (SEE = 0,52) çıkarıldığında 0,44 azalmıştır. En küçük SEE değerleri BF-MIRT ve MOSE yöntemleri için boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. AOSE yönteminde ise boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. Boyutlar arasındaki ilişki düzeyi arttıkça, SEE değerlerinde en büyük değişimler AOSE; en küçük değişimler BF-MIRT yönteminde gözlenmiştir.

Boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, tüm yöntemler için elde edilen SEE değerlerinin, örneklem büyüklüğü ve ortak madde oranı arttıkça azaldığı gözlenmiştir. Örneğin, boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında 0,42; örneklem büyüklüğü 1500'den 3000'e ve ortak



madde oranı %35'ten %50'ye çıkarıldığında 0,16 azalmıştır. En küçük SEE değerleri tüm yöntemler için örneklem büyüklüğü 3000 ve ortak madde oranı %50 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve ortak madde oranı arttıkça, SEE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler MOSE yönteminde gözlenmiştir.

Ortak madde oranı sabit tutulduğunda, tüm yöntemler için elde edilen SEE değerlerinin, örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça azaldığı gözlenmiştir. Örneğin, ortak madde oranı %50 olan veri setlerinde MOSE yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,14; örneklem büyüklüğü 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,06 azalmıştır. En küçük SEE değerleri tüm yöntemler için örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça, SEE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler MOSE yönteminde gözlenmiştir.

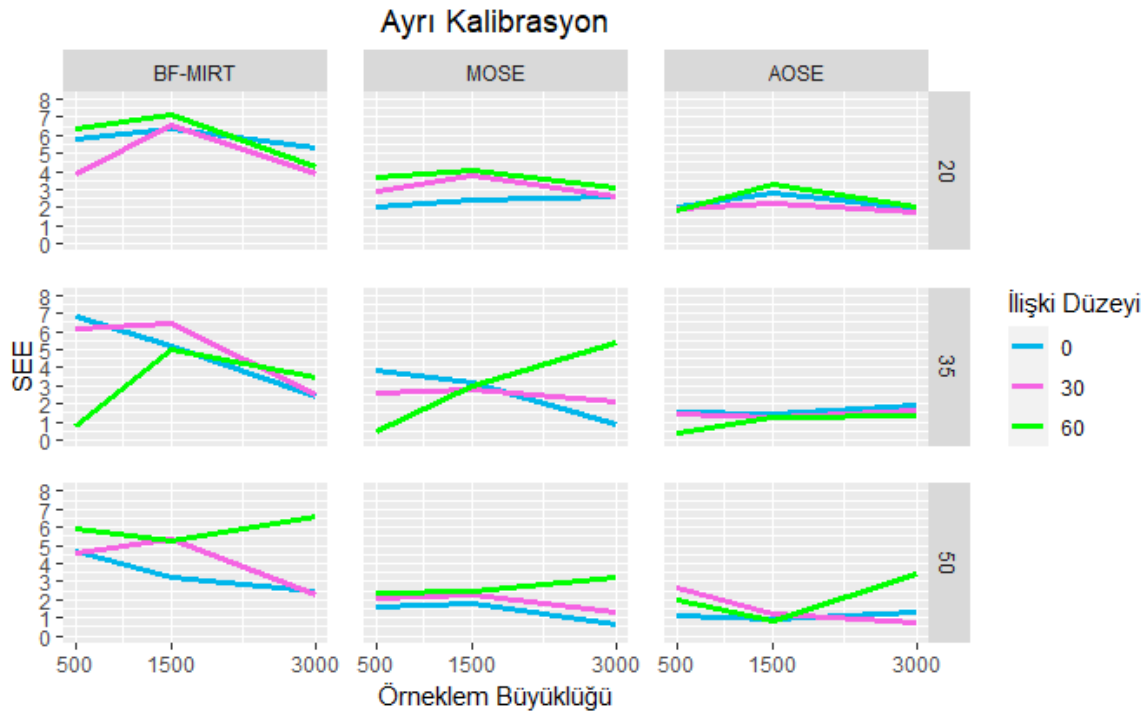
Örneklem büyüklüğü sabit tutulduğunda, tüm yöntemler için elde edilen SEE değerlerinin, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında arttığı; ortak madde oranı %35'ten %60'a ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında azaldığı gözlenmiştir. Örneğin, örneklem büyüklüğü 500 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen SEE değerleri, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,20 artarken; ortak madde oranı %35'ten %50'a ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,80 azalmıştır. En küçük SEE değerleri tüm yöntemler için ortak madde oranı %50 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. Ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça, SEE değerlerinde en büyük değişimler AOSE; en küçük değişimler MOSE yönteminde gözlenmiştir.

Tüm yöntemler için elde edilen SEE değerlerinin, örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça azaldığı gözlenmiştir. Örneğin BF-MIRT yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,24; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,37 azalmıştır. En küçük SEE değerleri tüm yöntemler için örneklem büyüklüğü 3000, ortak madde oranı %50 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça, SEE değerlerinde en büyük değişimler AOSE; en küçük değişimler MOSE yönteminde gözlenmiştir.

Ayrı kalibrasyon sonucunda elde edilen SEE değerlerine ait grafikler Şekil 4'te gösterilmiştir.

#### Şekil 4

Ayrı Kalibrasyon SEE Değerleri



Şekil 4 incelendiğinde, ayrı kalibrasyon sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme uygulamalarının sonuçlarına göre, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, BF-MIRT ve MOSE yöntemleri için elde edilen SEE değerlerinin, örneklem büyüklüğü 500'den 1500'e çıkarıldığında arttığı; 1500'den 3000'e çıkarıldığında azaldığı gözlenmiştir. AOSE yönteminde ise elde edilen SEE değerlerinin, örneklem büyüklüğü 500'den 1500'e çıkarıldığında azaldığı; 1500'den 3000'e çıkarıldığında arttığı gözlenmiştir. Örneğin, ortak madde oranı %50 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde MOSE yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den (SEE = 1,62) 1500'e (SEE = 1,75) çıkarıldığında 0,13 artarken; örneklem büyüklüğü 1500'den 3000'e (SEE = 0,66) çıkarıldığında 1.09 azalmıştır. Ortak madde oranı %50 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den (SEE = 1,14) 1500'e (SEE = 0,91) çıkarıldığında 0,23 azalırken; örneklem büyüklüğü 1500'den 3000'e (SEE = 1,27) çıkarıldığında 0,36 artmıştır. En küçük SEE değerleri BF-MIRT ve MOSE yöntemleri için örneklem büyüklüğü 3000 olan veri setlerinde gözlenmiştir. AOSE yönteminde ise örneklem büyüklüğü 1500 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü arttıkça, SEE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, tüm yöntemler için elde edilen SEE değerlerinin, ortak madde oranı arttıkça azaldığı gözlenmiştir. Örneğin, örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen SEE değerleri, ortak madde oranı %20'den (SEE = 1,71) %35'e (SEE = 1,59) çıkarıldığında 0,12; ortak madde oranı %35'ten %50'ye (SEE = 0,73) çıkarıldığında 0,86 azalmıştır. En küçük SEE değerleri tüm yöntemler için ortak madde oranı %50 olan veri setlerinde gözlenmiştir. Ortak madde oranı arttıkça, SEE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler MOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü ve ortak madde oranı sabit tutulduğunda, BF-MIRT yöntemi için elde edilen SEE değerlerinin, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında azaldığı; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında arttığı gözlenmiştir. MOSE yöntemi için elde edilen SEE değerlerinin, boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. AOSE yönteminde ise elde edilen SEE değerlerinin, boyutlar arasındaki ilişki düzeyi arttıkça azaldığı gözlenmiştir. Örneğin, örneklem büyüklüğü 500 ve ortak madde oranı %50 olan veri setlerinde MOSE yöntemi kullanılarak elde edilen SEE değerleri, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan (SEE = 1,62)  $\rho=0,30$ 'a (SEE = 2,11) çıkarıldığında 0,49; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a (SEE = 2,39) çıkarıldığında 0,28 artmıştır. Örneklem büyüklüğü 500 ve ortak madde oranı %35 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen SEE değerleri, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan (SEE = 1,50)  $\rho=0,30$ 'a (SEE = 1,43) çıkarıldığında 0,07; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a (SEE = 0,38) çıkarıldığında 1,05 azalmıştır. En küçük SEE değerleri BF-MIRT yöntemi için boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  ve MOSE yöntemi için boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. AOSE yönteminde ise boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. Boyutlar arasındaki ilişki düzeyi arttıkça, SEE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, tüm yöntemler için elde edilen SEE değerlerinin, örneklem büyüklüğü ve ortak madde oranı arttıkça azaldığı gözlenmiştir. Örneğin, boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında 0,73; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında 0,49 azalmıştır. En küçük SEE değerleri tüm yöntemler için örneklem büyüklüğü 3000 ve ortak madde oranı %50 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve ortak madde oranı arttıkça, SEE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler MOSE yönteminde gözlenmiştir.

Ortak madde oranı sabit tutulduğunda, BF-MIRT yöntemi için elde edilen SEE değerlerinin, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında arttığı; örneklem büyüklüğü 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında azaldığı gözlenmiştir. MOSE ve AOSE yöntemlerinde ise elde edilen SEE değerlerinin örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. Örneğin, ortak madde oranı %20 olan veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,76 artarken; örneklem büyüklüğü 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 1,55 azalmıştır. Ortak madde oranı %50 olan veri setlerinde MOSE yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,67; örneklem büyüklüğü 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,91 artmıştır. En küçük SEE değerleri, BF-MIRT yöntemi için örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. MOSE ve AOSE yöntemlerinde ise örneklem büyüklüğü 500 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça, SEE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü sabit tutulduğunda, BF-MIRT ve MOSE yöntemleri için elde edilen SEE değerlerinin, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında arttığı; ortak madde oranı %35'ten %60'a ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında azaldığı gözlenmiştir. AOSE yönteminde ise elde edilen SEE değerlerinin, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında azaldığı; ortak madde oranı %35'ten %60'a ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 1500 olan veri setlerinde MOSE yöntemi

kullanılarak elde edilen SEE değerleri, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,33 artarken; ortak madde oranı %35'ten %50'a ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,25 azalmıştır. Örneklem büyüklüğü 500 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen SEE değerleri, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,59 azalırken; ortak madde oranı %35'ten %50'a ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,50 artmıştır. En küçük SEE değerleri, BF-MIRT ve MOSE yöntemleri için ortak madde oranı %50 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. AOSE yönteminde ise ortak madde oranı %35 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde gözlenmiştir. Ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça, SEE değerlerinde en büyük değişimler AOSE; en küçük değişimler MOSE yönteminde gözlenmiştir.

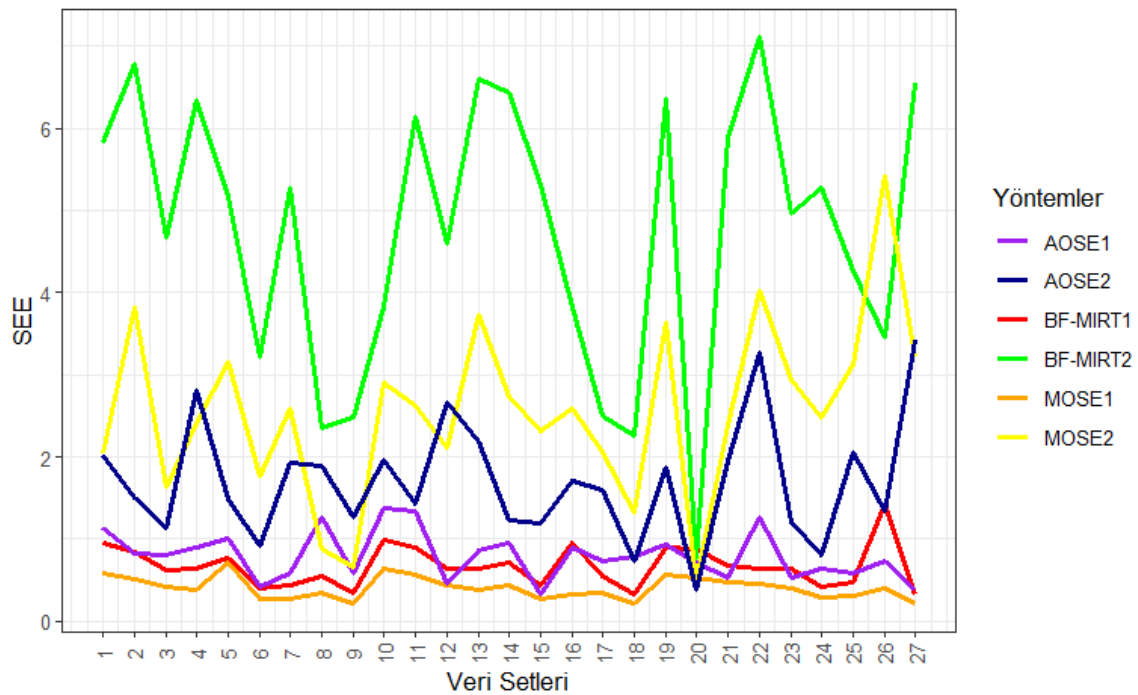
BF-MIRT ve MOSE yöntemleri için elde edilen SEE değerlerinin, örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. AOSE yönteminde ise elde edilen SEE değerlerinin, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında azaldığı; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında arttığı gözlenmiştir. Örneğin MOSE yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,70; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,48 artmıştır. AOSE yöntemi kullanılarak elde edilen SEE değerleri, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,80 azalırken; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki

düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 2,18 artmıştır. En küçük SEE değerleri BF-MIRT ve MOSE yöntemleri için örneklem büyüklüğü 500, ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. AOSE yönteminde ise örneklem büyüklüğü 1500, ortak madde oranı %35 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça, SEE değerlerinde en büyük değişimler AOSE; en küçük değişimler BF-MIRT yönteminde gözlenmiştir.

SEE değerlerinin eşitleme ve kalibrasyon yöntemlerine göre değişimi çizgi grafiği ile Şekil 5'te incelenmiştir. Bir ile belirtilen eşitleme yöntemleri (BF-MIRT1, MOSE1, AOSE1) eş zamanlı; iki ile belirtilen eşitleme yöntemleri (BF-MIRT2, MOSE2, AOSE2) ayrı (Stocking Lord) kalibrasyon sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme işleminde elde edilen SEE değerlerini her bir veri seti için göstermektedir.

## Şekil 5

SEE ve Yöntem Etkileşimi



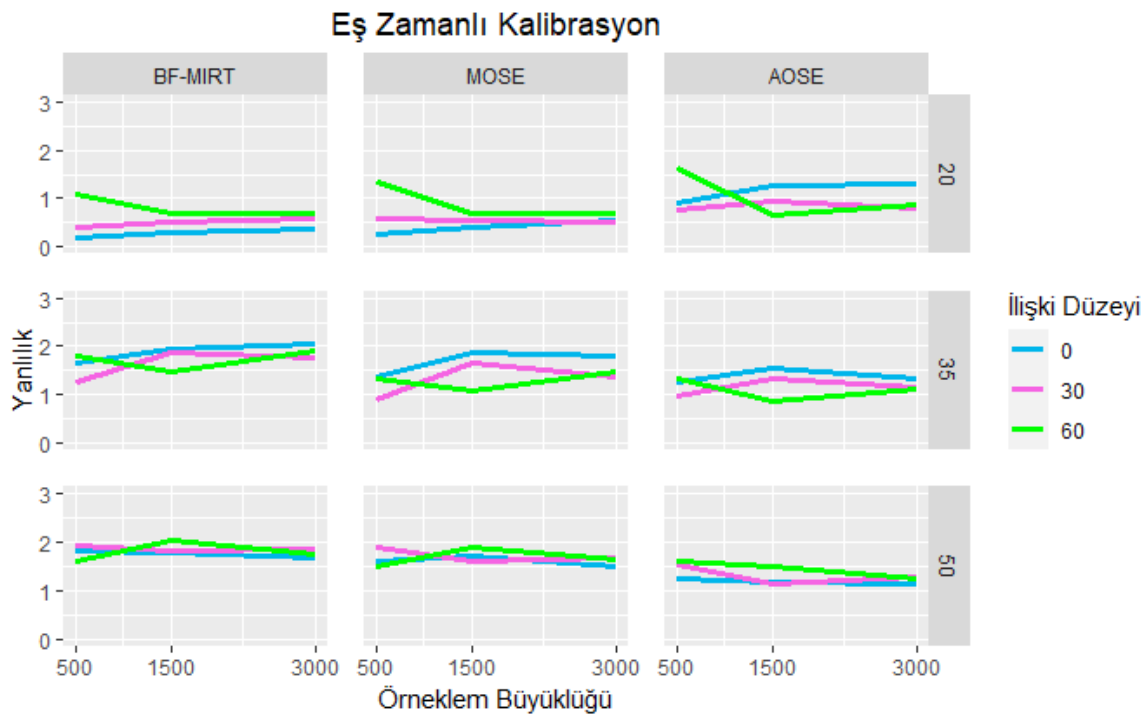
Şekil 5 incelendiğinde en büyük SEE değerleri ayrı (Stocking Lord) kalibrasyondan sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme uygulamalarında gözlenmiştir. Eş zamanlı kalibrasyon ile gerçekleştirilen test eşitleme uygulamalarında SEE değerleri daha küçüktür. Örneğin, örneklem büyüklüğü 1500, ortak madde oranı %50, boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri seti (koşul 14) için AOSE yönteminde ayrı kalibrasyon yöntemi kullanıldığında  $SEE = 0,81$ ; eş zamanlı kalibrasyon yöntemi kullanıldığında  $SEE = 0,64$  olarak gözlenmiştir.

Eş zamanlı kalibrasyon ile gerçekleştirilen test eşitleme uygulamalarında en küçük SEE değerleri MOSE; en büyük SEE değerleri AOSE yönteminde gözlenmiştir. Ayrı kalibrasyon yönteminde ise en küçük SEE değerleri AOSE; en büyük SEE değerleri BF-MIRT yönteminde gözlenmiştir.

Eş zamanlı kalibrasyon sonucunda elde edilen yanlılık değerlerine ait grafikler Şekil 6'da gösterilmiştir.

## Şekil 6

### Eş Zamanlı Kalibrasyon Yanlılık Değerleri





Şekil 6 incelendiğinde, eş zamanlı kalibrasyon sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme uygulamalarının sonuçlarına göre, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda BF-MIRT ve AOSE yöntemleri için elde edilen yanlılık değerlerine, örneklem büyüklüğünün etkisi ortak madde oranına göre değişmektedir. Ortak madde oranı %20 olan veri setlerinde BF-MIRT ve AOSE yöntemlerinden elde edilen yanlılık değerlerinin, örneklem büyüklüğü arttıkça arttığı gözlenmiştir. Ortak madde oranı %35 ve %50 olan veri setlerinde örneklem büyüklüğü arttıkça yanlılık değerlerinin azaldığı gözlenmiştir. MOSE yönteminde ise elde edilen yanlılık değerlerinin örneklem büyüklüğü 500'den 1500'e çıkarıldığında arttığı; 1500'den 3000'e çıkarıldığında azaldığı gözlenmiştir. Örneğin, ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den (Yanlılık = 0,90) 1500'e (Yanlılık = 1,27) çıkarıldığında 0,37; örneklem büyüklüğü 1500'den 3000'e (Yanlılık = 1,30) çıkarıldığında 0,03 artmıştır. Ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde MOSE yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den (Yanlılık = 0,89) 1500'e (Yanlılık = 1,63) çıkarıldığında 0,74 artarken; örneklem büyüklüğü 1500'den 3000'e (Yanlılık = 1,38) çıkarıldığında 0,25 azalmıştır. En küçük yanlılık değerleri BF-MIRT ve AOSE yöntemleri için ortak madde oranının %20 olduğu koşullarda örneklem büyüklüğü 500; ortak madde oranının %35 ve %50 olduğu koşullarda örneklem büyüklüğü 3000 olan veri setlerinde gözlenmiştir. MOSE yönteminde ise örneklem büyüklüğü 500 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü arttıkça, yanlılık değerlerinde en büyük değişimler MOSE; en küçük değişimler BF-MIRT yönteminde gözlenmiştir.

Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, tüm yöntemler için elde edilen yanlılık değerlerinin, ortak madde oranı arttıkça arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 500 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen YANLILIK değerleri, ortak

madde oranı %20'den (Yanlılık = 0,41) %35'e (Yanlılık = 1,25) çıkarıldığında 0,83; ortak madde oranı %35'ten %50'ye (Yanlılık = 1,93) çıkarıldığında 0,68 artmıştır. En küçük yanlılık değerleri tüm yöntemler için ortak madde oranı %20 olan veri setlerinde gözlenmiştir. Ortak madde oranı arttıkça, yanlılık değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Örnekleme büyüklüğü ve ortak madde oranı sabit tutulduğunda, BF-MIRT yöntemi için elde edilen yanlılık değerlerinin, boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. MOSE ve AOSE yöntemlerinde ise elde edilen yanlılık değerlerinin, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında azaldığı;  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 500 ve ortak madde oranı %20 olan veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen yanlılık değerleri, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan (Yanlılık = 0,19)  $\rho=0,30$ 'a (Yanlılık = 0,41) çıkarıldığında 0,21; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a (Yanlılık = 1,08) çıkarıldığında 0,66 artmıştır. Örneklem büyüklüğü 500 ve ortak madde oranı %35 olan veri setlerinde MOSE yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den (Yanlılık = 1,36) 1500'e (Yanlılık = 0,89) çıkarıldığında 0,46 azalırken; örneklem büyüklüğü 1500'den 3000'e (Yanlılık = 1,34) çıkarıldığında 0,44 artmıştır. En küçük yanlılık değerleri BF-MIRT yöntemi için boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. MOSE ve AOSE yöntemlerinde ise boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde gözlenmiştir. Boyutlar arasındaki ilişki düzeyi arttıkça, yanlılık değerlerinde en büyük değişimler AOSE; en küçük değişimler BF-MIRT yönteminde gözlenmiştir.

Boyutlar arasındaki ilişki düzeyi sabit tutulduğunda BF-MIRT ve AOSE yöntemleri için elde edilen yanlılık değerlerinin, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında arttığı; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında azaldığı gözlenmiştir. MOSE yönteminde ise elde edilen yanlılık değerlerinin, örneklem büyüklüğü ve ortak madde oranı arttıkça arttığı

gözlenmiştir. Örneğin, boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında 0,65 artarken; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında 0,43 azalmıştır. Boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde MOSE yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında 1,05; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'a çıkarıldığında 0,02 artmıştır. En küçük yanlılık değerleri tüm yöntemler için örneklem büyüklüğü 500 ve ortak madde oranı %20 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve ortak madde oranı arttıkça, yanlılık değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Ortak madde oranı sabit tutulduğunda, BF-MIRT ve MOSE yöntemleri için elde edilen yanlılık değerlerinin, örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. AOSE yönteminde ise elde edilen yanlılık değerlerinin, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında arttığı; 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında azaldığı gözlenmiştir. Örneğin, ortak madde oranı %20 olan veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen yanlılık değerleri örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,31; örneklem büyüklüğü 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,50 artmıştır. Ortak madde oranı %35 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,06 artarken; örneklem büyüklüğü 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,20 azalmıştır. En küçük yanlılık değerleri, BF-MIRT ve MOSE yöntemleri için örneklem büyüklüğü 500 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. AOSE yönteminde ise örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki

düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça, yanlışlık değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü sabit tutulduğunda, tüm yöntemler için elde edilen yanlışlık değerlerinin, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 500 olan veri setlerinde MOSE yöntemi kullanılarak elde edilen yanlışlık değerleri, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,65; ortak madde oranı %35'ten %50'a ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 1,01 artmıştır. En küçük yanlışlık değerleri tüm yöntemler için ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. Ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça, yanlışlık değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

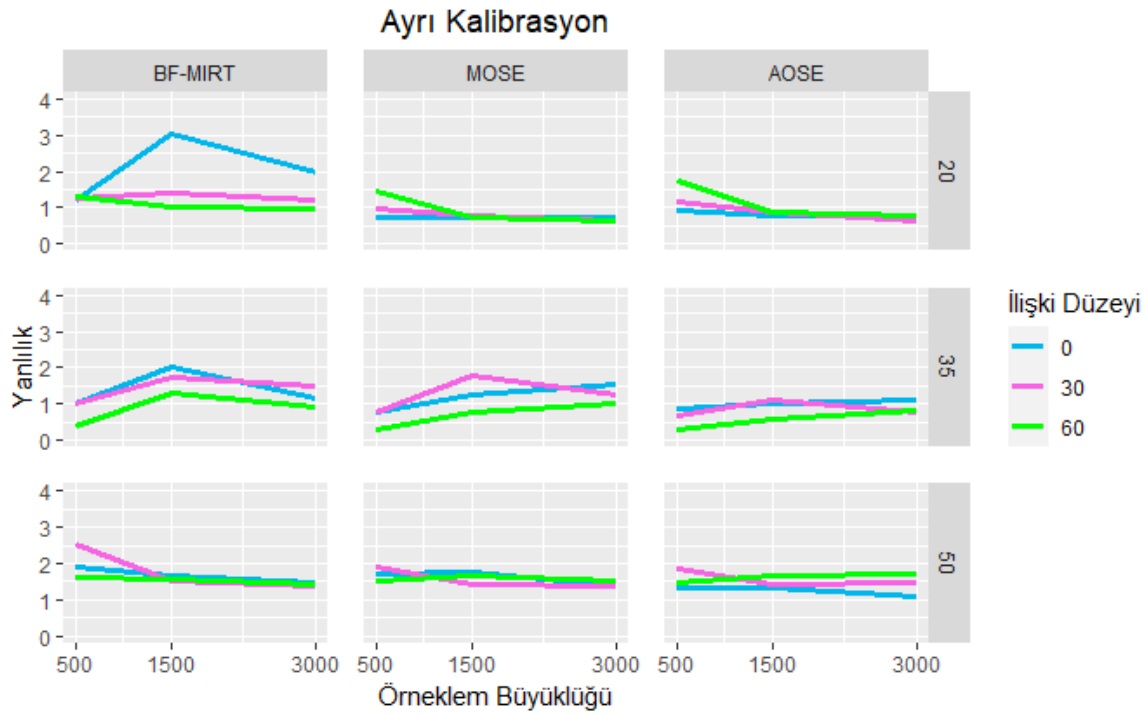
BF-MIRT ve AOSE yöntemleri için elde edilen yanlışlık değerlerinin, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında arttığı; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında azaldığı gözlenmiştir. MOSE yönteminde ise elde edilen yanlışlık değerlerinin, örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. Örneğin BF-MIRT yöntemi kullanılarak elde edilen yanlışlık değerleri, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 1,67 artarken; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,12 azalmıştır. MOSE yöntemi kullanılarak elde edilen yanlışlık değerleri, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 1,39; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki düzeyi

$\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,01 artmıştır. En küçük yanlılık değerleri tüm yöntemler için örneklem büyüklüğü 500, ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça, yanlılık değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Ayrı kalibrasyon sonucunda elde edilen yanlılık değerlerine ait grafikler Şekil 7'de gösterilmiştir.

### Şekil 7

#### Ayrı Kalibrasyon Yanlılık Değerleri



Şekil 7 incelendiğinde, ayrı kalibrasyon sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme uygulamalarının sonuçlarına göre, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, BF-MIRT ve MOSE yöntemleri için elde edilen yanlılık değerlerinin, örneklem büyüklüğü 500'den 1500'e çıkarıldığında arttığı; 1500'den 3000'e çıkarıldığında azaldığı gözlenmiştir. AOSE yönteminde ise örneklem büyüklüğünün etkisi ortak madde oranına göre değişmektedir.

Ortak madde oranı %20 olan veri setlerinde AOSE yönteminden elde edilen yanlılık değerlerinin örneklem büyüklüğü arttıkça azaldığı gözlenmiştir. Ortak madde oranı %35 ve %50 olan veri setlerinde örneklem büyüklüğü arttıkça yanlılık değerlerinin arttığı gözlenmiştir. Örneğin, ortak madde oranı %35 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den (Yanlılık = 0,37) 1500'e (Yanlılık = 1,28) çıkarıldığında 0,91 artarken; örneklem büyüklüğü 1500'den 3000'e (Yanlılık = 0,91) çıkarıldığında 0,37 azalmıştır. Ortak madde oranı %35 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den (Yanlılık = 0,29) 1500'e (Yanlılık = 0,58) çıkarıldığında 0,29; örneklem büyüklüğü 1500'den 3000'e (Yanlılık = 0,82) çıkarıldığında 0,24 artmıştır. En küçük yanlılık değerleri BF-MIRT ve MOSE yöntemleri için örneklem büyüklüğü 3000 olan veri setlerinde gözlenmiştir. AOSE yönteminde ise ortak madde oranının %20 olduğu koşullarda 3000; ortak madde oranının %35 ve %50 olduğu koşullarda 500 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü arttıkça, yanlılık değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, BF-MIRT yöntemi için elde edilen yanlılık değerlerinin, ortak madde oranı %20'den %35'e çıkarıldığında azaldığı; %35'ten %50'ye çıkarıldığında arttığı gözlenmiştir. MOSE ve AOSE yöntemlerinde ise elde edilen yanlılık değerlerinin, ortak madde oranı arttıkça arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen yanlılık değerleri, ortak madde oranı %20'den (Yanlılık = 0,98) %35'e (Yanlılık = 0,91) çıkarıldığında 0,07 azalırken; ortak madde oranı %35'ten %50'ye (Yanlılık = 1,44) çıkarıldığında 0,53 artmıştır. Örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde MOSE yöntemi kullanılarak elde edilen yanlılık değerleri, ortak madde oranı %20'den (Yanlılık = 0,62) %35'e (Yanlılık = 1,25) çıkarıldığında 0,63; ortak madde oranı %35'ten %50'ye

(Yanlılık = 1,37) çıkarıldığında 0,12 artmıştır. En küçük yanlılık değerleri BF-MIRT yöntemi için ortak madde oranı %35 olan veri setlerinde gözlenmiştir. MOSE ve AOSE yöntemlerinde ise ortak madde oranı %20 olan veri setlerinde gözlenmiştir. Ortak madde oranı arttıkça, yanlılık değerlerinde en büyük değişimler MOSE; en küçük değişimler AOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü ve ortak madde oranı sabit tutulduğunda, BF-MIRT ve MOSE yöntemleri için elde edilen yanlılık değerlerinin, boyutlar arasındaki ilişki düzeyi arttıkça azaldığı gözlenmiştir. AOSE yönteminde ise elde edilen yanlılık değerlerinin, boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 3000 ve ortak madde oranı %20 olan veri setlerinde MOSE yöntemi kullanılarak elde edilen yanlılık değerleri, boyutlar arasındaki ilişki düzeyi  $p=0$ 'dan (Yanlılık = 0,71)  $p=0,30$ 'a (Yanlılık = 0,627) çıkarıldığında 0,083; boyutlar arasındaki ilişki düzeyi  $p=0,30$ 'dan  $p=0,60$ 'a (Yanlılık = 0,625) çıkarıldığında 0,002 azalmıştır. Örneklem büyüklüğü 1500 ortak madde oranı %20 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen yanlılık değerleri, boyutlar arasındaki ilişki düzeyi  $p=0$ 'dan (Yanlılık = 0,76)  $p=0,30$ 'a (Yanlılık = 0,85) çıkarıldığında 0,09; boyutlar arasındaki ilişki düzeyi  $p=0,30$ 'dan  $p=0,60$ 'a (Yanlılık = 0,86) çıkarıldığında 0,01 artmıştır. En küçük yanlılık değerleri BF-MIRT ve MOSE yöntemleri için boyutlar arasındaki ilişki düzeyi  $p=0,60$  olan veri setlerinde gözlenmiştir. AOSE yönteminde ise boyutlar arasındaki ilişki düzeyi  $p=0$  olan veri setlerinde gözlenmiştir. Boyutlar arasındaki ilişki düzeyi arttıkça, yanlılık değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler MOSE yönteminde gözlenmiştir.

Boyutlar arasındaki ilişki düzeyi sabit tutulduğunda BF-MIRT yöntemi için elde edilen yanlılık değerlerinin, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında arttığı; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında azaldığı gözlenmiştir. MOSE yöntemi için elde edilen yanlılık değerlerinin, örneklem büyüklüğü ve ortak madde oranı arttıkça arttığı gözlenmiştir. AOSE yönteminde ise elde edilen yanlılık değerlerinin, örneklem büyüklüğü 500'den 1500'e ve

ortak madde oranı %20'den %35'e çıkarıldığında azaldığı; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında arttığı gözlenmiştir. Örneğin, boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında 0,78 artarken; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında 0,49 azalmıştır. Boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında 1,15 azalırken; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında 1,11 artmıştır. En küçük yanlılık değerleri BF-MIRT ve MOSE yöntemleri için örneklem büyüklüğü 500 ve ortak madde oranı %20 olan veri setlerinde gözlenmiştir. AOSE yönteminde ise örneklem büyüklüğü 1500 ve ortak madde oranı %35 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve ortak madde oranı arttıkça, yanlılık değerlerinde en büyük değişimler MOSE; en küçük değişimler AOSE yönteminde gözlenmiştir.

Ortak madde oranı sabit tutulduğunda, tüm yöntemler için elde edilen yanlılık değerlerinin, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında arttığı; 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında azaldığı gözlenmiştir. Örneğin, ortak madde oranı %20 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,26 artarken; örneklem büyüklüğü 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,29 azalmıştır. En küçük yanlılık değerleri tüm yöntemler için örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça, yanlılık değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler MOSE yönteminde gözlenmiştir.



Örneklem büyüklüğü sabit tutulduğunda, BF-MIRT yöntemi için elde edilen yanlılık değerlerinin, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça azaldığı gözlenmiştir. MOSE ve AOSE yöntemlerinde ise elde edilen yanlılık değerlerinin, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 1500 veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen yanlılık değerleri, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 1,30; ortak madde oranı %35'ten %50'a ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,17 azalmıştır. Örneklem büyüklüğü 1500 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen yanlılık değerleri, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,35; ortak madde oranı %35'ten %50'a ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,52 artmıştır. En küçük yanlılık değerleri BF-MIRT yöntemi için koşullarda ortak madde oranı %50 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. MOSE ve AOSE yöntemlerinde ise ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. Ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça, yanlılık değerlerinde en büyük değişimler MOSE; en küçük değişimler AOSE yönteminde gözlenmiştir.

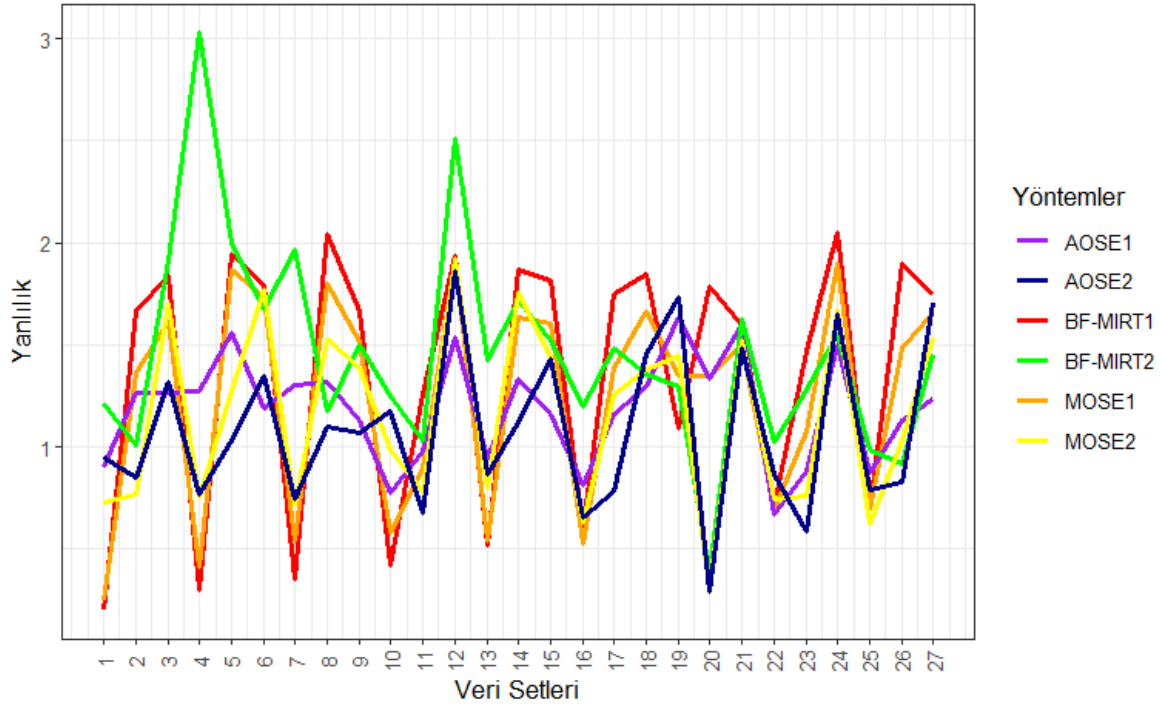
BF-MIRT ve MOSE yöntemleri için elde edilen yanlılık değerlerinin, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında arttığı; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında azaldığı gözlenmiştir. AOSE yönteminde ise elde edilen yanlılık değerlerinin, örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. Örneğin MOSE yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 1,02 artarken; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan

$\rho=0,60$ 'a çıkarıldığında 0,21 azalmıştır. AOSE yöntemi kullanılarak elde edilen yanlılık değerleri, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,17; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,57 artmıştır. En küçük yanlılık değerleri tüm yöntemler için örneklem büyüklüğü 500, ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça, yanlılık değerlerinde en büyük değişimler MOSE; en küçük değişimler BF-MIRT yönteminde gözlenmiştir.

Yanlılık değerlerinin eşitleme ve kalibrasyon yöntemlerine göre değişimi çizgi grafiği ile Şekil 8'de incelenmiştir. Bir ile belirtilen eşitleme yöntemleri (BF-MIRT1, MOSE1, AOSE1) eş zamanlı; iki ile belirtilen eşitleme yöntemleri (BF-MIRT2, MOSE2, AOSE2) ayrı (Stocking Lord) kalibrasyon sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme işleminde elde edilen yanlılık değerlerini her bir veri seti için göstermektedir.

## Şekil 8

### Yanlılık ve Yöntem Etkileşimi

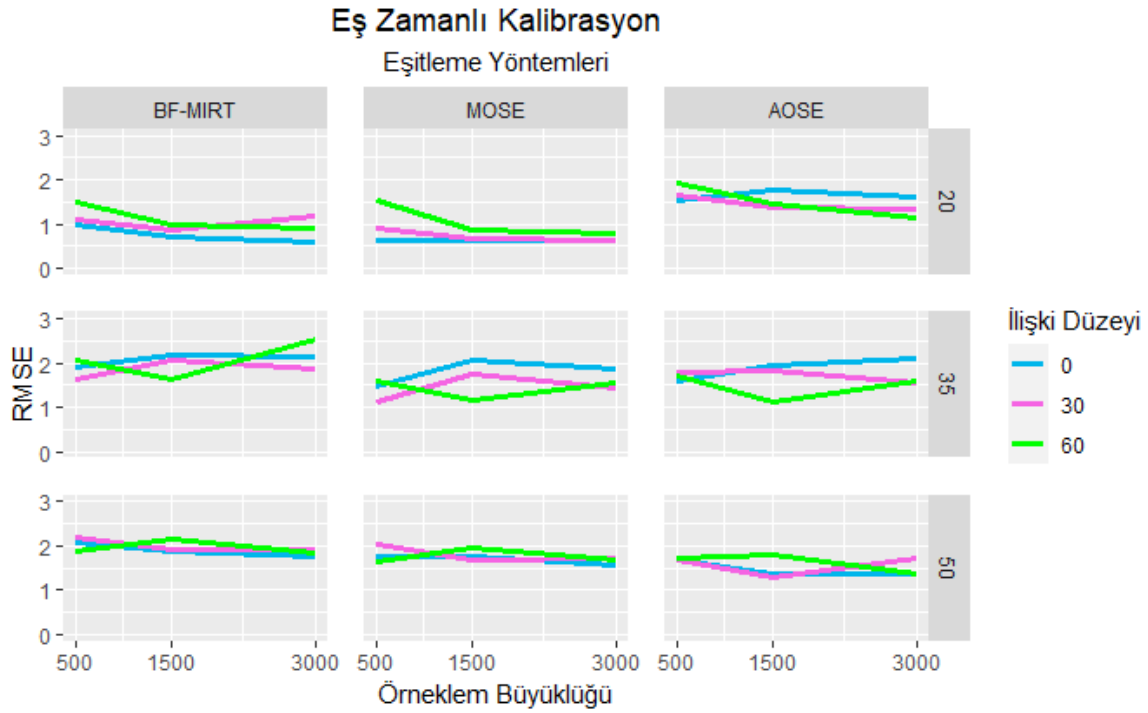


Şekil 8 incelendiğinde en yüksek yanlılık değerleri ayrı (Stocking Lord) kalibrasyondan sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme uygulamalarında gözlenmiştir. Eş zamanlı kalibrasyon ile gerçekleştirilen test eşitleme uygulamalarında yanlılık değerleri daha küçüktür. Örneğin, örneklem büyüklüğü 1500, ortak madde oranı %20, boyutlar arasındaki ilişki düzeyi  $p=0$  olan veri seti (koşul 4) için MOSE yönteminde ayrı kalibrasyon yöntemi kullanıldığında yanlılık = 0,74; eş zamanlı kalibrasyon yöntemi kullanıldığında yanlılık = 0,41 olarak gözlenmiştir. Eş zamanlı ve ayrı kalibrasyon ile gerçekleştirilen test eşitleme uygulamalarında en küçük yanlılık değerleri AOSE; en büyük yanlılık değerleri BF-MIRT yönteminde gözlenmiştir.

Eş zamanlı kalibrasyon sonucunda elde edilen RMSE değerlerine ait grafikler Şekil 9'da gösterilmiştir.

## Şekil 9

### Eş Zamanlı Kalibrasyon RMSE Değerleri



Şekil 9 incelendiğinde, eş zamanlı kalibrasyon sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme uygulamalarının sonuçlarına göre, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, tüm yöntemler için elde edilen RMSE değerlerinin, örneklem büyüklüğü arttıkça azaldığı gözlenmiştir. Örneğin, ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen RMSE değerleri, örneklem büyüklüğü 500'den (RMSE = 1,97) 1500'e (RMSE = 1,47) çıkarıldığında 0,46; örneklem büyüklüğü 1500'den 3000'e (RMSE = 1,15) çıkarıldığında 0,31 azalmıştır. En küçük RMSE değerleri tüm yöntemler için örneklem büyüklüğü 3000 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü arttıkça, RMSE değerlerinde en büyük değişimler AOSE; en küçük değişimler MOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, BF-MIRT ve AOSE yöntemleri için elde edilen RMSE değerlerinin, ortak madde oranı %20'den %35'e çıkarıldığında arttığı; %35'ten %50'ye çıkarıldığında azaldığı gözlenmiştir. MOSE

yönteminde ise elde edilen RMSE değerlerinin, ortak madde oranı arttıkça arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen RMSE değerleri, ortak madde oranı %20'den (RMSE = 0,59) %35'e (RMSE = 2,15) çıkarıldığında 1,55 artarken; ortak madde oranı %35'ten %50'ye (RMSE = 1,75) çıkarıldığında 0,40 azalmıştır. Örneklem büyüklüğü 500 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde BF-MIRT yöntemi kullanılarak elde edilen RMSE değerleri, ortak madde oranı %20'den (RMSE = 0,64) %35'e (RMSE = 1,49) çıkarıldığında 0,85; ortak madde oranı %35'ten %50'ye (RMSE = 1,75) çıkarıldığında 0,26 artmıştır. En küçük RMSE değerleri tüm yöntemler için ortak madde oranı %20 olan veri setlerinde gözlenmiştir. Ortak madde oranı arttıkça, RMSE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü ve ortak madde oranı sabit tutulduğunda, BF-MIRT ve MOSE yöntemleri için elde edilen RMSE değerlerinin, boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. AOSE yönteminde ise elde edilen RMSE değerlerinin, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında azaldığı;  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 500 ve ortak madde oranı %20 olan veri setlerinde MOSE yöntemi kullanılarak elde edilen RMSE değerleri, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan (RMSE = 0,64)  $\rho=0,30$ 'a (RMSE = 0,90) çıkarıldığında 0,26; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a (RMSE = 1,53) çıkarıldığında 0,62 artmıştır. Örneklem büyüklüğü 1500 ve ortak madde oranı %35 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen RMSE değerleri, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan (RMSE = 1,76)  $\rho=0,30$ 'a (RMSE = 1,37) çıkarıldığında 0,38 azalırken; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a (RMSE = 1,47) çıkarıldığında 0,09 artmıştır. En küçük RMSE değerleri BF-MIRT ve MOSE yöntemleri için boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. AOSE yönteminde ise boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde gözlenmiştir. Boyutlar arasındaki ilişki düzeyi arttıkça, RMSE

değerlerinde en büyük değişimler MOSE; en küçük değişimler BFMIRT yönteminde gözlenmiştir.

Boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, tüm yöntemler için elde edilen RMSE değerlerinin, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında arttığı; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında azaldığı gözlenmiştir. Örneğin, boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen RMSE değerleri, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında 0,40 artarken; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında 0,55 azalmıştır. En küçük RMSE değerleri tüm yöntemler için örneklem büyüklüğü 500 ve ortak madde oranı %20 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve ortak madde oranı arttıkça, RMSE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Ortak madde oranı sabit tutulduğunda, tüm yöntemler için elde edilen RMSE değerlerinin, örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça azaldığı gözlenmiştir. Örneğin, ortak madde oranı %20 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen RMSE değerleri, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,15; örneklem büyüklüğü 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,22 azalmıştır. En küçük RMSE değerleri tüm yöntemler için örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça, RMSE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler MOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü sabit tutulduğunda, tüm yöntemler için elde edilen RMSE değerlerinin, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 1500 olan veri setlerinde MOSE yöntemi kullanılarak elde edilen RMSE değerleri, ortak madde oranı %20'den %35'e ve boyutlar

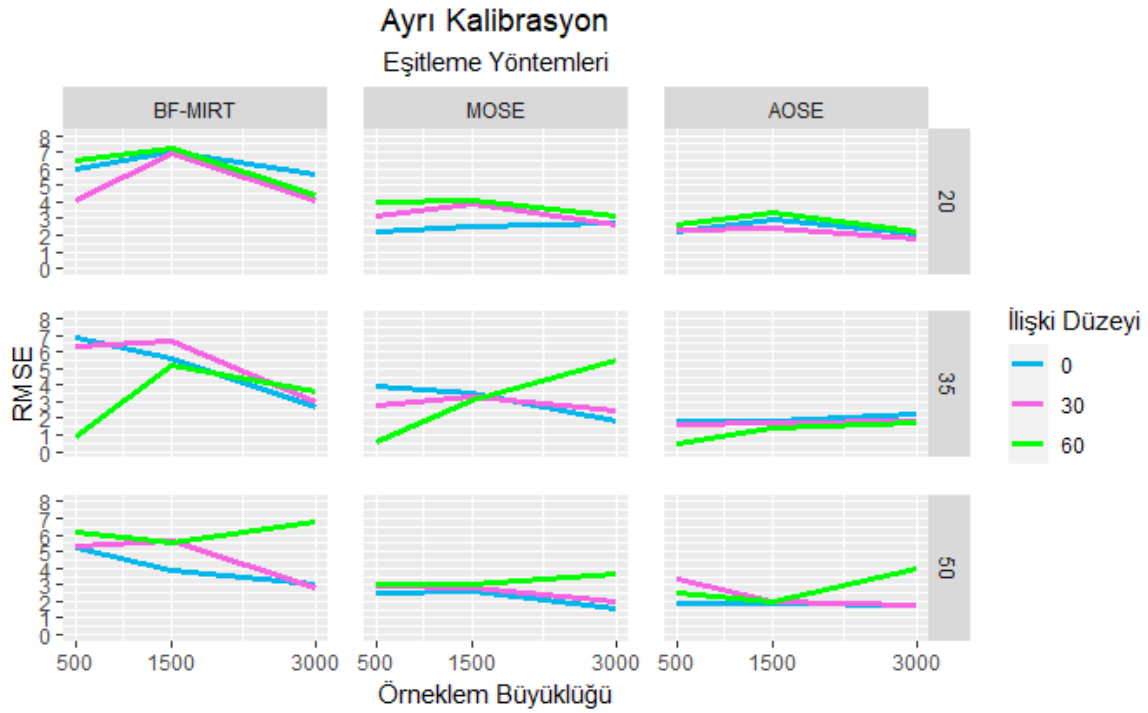
arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 1,11; ortak madde oranı %35'ten %50'a ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,22 artmıştır. En küçük RMSE değerleri tüm yöntemler için ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. Ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça, RMSE değerlerinde en büyük değişimler MOSE; en küçük değişimler AOSE yönteminde gözlenmiştir.

Tüm yöntemler için elde edilen RMSE değerlerinin, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında arttığı; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında azaldığı gözlenmiştir. Örneğin AOSE yöntemi kullanılarak elde edilen RMSE değerleri, örneklem büyüklüğü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,28 artarken; örneklem büyüklüğü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,42 azalmıştır. En küçük RMSE değerleri BF-MIRT ve MOSE yöntemleri için örneklem büyüklüğü 500, ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. AOSE yönteminde ise örneklem büyüklüğü 3000, ortak madde oranı %50 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça, RMSE değerlerinde en büyük değişimler MOSE; en küçük değişimler AOSE yönteminde gözlenmiştir.

Ayrı kalibrasyon sonucunda elde edilen RMSE değerlerine ait grafikler Şekil 10'da gösterilmiştir.

## Şekil 10

### Ayrı Kalibrasyon RMSE Değerleri



Şekil 10 incelendiğinde, ayrı kalibrasyon sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme uygulamalarının sonuçlarına göre, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, tüm yöntemler için elde edilen RMSE değerlerinin, örneklem büyüklüğü 500'den 1500'e çıkarıldığında arttığı; 1500'den 3000'e çıkarıldığında azaldığı gözlenmiştir. Örneğin, ortak madde oranı %50 ve boyutlar arasındaki ilişki düzeyi  $p=0$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen RMSE değerleri, örneklem büyüklüğü 500'den (RMSE = 1,83) 1500'e (RMSE = 1,84) çıkarıldığında 0,01 artarken; örneklem büyüklüğü 1500'den 3000'e (RMSE = 1,72) çıkarıldığında 0,12 azalmıştır. En küçük RMSE değerleri tüm yöntemler için örneklem büyüklüğü 3000 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü arttıkça, RMSE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda, BF-MIRT ve MOSE yöntemleri için elde edilen RMSE değerlerinin, ortak madde oranı arttıkça



azaldığı gözlenmiştir. AOSE yönteminde ise elde edilen RMSE değerlerinin, ortak madde oranı %20'den %35'e çıkarıldığında azaldığı; %35'ten %50'ye çıkarıldığında arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde MOSE yöntemi kullanılarak elde edilen RMSE değerleri, ortak madde oranı %20'den (RMSE = 2,70) %35'e (RMSE = 1,79) çıkarıldığında 0,91; ortak madde oranı %35'ten %50'ye (RMSE = 1,61) çıkarıldığında 0,18 azalmıştır. Örneklem büyüklüğü 500 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen RMSE değerleri, ortak madde oranı %20'den (RMSE = 2,60) %35'e (RMSE = 0,52) çıkarıldığında 2,08 azalırken; ortak madde oranı %35'ten %50'ye (RMSE = 2,51) çıkarıldığında 0,44 artmıştır. En küçük RMSE değerleri BF-MIRT ve MOSE yöntemleri için ortak madde oranı %50 olan veri setlerinde gözlenmiştir. AOSE yönteminde ise ortak madde oranı %35 olan veri setlerinde gözlenmiştir. Ortak madde oranı arttıkça, RMSE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler MOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü ve ortak madde oranı sabit tutulduğunda, BF-MIRT yöntemi için elde edilen RMSE değerlerinin, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında azaldığı; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında arttığı gözlenmiştir. MOSE yöntemi için elde edilen RMSE değerlerinin, boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. AOSE yönteminde ise elde edilen RMSE değerlerinin, boyutlar arasındaki ilişki düzeyi arttıkça azaldığı gözlenmiştir. Örneğin, örneklem büyüklüğü 3000 ve ortak madde oranı %50 olan veri setlerinde MOSE yöntemi kullanılarak elde edilen RMSE değerleri, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan (RMSE = 1,61)  $\rho=0,30$ 'a (RMSE = 1,98) çıkarıldığında 0,37; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a (RMSE = 3,63) çıkarıldığında 1,65 artmıştır. Örneklem büyüklüğü 500 ve ortak madde oranı %35 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen RMSE değerleri, boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan (RMSE = 1,83)  $\rho=0,30$ 'a (RMSE = 1,61) çıkarıldığında 0,22; boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a (RMSE =

0,52) çıkarıldığında 1,09 azalmıştır. En küçük RMSE değerleri BF-MIRT yöntemi için boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  ve MOSE yöntemi için boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. AOSE yönteminde ise boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. Boyutlar arasındaki ilişki düzeyi arttıkça, RMSE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Boyutlar arasındaki ilişki düzeyi sabit tutulduğunda BF-MIRT yöntemi için elde edilen RMSE değerlerinin, örneklem büyüklüğü ve ortak madde oranı arttıkça azaldığı gözlenmiştir. MOSE yöntemi için elde edilen RMSE değerlerinin, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında arttığı; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında azaldığı gözlenmiştir. AOSE yönteminde ise elde edilen RMSE değerlerinin, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında azaldığı; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında arttığı gözlenmiştir. Örneğin, boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde MOSE yöntemi kullanılarak elde edilen RMSE değerleri, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında 0,15 artarken; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında 1,32 azalmıştır. Boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde AOSE yöntemi kullanılarak elde edilen RMSE değerleri, örneklem büyüklüğü 500'den 1500'e ve ortak madde oranı %20'den %35'e çıkarıldığında 0,63 azalırken; örneklem büyüklüğü 1500'den 3000'e ve ortak madde oranı %35'ten %50'ye çıkarıldığında 0,04 artmıştır. En küçük RMSE değerleri BF-MIRT ve MOSE yöntemleri için örneklem büyüklüğü 3000 ve ortak madde oranı %50 olan veri setlerinde gözlenmiştir. AOSE yönteminde ise örneklem büyüklüğü 1500 ve ortak madde oranı %35 olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve ortak madde oranı arttıkça, RMSE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler MOSE yönteminde gözlenmiştir.

Ortak madde oranı sabit tutulduğunda, BF-MIRT ve AOSE yöntemleri için elde edilen RMSE değerlerinin, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında arttığı; örneklem büyüklüğü 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında azaldığı gözlenmiştir. MOSE yönteminde ise elde edilen RMSE değerlerinin örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. Örneğin, ortak madde oranı %20 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen RMSE değerleri, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,12 artarken; örneklem büyüklüğü 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,14 azalmıştır. Ortak madde oranı %50 olan veri setlerinde MOSE yöntemi kullanılarak elde edilen RMSE değerleri, örneklem büyüklüğü 500'den 1500'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında 0,34; örneklem büyüklüğü 1500'den 3000'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 0,79 artmıştır. En küçük RMSE değerleri, BF-MIRT ve AOSE yöntemleri için örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde gözlenmiştir. MOSE yönteminde ise örneklem büyüklüğü 500 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça, RMSE değerlerinde en büyük değişimler BF-MIRT; en küçük değişimler AOSE yönteminde gözlenmiştir.

Örneklem büyüklüğü sabit tutulduğunda, BF-MIRT yöntemi için elde edilen RMSE değerlerinin, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça azaldığı gözlenmiştir. MOSE yöntemi için elde edilen RMSE değerlerinin, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça arttığı gözlenmiştir. AOSE yönteminde ise elde edilen RMSE değerlerinin, ortak madde oranı %20'den %35'e ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldığında azaldığı; ortak madde oranı %35'ten %60'a ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında arttığı gözlenmiştir. Örneğin, örneklem büyüklüğü 500 olan veri setlerinde BF-MIRT yöntemi kullanılarak elde

edilen RMSE deęerleri, ortak madde oranı %20'den %35'e ve boyutlar arasındaki iliřki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldıęında 0,36; ortak madde oranı %35'ten %50'a ve boyutlar arasındaki iliřki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldıęında 1,11 azalmıřtır. Örneklem büyüklüęü 500 olan veri setlerinde AOSE yöntemi kullanılarak elde edilen SEE deęerleri, ortak madde oranı %20'den %35'e ve boyutlar arasındaki iliřki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldıęında 0,64 azalırken; ortak madde oranı %35'ten %50'a ve boyutlar arasındaki iliřki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldıęında 0,89 artmıřtır. En küçük RMSE deęerleri BF-MIRT yöntemi için ortak madde oranı %50 ve boyutlar arasındaki iliřki düzeyi  $\rho=0,60$  ve MOSE yöntemi için ortak madde oranı %35 ve boyutlar arasındaki iliřki düzeyi  $\rho=0,30$  olan veri setlerinde gözlenmiřtir. AOSE yönteminde ise ortak madde oranı %20 ve boyutlar arasındaki iliřki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiřtir. Ortak madde oranı ve boyutlar arasındaki iliřki düzeyi arttıkça, RMSE deęerlerinde en büyük deęişimler AOSE; en küçük deęişimler MOSE yönteminde gözlenmiřtir.

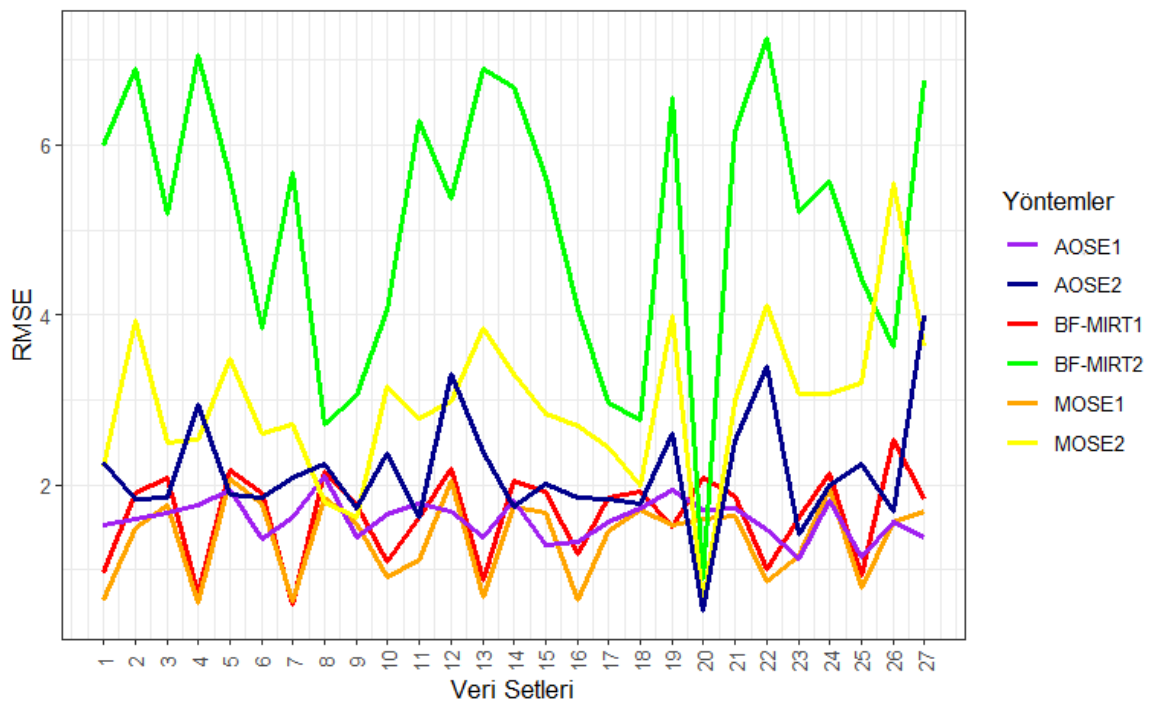
BF-MIRT ve MOSE yöntemleri için elde edilen RMSE deęerlerinin, örneklem büyüklüęü, ortak madde oranı ve boyutlar arasındaki iliřki düzeyi arttıkça arttıęı gözlenmiřtir. AOSE yönteminde ise elde edilen RMSE deęerlerinin, örneklem büyüklüęü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki iliřki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldıęında azaldıęı; örneklem büyüklüęü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki iliřki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldıęında arttıęı gözlenmiřtir. Örneęin BF-MIRT yöntemi kullanılarak elde edilen RMSE deęerleri, örneklem büyüklüęü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki iliřki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldıęında 0,69; örneklem büyüklüęü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar arasındaki iliřki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldıęında 0,05 artmıřtır. AOSE yöntemi kullanılarak elde edilen RMSE deęerleri, örneklem büyüklüęü 500'den 1500'e, ortak madde oranı %20'den %35'e ve boyutlar arasındaki iliřki düzeyi  $\rho=0$ 'dan  $\rho=0,30$ 'a çıkarıldıęında 0,52 azalırken; örneklem büyüklüęü 1500'den 3000'e, ortak madde oranı %35'ten %50'ye ve boyutlar

arasındaki ilişki düzeyi  $\rho=0,30$ 'dan  $\rho=0,60$ 'a çıkarıldığında 2,24 artmıştır. En küçük RMSE değerleri BF-MIRT ve MOSE yöntemleri için örneklem büyüklüğü 500, ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinde gözlenmiştir. AOSE yönteminde ise örneklem büyüklüğü 1500, ortak madde oranı %35 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,30$  olan veri setlerinde gözlenmiştir. Örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça, RMSE değerlerinde en büyük değişimler AOSE; en küçük değişimler BF-MIRT yönteminde gözlenmiştir.

RMSE değerlerinin eşitleme ve kalibrasyon yöntemlerine göre değişimi çizgi grafiği ile Şekil 11'de incelenmiştir. Bir ile belirtilen eşitleme yöntemleri (BF-MIRT1, MOSE1, AOSE1) eş zamanlı; iki ile belirtilen eşitleme yöntemleri (BF-MIRT2, MOSE2, AOSE2) ayrı (Stocking Lord) kalibrasyon sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme işleminde elde edilen RMSE değerlerini her bir veri seti için göstermektedir.

### Şekil 11

*RMSE ve Yöntem Etkileşimi*



Şekil 11 incelendiğinde en yüksek RMSE değerleri ayrı (Stocking Lord) kalibrasyondan sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen test eşitleme uygulamalarında gözlenmiştir. Eş zamanlı kalibrasyon ile gerçekleştirilen test eşitleme uygulamalarında RMSE değerleri daha küçüktür. Örneğin, örneklem büyüklüğü 1500, ortak madde oranı %20, boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri seti (koşul 4) için MOSE yönteminde ayrı kalibrasyon yöntemi kullanıldığında  $RMSE = 2,54$ ; eş zamanlı kalibrasyon yöntemi kullanıldığında  $RMSE = 0,61$  olarak gözlenmiştir.

Eş zamanlı kalibrasyon ile gerçekleştirilen test eşitleme uygulamalarında en küçük RMSE değerleri MOSE; en büyük RMSE değerleri BF-MIRT yönteminde gözlenmiştir. Ayrı kalibrasyon yönteminde ise en küçük RMSE değerleri AOSE; en büyük RMSE değerleri BF-MIRT yönteminde gözlenmiştir.

### **Üçüncü Alt Probleme Ait Bulgular**

**Üçüncü Alt Problem.** “Denk olmayan gruplarda ortak madde deseni altında tam ÇB-MTK gözlenen puan eşitleme, MIRT gözlenen puan eşitleme tek boyutlu yaklaşım ve bi-faktör ÇB-MTK gözlenen puan eşitleme yöntemleri kullanılarak testin kendisi ile eşitlenmesinden elde edilen eşitlenmiş puanlar farklılaşmakta mıdır?”.

Üçüncü alt problem doğrultusunda X testi Y grubu bireyelerine uygulanmıştır. Daha sonra eşitleme yapılmıştır. Yöntemlerden elde edilen eşitlenmiş puanların istatistiksel olarak anlamlı şekilde farklılaşıp farklılaşmadığını incelemek amacıyla iki yönlü ANOVA testi uygulanmıştır. İlgili analizlerin sonuçları Tablo 3’te yer almaktadır.

### **Tablo 3**

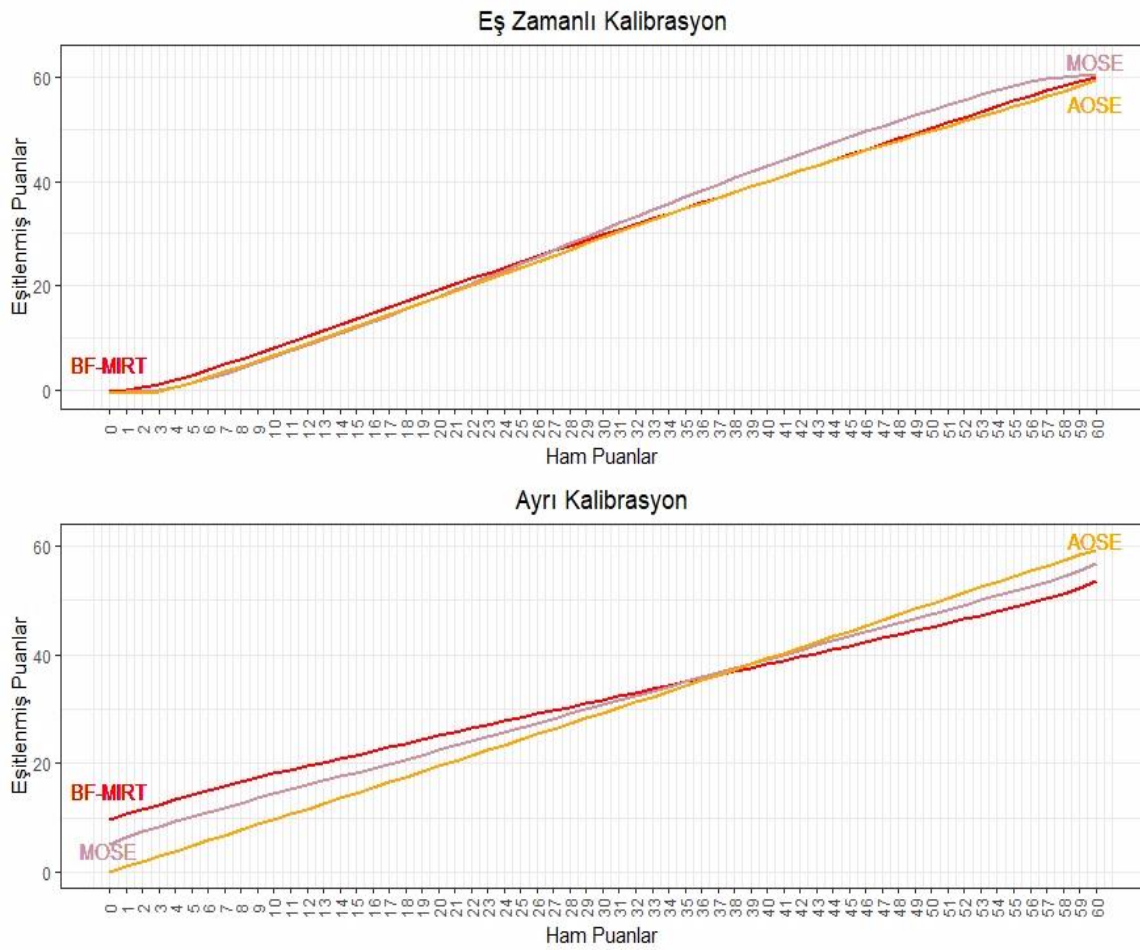
*Testin Kendisi ile Eşitlenmesi İki yönlü ANOVA Sonuçları*

		Ortalama Farkı			F	p	$\eta^2$
		BF-MIRT MOSE	BF-MIRT AOSE	MOSE AOSE			
Eş Zamanlı Kalibrasyon	Yöntem	1,237	1,1843	0,05	0,318	0,728	0,0012
	Grup	0,8622	0,7456	1,6078	175,703	<2e-16*	0,667
	Yöntem x Grup	2,4469	0,8891	3,3361	0,09	0,987	0,0006
Ayrı Kalibrasyon	Yöntem	4,0399	9,2596	5,2196	0,993	0,372	0,012
	Grup	0,8489	2,2804	1,4315	191,028	<2e-16*	0,687
	Yöntem x Grup	2,9081	5,5530	2,6449	1,902	0,112	0,041

Tablo 9 incelendiğinde eş zamanlı ve ayrı kalibrasyon sonucunda elde edilen madde ve yetenek parametreleri kullanılarak gerçekleştirilen tam ÇB-MTK gözlenen puan (MOSE), ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım (AOSE) ve bi-faktör ÇB-MTK gözlenen puan eşitleme (BF-MIRT) uygulamalarından elde edilen eşitlenmiş puanlar arasında istatistiksel olarak anlamlı fark bulunmamıştır ( $p>0,05$ ). Gruplardan elde edilen eşitlenmiş puanlar arasında ise istatistiksel olarak anlamlı fark bulunmaktadır ( $p<0,05$ ). Eta kare ( $\eta^2$ ) değeri incelendiğinde grup değişkeninin eşitlenmiş puanlar üzerinde etkisinin büyük olduğu gözlenmiştir ( $\eta^2=0,667>0,14$ ;  $\eta^2=0,687>0,14$ ). Puanların hangi gruplar arasında farklılaştığı ile ilgili Post-Hoc Tukey testi sonuçları (Ek-B) incelendiğinde alt, orta ve üst gruplarda istatistiksel olarak anlamlı fark bulunmadığı belirlenmiştir ( $p>0,05$ ). Puanlar Şekil 9'da çizgi grafiği ile incelenmiştir.

## Şekil 12

### Kendi ile Eşitlenmiş Puanlar



Şekil 12 incelendiğinde genel olarak, ham puan ve eşitlenmiş puan arasında iki kalibrasyon yöntemi ve her üç yöntem içinde benzer bir desen bulunmaktadır. Ham puanlar ile eşitleme sonucu elde edilen puanlar arasındaki ilişkinin doğrusal olduğu görülmektedir. Eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamaları sonucunda elde edilen eşitlenmiş puanların BF-MIRT ve AOSE eşitleme yöntemleri için benzer olduğu görülmüştür. MOSE yönteminden elde edilen eşitlenmiş puanların ise  $x = 31$  ile  $x = 60$  ham puanları arasında diğer yöntemlerden elde edilen eşitlenmiş puanlardan farklılaştığı görülmektedir. Örneğin  $x = 48$  ham puanında MOSE yönteminden elde edilen eşitlenmiş puan (51,60) ile BF-MIRT yönteminden elde edilen eşitlenmiş puan (48,12) arasındaki fark (3,47) bir puandan fazladır.  $x = 31$  ham puanından  $x = 48$  ham puanına kadar farkların monoton arttığı;  $x = 49$  ham puanından  $x = 60$  ham puanına kadar farkların



monoton azaldığı belirlenmiştir.  $x = 0$  ile  $x = 26$  ham puanları arasında en büyük değerler BF-MIRT;  $x = 27$  ve  $x = 60$  ham puanları arasında en büyük değerler MOSE yönteminde gözlenmiştir. Her ham puan düzeyinde en küçük değerler AOSE yönteminde gözlenmiştir. Ham puanlar 0 ila 60 aralığında değerler alırken eşitlenmiş puanlar BF-MIRT yönteminde -0,24 ila 59,83; MOSE yönteminde -0,46 ila 60,44 ve AOSE yönteminde -0,5 ila 59,44 arasında değerler almıştır.

Ayrı kalibrasyon (Stocking Lord) yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamaları sonucunda elde edilen eşitlenmiş puanların ise benzer olmadığı görülmüştür.  $x = 37$  ham puanın üst ve altındaki değerler için fark bir puandan fazladır. Farkların monoton bir şekilde uç değerlere doğru arttığı belirlenmiştir.  $x = 37$  ham puanın altındaki değerler için AOSE;  $x = 37$  ham puanın üstündeki değerler için BF-MIRT yönteminin diğer yöntemlerden farklılaştığı belirlenmiştir. Örneğin  $x = 14$  ham puanında MOSE yönteminden elde edilen eşitlenmiş puan (17,51) ile AOSE yönteminden elde edilen eşitlenmiş puan (13,53) arasındaki fark (3,98) bir puandan fazladır.  $x = 54$  ham puanında MOSE yönteminden elde edilen eşitlenmiş puan (50,76) ile BF-MIRT yönteminden elde edilen eşitlenmiş puan (47,93) arasındaki fark (2,83) bir puandan fazladır.  $x = 0$  ile  $x = 37$  ham puanları arasında en büyük değerler BF-MIRT;  $x = 38$  ile  $x = 60$  ham puanları arasında en büyük değerler AOSE yönteminde gözlenmiştir.  $x = 0$  ile  $x = 37$  ham puanları arasında en küçük değerler AOSE;  $x = 38$  ile  $x = 60$  ham puanları arasında en küçük değerler BF-MIRT yönteminde gözlenmiştir. Ham puanlar 0 ila 60 aralığında değerler alırken eşitlenmiş puanlar BF-MIRT yönteminde 9,52 ila 53,59; MOSE yönteminde 5,05 ila 56,71 ve AOSE yönteminde 0,006 ila 59,26 arasında değerler almıştır.

### **Yorumlar ve Tartışma**

Bu araştırmada, çok boyutlu testlerden elde edilen puanların eşitlenmesinde kullanılan bi-faktör ÇB-MTK gözlenen puan eşitleme (BF-MIRT), tam ÇB-MTK gözlenen puan (MOSE) ve ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım (AOSE)

yöntemlerinden denk olmayan gruplarda ortak madde deseni altında elde edilen eşitlenmiş puanlar ve bu puanlara ait eşitlemenin standart hatası (SEE), yanlılık (BIAS) ve hata kareler ortalamasının karekökü (RMSE) değerlerinin çeşitli faktörlere göre karşılaştırılması amaçlanmıştır. Sonuçların, ÇB-MTK test eşitleme yöntemlerinden hangi koşullarda daha doğru sonuçlar elde edilebileceğine dair bazı yönergeler sağlaması hedeflenmiştir.

Brossman (2010) MOSE, AOSE yöntemlerinden elde edilen eşitlenmiş puanların benzer olduğunu göstermiştir. Bu bulguyla tutarlı olarak alt, orta ve üst gruplar incelendiğinde eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamaları sonucunda elde edilen eşitlenmiş puanların benzer olduğu görülmüştür. Bu nedenle eş zamanlı kalibrasyon yöntemi kullanıldığında eşitleme uygulamalarını gerçekleştirmek için BF-MIRT, MOSE veya AOSE yöntemlerinden herhangi biri kullanılabilir. Ayrı kalibrasyon (Stocking Lord) yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamaları sonucunda elde edilen eşitlenmiş puanların ise benzer olmadığı görülmüştür. Choi (2019) ile benzer olarak ayrı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamaları sonucunda elde edilen eşitlenmiş puanlar arasındaki fark eş zamanlı kalibrasyondan daha fazladır. Özellikle uç değerlere gidildikçe farkların arttığı görülmüştür. Bunun olası nedeni bu puanların frekanslarının düşük veya sıfır olmasıdır (Kim, 2019). Çünkü eşzamanlı kalibrasyon yöntemi, ortak maddeler için madde parametrelerini kestirirken iki grubun da yanıtlarını kullanır. Ayrı kalibrasyon yöntemi ise sadece bir grubun yanıtlarını kullanarak kestirim yapar (Kim & Lee, 2018). Eş zamanlı kalibrasyonda BF-MIRT; ayrı kalibrasyonda AOSE yöntemi kullanılarak elde edilen eşitlenmiş puanlar ham puanlara daha yakındır.

Test eşitlemede doğru sonuçlar elde etmek için örneklem büyüklüğünün yeterli olması gerekmektedir. Küçük örneklerle test eşitleme uygulamalarının gerçekleştirilmesi sonucunda, oldukça büyük eşitleme hataları gözlenebilmektedir. Yapılan çalışmalar, örneklem büyüklüğündeki artışların tek boyutlu test eşitleme uygulamalarının doğruluğunu artırdığını ve eşitlemenin standart hatasını azalttığını göstermektedir (Çokluk vd., 2022; Gök

& Keleciođlu, 2014; Lee vd., 2014; Kilmen & Demirtařlı, 2012; Kolen & Brennan, 2014; Wang, 2006). Lee (2013) ok boyutlu test eřitleme yntemlerini (BF-MIRT, AOSE ve ATSE) incelediđi alıřmasında rnekleme byklđ arttıka ađırlıklandırılmıř hata deđerlerinin azaldıđını gzlemlemiřtir. Bu arařtırmada da alan yazın ile benzer olarak eř zamanlı ve ayrı kalibrasyon yntemi kullanılarak gerekleřtirilen test eřitleme uygulamaları sonucunda tm yntemler iin rnekleme byklđ arttıka hata deđerlerinin azaldıđı gzlenmiřtir.

Denk olmayan grupta ortak madde deseni kullanılan test eřitleme uygulamalarında ortak madde oranı test eřitleme sonularını etkileyen nemli bir faktrdr. Kim ve Lee (2018), ortak madde oranı arttıka hata deđerlerinin azaldıđını gzlemlemiřtir. Meng (2007), Panidvadtana vd. (2019), eř zamanlı kalibrasyon yntemi kullanılarak gerekleřtirilen test eřitleme uygulamaları sonucunda tm yntemler iin %20 ortak madde oranının daha verimli olduđunu gzlemlemiřtir. Bununla beraber ortak madde oranının arttıka hatanın azaldıđını belirlemiř ve %30 ortak madde oranının da kullanılabileceđini belirtmiřtir. Arařtırma sonularına gre literatrle benzer olarak eř zamanlı kalibrasyon ynteminde ortak madde oranı %20 olduđunda en kk hata deđerleri gzlenmiřtir. SEE deđerleri incelendiđinde ise %50 ortak madde oranının da kullanılabileceđi gzlenmiřtir. Ayrı kalibrasyon ynteminde ise ortak madde oranı arttıka hata deđerlerinin azaldıđı ve %50 ortak madde oranının daha verimli olduđu gzlenmiřtir. Ortak madde oranı arttıka, ortak madde ile tm test arasındaki iliřki daha gvenilir hale gelmektedir. Dolayısıyla daha az eřitleme hatası elde edilir (Hou, 2007).

Bi-faktr modelde boyutlar arasındaki iliřki dzeyi yani ok boyutluluk derecesi ile test eřitleme uygulamalarının sonularına etkisi iki kalibrasyon yntemi iinde incelendiđinde BF-MIRT ve MOSE yntemlerinin AOSE ynteminden farklılařtıđı gzlenmiřtir. Lee ve Lee (2016) boyutlar arasındaki iliřki dzeyi arttıka (ok boyutluluk derecesi azaldıka) BF-MIRT ynteminden elde edilen RMSD deđerlerinin arttıđını gzlemlemiřtir. Bu bulgu ile benzer olarak BF-MIRT ve MOSE yntemlerinden elde edilen hata deđerleri boyutlar arasındaki iliřki dzeyi arttıka yani ok boyutluluktan tek

boyutluluğa gidildikçe artmaktadır. Alan yazındaki çalışmalar (Brossman ve Lee,2012; Lee,2013) ve bu araştırmanın bulgularına göre boyutlar arasındaki ilişki düzeyi arttıkça yanlılık değerinde artışlar gözlenmiştir. AOSE yönteminde ise boyutlar arasındaki ilişki düzeyi arttıkça yani çok boyutluluktan tek boyutluluğa gidildikçe hata değerlerinin azaldığı gözlenmiştir. AOSE yöntemi tek boyutlu gözlenen puan eşitleme yönteminin uzantısı olduğu için çok boyutluluk derecesi azaldıkça hata değerlerinin düşmesi beklenen bir sonuçtur.

Örneklem büyüklüğü ve ortak madde oranının eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarının sonuçlarına etkisi incelendiğinde iki faktörün birlikte artmasının hata değerlerinde artışa neden olduğu gözlenmiştir. Ortak madde oranının, örneklem büyüklüğündeki artışların hata değerlerine olan etkisini sınırladığı düşünülmektedir. Ortak madde oranı ve boyutlar arasındaki ilişki düzeyi sabit tutulduğunda örneklem büyüklüğü arttıkça hata miktarı azalmaktadır. Fakat ortak madde oranındaki artışlar bu durumun tersi bir etkinin oluşmasına neden olmaktadır. Örneklem büyüklüğü 3000 ve ortak madde oranı %50 olan veri setlerinde hata değerleri örneklem büyüklüğü 3000 ve ortak madde oranı %20 olan veri setlerinden büyüktür. Bunun sonucunda örneklem büyüklüğü arttırıldığında ortak madde oranının arttırılmaması gerektiği belirlenmiştir. Bunun olası nedeninin kullanılan test eşitleme deseni olduğu düşünülmektedir. İki kalibrasyon yöntemi içinde ortak madde oranının eşitleme sonuçlarına etkisinin örneklem büyüklüğünden fazla olduğu gözlenmiştir. Ayrı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında ise bu durum gözlenmemiştir. Örneklem büyüklüğü ve ortak madde oranının birlikte artmasının hata değerlerinde azalışa neden olduğu gözlenmiştir.

Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyinin test eşitleme uygulamalarının sonuçlarına etkisi iki kalibrasyon yöntemi içinde incelendiğinde iki faktörün birlikte artmasının hata değerlerinde azalışa neden olduğu gözlenmiştir. Örneklem büyüklüğünün, boyutlar arasındaki ilişki düzeyindeki artışların hata değerlerine olan etkisini sınırladığı düşünülmektedir. Örneklem büyüklüğü ve ortak madde oranı sabit tutulduğunda

boyutlar arasındaki ilişki düzeyi arttıkça hata miktarı artmaktadır. Fakat örneklem büyüklüğündeki artışlar bu durumun tersi bir etkinin oluşmasına neden olmaktadır. Örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde hata değerleri örneklem büyüklüğü 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$  olan veri setlerinden büyüktür. Bunun sonucunda örneklem büyüklüğü artırıldığında boyutlar arasındaki ilişki düzeyinin artırılmaması gerektiği belirlenmiştir. Örneklem büyüklüğün eşitleme sonuçlarına etkisinin boyutlar arasındaki ilişki düzeyinden fazla olduğu gözlenmiştir.

Ortak madde oranı ve boyutlar arasındaki ilişki düzeyinin test eşitleme uygulamalarının sonuçlarına etkisi iki kalibrasyon yöntemi içinde incelendiğinde birbirlerinin etkilerini değiştirmedeği gözlenmiştir. Eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde en küçük hata değerleri gözlenmiştir. Ortak madde oranı azaldıkça ve boyutlar arasındaki ilişki düzeyi arttıkça hata değerleri azalmıştır. Ayrı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında ortak madde oranı %50 ve boyutlar arasındaki ilişki düzeyi  $\rho=0,60$  olan veri setlerinde en küçük hata değerleri gözlenmiştir. Ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça hata değerleri azalmıştır. Ortak madde oranı ve boyutlar arasındaki ilişki düzeyi testin yapısı ile ilgili faktörlerdir. Bu nedenle birbirlerinin etkilerini sınırlamadığı düşünülmektedir.

Örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyinin test eşitleme uygulamalarının sonuçlarına etkisi iki kalibrasyon yöntemi içinde incelendiğinde yukarıda bahsedilen etkileşimlerin değişmediği belirlenmiştir. Ortak madde oranı örneklem büyüklüğünün etkisini; örneklem büyüklüğü boyutlar arasındaki ilişki düzeyinin etkisini sınırladığı düşünülmektedir.

Eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında eşitleme sonuçları üzerinde en büyük etkiye ortak madde oranı; en küçük

etkiye boyutlar arasındaki ilişki düzeyi sahiptir. Aynı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında ise eşitleme sonuçları üzerinde en büyük etkiye örneklem büyüklüğü; en küçük etkiye boyutlar arasındaki ilişki düzeyi sahiptir. Faktörlerin eşitleme sonuçları üzerinde etkisi arttıkça birbirlerinin etkilerini sınırladığı düşünülmektedir.

Choi (2019), Kim (2017), Kim ve Lee (2018) eş zamanlı kalibrasyon yönteminin daha küçük hata değerleri ürettiğini belirtmişlerdir. İki kalibrasyon yöntemi karşılaştırıldığında, bu bulgu ile tutarlı olarak eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında, ayrı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarına göre daha küçük hata değerleri gözlenmiştir. Kalibrasyon yönteminden en çok etkilenen SEE; en az etkilenen yanlılık değerleridir. Zhang (2012) kalibrasyon yönteminin eşitleme sonuçlarına çok büyük etkisinin olduğunu belirtmiştir. Hata değerlerindeki artış incelendiğinde kalibrasyon yönteminden en çok etkilenen BF-MIRT; en az etkilenen AOSE yöntemidir. AOSE yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamaları sonucunda elde edilen hata değerleri, iki kalibrasyon yöntemi için benzer sonuçlar göstermiştir. Kalibrasyon yönteminin eşitleme sonuçlarına etkisinin yöntemlerde kullanılan veri yapısından kaynaklandığı düşünülmektedir. BF-MIRT yönteminde karışık, MOSE yönteminde basit yapı kullanılmaktadır. Madde parametreleri çok boyutlu basit yapılarda çok boyutlu karmaşık yapılardan daha düşük hata ile kestirilir (Eser, 2015). Basit yapının test eşitleme sonuçları üzerinde etkisi olmadığı da gösterilmiştir (Lee, 2013; Panidvadtana vd., 2019; Peterson, 2014; Zhang, 2012). Bu nedenle BF-MIRT yönteminin kalibrasyon yönteminden daha çok etkilendiği düşünülmektedir. AOSE yönteminde ise çok boyutlu madde parametreleri tek boyutlu madde parametrelerine dönüştürülmektedir. Tek boyutlu yapılarda çok boyutlu yapılardan daha kararlı kestirimler yapılmaktadır. AOSE yöntemi tek boyutlu yapıyı kullandığı için kalibrasyon yöntemlerinden en az etkilenen yöntem olduğu düşünülmektedir.

Lee (2013) MOSE yönteminin diğer eşitleme yöntemlerinden daha iyi performans gösterdiğini belirlemiştir. Bununla benzer olarak eş zamanlı kalibrasyon yönteminde çalışmadaki tüm eşitleme yöntemleri arasında, MOSE yöntemi çoğu zaman hata değerleri karşılaştırıldığında en iyi performansı göstermiştir. Bir başka ifadeyle en küçük hata değerleri MOSE yöntemi kullanılan test eşitleme uygulamalarında gözlenmiştir. Ayrı kalibrasyon yönteminde ise AOSE yöntemi en iyi performansı göstermiştir. Her iki kalibrasyon yöntemi içinde BF-MIRT yöntemi en kötü performansı göstermiştir. En büyük hata değerleri iki kalibrasyon yöntemi için de BF-MIRT yönteminde gözlenmiştir. İki kalibrasyon yönteminde benzer durumla karşılaşıldığı için hata değerlerinin diğer yöntemlere nazaran yüksek olmasının sebebinin kalibrasyon yönteminden bağımsız olduğu, koşullu gözlenen puan hesaplanırken kullanılan Lord-Wingersky 2.0 algoritmasından kaynaklandığı düşünülmektedir.

## Bölüm 5

### Sonuç ve Öneriler

Bu bölümde araştırmada ulaşılan sonuçlara ve bu sonuçlardan yola çıkarak sunulan önerilere yer verilmiştir.

#### Sonuçlar

Araştırmadan elde edilen sonuçlar aşağıda sıralanmıştır.

1. Eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında

- Elde edilen eşitlenmiş puanlar benzer,
- Örneklem büyüklüğü arttıkça hata değerlerinin azaldığı,
- Ortak madde oranı arttıkça hata değerlerinin arttığı,
- BF-MIRT ve MOSE yöntemlerinde boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin arttığı; AOSE yönteminde boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin azaldığı,
- Örneklem büyüklüğü ve ortak madde oranı arttıkça hata değerlerinin arttığı,
- Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin azaldığı,
- Ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin arttığı,
- Örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin arttığı,
- Eşitleme sonuçları üzerinde en büyük etkiye ortak madde oranı; en küçük etkiye boyutlar arasındaki ilişki düzeyinin sahip olduğu,
- MOSE yönteminin en az hata değerleri gözleendiği için en iyi performansı gösterdiği gözlenmiştir.



2. Ayrı kalibrasyon (Stocking Lord) yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında

- Elde edilen eşitlenmiş puanların farklılaştığı,
- Örneklem büyüklüğü arttıkça hata değerlerinin azaldığı,
- Ortak madde oranı arttıkça hata değerlerinin azaldığı,
- BF-MIRT ve MOSE yöntemlerinde boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin arttığı; AOSE yönteminde boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin azaldığı,
- Örneklem büyüklüğü ve ortak madde oranı arttıkça hata değerlerinin azaldığı,
- Örneklem büyüklüğü ve boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin azaldığı,
- Ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin azaldığı,
- Örneklem büyüklüğü, ortak madde oranı ve boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin arttığı,
- Eşitleme sonuçları üzerinde en büyük etkiye örneklem büyüklüğü; en küçük etkiye boyutlar arasındaki ilişki düzeyinin sahip olduğu,
- MOSE yönteminin en az hata değerleri gözleendiği için en iyi performansı gösterdiği gözlenmiştir.

3. İki kalibrasyon yöntemi içinde en yüksek hata değerleri BF-MIRT yönteminde gözleendiği için en kötü performansı BF-MIRT yöntemi göstermiştir.

4. Eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında ayrı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarına göre daha küçük hata değerleri gözlenmiştir.

5. Kalibrasyon yönteminden en çok etkilenen yöntem BF-MIRT, en az etkilenen AOSE yöntemidir.

## **Öneriler**

Araştırmadan elde edilen sonuçlara göre öneriler şu şekildedir.

### ***Uygulayıcılara Yönelik Öneriler***

Uygulayıcılara yönelik öneriler aşağıda sıralanmıştır.

1. Eş zamanlı kalibrasyon yöntemi kullanıldığında eşitleme uygulamalarında BF-MIRT, MOSE veya AOSE yöntemleri benzer performans gösterdiği için yöntemlerden herhangi birinden elde edilen eşitlenmiş puanlar birbirinin yerine kullanılabilir.

2. İki kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamaları için de örneklem büyüklüğü 3000 ve üzeri olmalıdır.

3. Yanlılık miktarını en aza indirebilen bir eşitleme prosedürü, SEE miktarını azaltabilen bir prosedüre tercih edilir (Kim vd., 2019). Eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında ortak madde oranı %20 olduğunda en küçük yanlılık değerleri gözlenmiştir. Ortak madde oranını en çok %20 olmalıdır.

4. Ayrı zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında ortak madde oranı %50 olduğunda en hata değerleri gözlenmiştir. Ortak madde oranı %50 ve üzeri olmalıdır.

5. Eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında boyutlar arasındaki ilişki düzeyi BF-MIRT ve MOSE yöntemleri kullanıldığında  $p=0$ 'a; AOSE yöntemi kullanıldığında  $p=0,30$ 'a yakın olmalıdır.

6. Ayrı zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında boyutlar arasındaki ilişki düzeyi BF-MIRT ve MOSE yöntemleri kullanıldığında  $\rho=0$ 'a yakın olmalıdır. AOSE yönteminde ise en az  $\rho=0,60$  olmalıdır.

7. Boyutlar arasındaki ilişki düzeyinin serbestçe tahmin edilmesi önerilmektedir. İlişki düzeyi  $\rho=0$ 'a yakın olduğunda eş zamanlı; ilişki düzeyi  $\rho=0,30$  ve üzerinde olduğunda ayrı kalibrasyon yöntemi tercih edilebilir.

8. Eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında

- Örneklem büyüklüğü ile ortak madde oranı birlikte incelendiğinde örneklem büyüklüğü en az 3000 ve ortak madde oranı en çok %20,
- Örneklem büyüklüğü ile boyutlar arasındaki ilişki düzeyi birlikte incelendiğinde örneklem büyüklüğü en az 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'a yakın,
- Ortak madde oranı ile boyutlar arasındaki ilişki düzeyi birlikte incelendiğinde ortak madde oranı en çok %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'a yakın,
- Örneklem büyüklüğü, ortak madde oranı ile boyutlar arasındaki ilişki düzeyi birlikte incelendiğinde örneklem büyüklüğü en az 3000, ortak madde oranı en çok %20 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'a yakın olmalıdır.

9. Ayrı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında

- Örneklem büyüklüğü ile ortak madde oranı birlikte incelendiğinde örneklem büyüklüğü en az 3000 ve ortak madde oranı en az %50,
- Örneklem büyüklüğü ile boyutlar arasındaki ilişki düzeyi birlikte incelendiğinde örneklem büyüklüğü en az 3000 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'a yakın,

- Ortak madde oranı ile boyutlar arasındaki ilişki düzeyi birlikte incelendiğinde ortak madde oranı en az %50 ve boyutlar arasındaki ilişki düzeyi en az  $\rho=0,60$ ,
- Örneklem büyüklüğü, ortak madde oranı ile boyutlar arasındaki ilişki düzeyi birlikte incelendiğinde örneklem büyüklüğü en az 3000, ortak madde oranı en az %50 ve boyutlar arasındaki ilişki düzeyi  $\rho=0$ 'a yakın olmalıdır.

10. Eş zamanlı kalibrasyonda ayrı kalibrasyon yönteminden daha küçük hata değerleri gözlenmiştir. Eş zamanlı kalibrasyon yöntemi tek bir analiz gerektirir. Küçük hata değerleri ve uygulama kolaylığı göz önünde bulundurulduğunda eş zamanlı kalibrasyon tercih edilebilir.

11. Eş zamanlı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında MOSE; ayrı kalibrasyon yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında AOSE yöntemi tercih edilebilir.

### ***Araştırmacılara Yönelik Öneriler***

İleride yapılacak araştırmalara yönelik öneriler aşağıda sıralanmıştır.

1. Bu çalışmada BF-MIRT, MOSE veya AOSE yöntemleri kullanılmıştır. Alan yazın incelendiğinde tüm çok boyutlu test eşitleme yöntemlerini karşılaştıran herhangi bir çalışmaya rastlanmamıştır. Çok boyutlu test eşitleme yöntemlerinin tamamı çeşitli faktörlere göre karşılaştırılabilir.

2. Çalışmada belirli koşullar altında üretilen simülasyon verileri kullanılmıştır. Gerçek veri setlerinden elde edilen madde parametreleri istenen koşullara göre manipüle edilerek çalışma tekrarlanabilir.

3. BF-MIRT yöntemi uygulanırken Lord-Wingersky 2.0 algoritması kullanılmıştır. Çalışmada Lord-Wingersky 2.0 algoritmasının BF-MIRT yöntemi kullanılarak gerçekleştirilen test eşitleme uygulamalarında elde edilen hata değerlerinin yükselmesine neden olduğu düşünülmektedir. Lord-Wingersky 2.0 algoritması yeni geliştirilen bir

yöntemdir. Algoritma veya algoritmanın farklı versiyonları kullanılarak BF-MIRT yöntemi uygulanabilir. Algoritmanın hata değerleri incelenebilir.

4. Çalışmada grupların yetenek dağılımlarının ortalaması arasındaki farkı 0,05 olarak ele alınmıştır. Grupların yetenek dağılımlarının ortalaması arasındaki farkların yöntemlere etkisi incelenebilir.

5. Ayrı kalibrasyonun yönteminin eşitleme sonuçlarına etkisi incelenirken Stocking Lord yöntemi kullanılmıştır. Ortalama-ortalama, ortalama-sigma vb. yöntemler kullanılarak ayrı kalibrasyon yöntemleri karşılaştırılabilir.

6. Boyutlar arasındaki ilişki düzeyinin maksimum değeri  $\rho=0,60$  ile sınırlandırılmıştır. Çok boyutlu test eşitleme yöntemlerinin tek boyutluluğa dayanıklılığı boyutlar arasındaki ilişki düzeyi artırılarak yöntemlerin performansları incelenebilir.

7. Her bir koşul altında her grup için cevaplar telafi edici çok boyutlu üç parametrelili lojistik (M-3PLM) madde tepki kuramı modeli kullanılarak üretilmiştir. Kısmi telafi edici çok boyutlu madde tepki kuramı modeli kullanılabilir.

8. Çalışmada veri setlerini oluşturmak için, bir genel faktör (g) ve iki spesifik faktörden (f1 ve f2) oluşan bi-faktör model kullanılmıştır. Basit ve karmaşık yapı modeli kullanılabilir.

9. Bir genel faktör (g) ve iki spesifik faktör (f1 ve f2) olmak üzere üç boyutlu bir model kullanılmıştır. Boyut sayısı artırılabilir.

10. Boyutlarda yer alan madde sayısı eşit tutulmuştur. Boyutlarda yer alan madde sayısının eşit olmadığı durumlar incelenebilir.

11. Çalışmada ikili puanlanan veri setleri üretilmiştir. Çoklu puanlanan veri setleri kullanılabilir.

12. Ayrı kalibrasyon yöntemi kullanıldığında örneklem büyüklüğü arttıkça hata değerlerinin artması gelecek çalışmalarda incelenebilir.

## Kaynaklar

- Ackerman, T.A. (1989). Unidimensional IRT calibration of compensatory and non-compensatory multidimensional items. *Applied Psychological Measurement*, 13, 113–127.
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 508–600). American Council on Education.
- Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Educational Testing Service.
- Atar, B., & Yeşiltaş, G. (2017). Çok boyutlu eşitleme yöntemlerinin eşdeğer olmayan gruplarda ortak madde deseni için performanslarının incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(4), 421-434.
- Baker, F. B., & Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. *Journal of Educational Measurement*, 28(2), 147-162.
- Bèguin, A. A., Hanson, B. A., & Glas, C. A. W. (2000, January). Effect of multidimensionality on separate and concurrent estimation in IRT equating. Presented at the National Council of Measurement in Education Meeting, New Orleans, USA.
- Bolt, D. M. (1999). Evaluating the effects of multidimensionality on IRT true-score equating. *Applied Measurement in Education*, 12(4), 383-407.
- Brossman, B. G. (2010). *Observed score and true score equating procedures for multidimensional item response theory* (Doctoral Dissertation). University of Iowa, USA.

- Brossman, B. G., & Lee, W. (2013). Observed score and true score equating procedures for multidimensional item response theory. *Applied Psychological Measurement, 37*, 460-481.
- Cai, L. (2015). Lord-Wingsky algorithm version 2.0 for hierarchical item factor models with applications in test scoring, scale alignment, and model fit testing. *Psychometrika, 80* (2), 535-559.
- Camilli, G., Wang, M., & Fesq, J. (1995). The effects of dimensionality on equating the law school admission test. *Journal of Educational Measurement, 32*(1), 79-96.
- Cattell, R. B. (1966). The scree plot test for the number of factors. *Multivariate Behavioral Research, 1*, 140-161.
- Choi, J. (2019). *Comparison of MIRT observed score equating methods under the common-item nonequivalent groups design* (Doctoral Dissertation). University of Iowa, USA.
- Cook, L. L., & Petersen, N. S. (1987). Problems related to the use of conventional and item response theory equating methods in less than optimal circumstances. *Applied Psychological Measurement, 11*, 225-244.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.
- Çokluk, Ö., Uçar, A., & Balta, E. (2022). Madde tepki kuramına dayalı gerçek puan eşitlemede ölçek dönüştürme yöntemlerinin incelenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 55*(1), 1-36.
- Davey, T., Oshima, T. C., & Lee, K. (1996). Linking multidimensional item calibrations. *Applied Psychological Measurement, 20*, 405-416.
- Davey, T., Oshima, T. C., & Lee, K. (2000). Multidimensional linking: Four practical approaches. *Journal of Educational Measurement, 37* (4), 357-373.

- Divgi, D. R. (1985). A minimum chi-square method for developing a common metric in item response theory. *Applied Psychological Measurement*, 9, 413-415.
- Dorans, N. J., & Holland P. W. (2000). Population invariance and the equatability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, 37 (4), 281-306.
- Ellis, P. D. (2010). *The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press.
- Eser, D. Ç. (2015). *Çok boyutlu madde tepki kuramının farklı modellerinden çeşitli koşullar altında kestirilen parametrelerin incelenmesi* (Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Felan, G. D. (2002, February 14-16). Test equating: Mean, linear, equipercentile, and item response theory. Presented at the Southwest Educational Research Association (SERA), Austin, TX, USA.
- Froelich, A. G. (2000). *Assessing unidimensionality of test items and some asymptotics of parametric item response theory* [Unpublished Doctoral Dissertation]. University of Illinois, Iowa, USA.
- Gao, F. (1997). *DIMTEST enhancements and some parametric irt asymptotics* [Unpublished Doctoral Dissertation]. University of Illinois, Iowa, USA.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Sheipl, F.& Hothorn, T. (2021). *mvtnorm: Multivariate Normal and t Distribution*. R package version 1.1-3, URL <http://CRAN.R-project.org/package=mvtnorm>
- Goldstein, H., & Wood, R. (1989). Five decades of item response modelling. *British Journal of Mathematical and Statistical Psychology*, 42, 139-167.
- Gorsuch, R. L. (2015). *Factor analysis*. Routledge.



- Gök, B. & Kelecioğlu, H. (2014). Comparison of IRT equating methods using the common-item nonequivalent groups design. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 10(1), 120-136.
- Güler, G. (2017). *Testin boyutluluğunun belirlenmesinde kullanılan yöntemlerde 1. tip hata ve güç oranlarının karşılaştırılması* (Doktora Tezi). Ankara Üniversitesi, Ankara.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Han, K. T. (2008). *Impact of item parameter drift on test equating and proficiency estimates* (Doctoral Dissertation). University of Massachusetts, USA.
- Hirsch, T. (1989). Multidimensional equating. *Journal of Educational Measurement*, 26, 337-349.
- Hou, J. (2007). *Effectiveness of the hybrid Levine equipercentile and modified frequency estimation equating methods under the common-item nonequivalent groups design* (Doctoral Dissertation). University of Iowa, USA.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141-151.
- Kaplan, D. (Ed.). (2004). *The SAGE handbook of quantitative methodology for the social sciences*. Sage.
- Karasar, N. (2005). *Bilimsel araştırma yöntemi*. Nobel Yayın Dağıtım.
- Kelecioğlu, H. (1994). *Öğrenci seçme sınavının eşitlenmesi üzerine bir çalışma* (Doktora Tezi). Hacettepe Üniversitesi, Ankara.

- Kilmen, S., & Demirtaşlı, N. (2012). Comparison of test equating methods based on item response theory according to the sample size and ability distribution. *Procedia-Social and Behavioral Sciences*, 46, 130-134.
- Kim, K. (2017). *IRT linking methods for the bifactor model: a special case of the two-tier item factor analysis model* (Doctoral Dissertation). The University of Iowa, USA.
- Kim, K. Y., & Lee, W. C. (2018). Linking methods for the full-information bifactor model under the common-item nonequivalent groups design. In M. J. Kolen & W. Lee (Eds.), *Mixed-format tests: Psychometric properties with a primary focus on equating (Volume 5)*. (CASMA Monograph Number 2.5) (pp. 243–261). Iowa City, IA: CASMA, The University of Iowa.
- Kim, K. Y. (2022). Item response theory true score equating for the bifactor model under the common-item nonequivalent groups design. *Applied Psychological Measurement*, 46(6), 479-493.
- Kim, S. H., & Cohen, A. S. (2002). A comparison of linking and concurrent calibration under the graded response model. *Applied Psychological Measurement*, 26(1), 25-41.
- Kim, S. Y. (2018). *Simple structure MIRT equating for multidimensional tests* (Doctoral Dissertation). University of Iowa, USA.
- Kim, S. Y., Lee, W., & Kolen, M. J. (2019). Simple-structure multidimensional item response theory equating for multidimensional test. *Educational and Psychological Measurement*, 80(1), 91-125.
- Kim, S. Y., & Lee, W. (2022). Several variations of simple-structure MIRT equating. *Journal of Educational Measurement*, 60(1), 76-105.
- Kılıç, A. F., & Uysal, İ. (2021). Faktör çıkarma yöntemlerinin paralel analiz sonuçlarına etkisi. *Trakya Eğitim Dergisi*, 11(2), 926-942.
- Kline R. B. (2005). *Principles and practice of structural equation modeling*. The Guilford Press.

- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices*. Springer Science and Business Media.
- Kumlu, G. (2019). *Test ve alt testlerde eşitlemenin farklı koşullar açısından incelenmesi* (Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Ledesma, D. R. & Mora, P. V. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research and Evaluation*, 12 (2).
- Ledesma, D. R., Mora, P. V., & Macbeth, G. (2015). The scree test and the number of factors: A dynamic graphics approach. *The Spanish Journal of Psychology*, 18(11).
- Lee, W., & Brossman, B. G. (2012). Observed score equating for mixed-format tests using a simple-structure multidimensional IRT framework. In M. J. Kolen & W. Lee (Eds.), *Mixed-format tests: Psychometric properties with a primary focus on equating* (vol 2.2, s. 115-142). Center for Advanced Studies in Measurement and Assessment.
- Lee, E. (2013). *Equating multidimensional test under a random groups design: A comparison of various equating procedures* (Doctoral Dissertation). University of Iowa, USA.
- Lee, E., Lee, W. C., & Brennan, R. L. (2014). *Equating multidimensional tests under a random groups design: A comparison of various equating procedures*. (CASMA Research Report No. 40). Center for Advanced Studies in Measurement and Assessment, The University of Iowa
- Lee, G., Lee, W., Kolen, M. J., Park, I. Y., Kim, D. I., & Yang, J. S. (2015). Bi-factor MIRT true-score equating for testlet-based tests. *Journal of Educational Evaluation*, 28, 681-700.
- Lee, G., & Lee, W. (2016). Bi-factor MIRT observed-score equating for mixed-format tests. *Applied Measurement in Education*, 29, 224-241.

- Li, H. (1995). *New nonparametric statistical procedures for analyzing BIAS/DIF and dimensionality in item response data* (Doctoral Dissertation). University of Iowa, USA.
- Li, Y. H., & Lissitz, R. W. (2000). An evaluation of the accuracy of multidimensional IRT linking. *Applied Psychological Measurement*, 24(2), 115-138.
- Livingston, S. A., Dorans, N. J., & Wright, N. K. (1990). What combination of sampling and equating methods works best?. *Applied Measurement in Education*, 3(1), 73–95.
- Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score equatings. *Applied Psychological Measurement*, 8, 452-461.
- Meng, H. (2007). *A comparison study of IRT calibration methods for mixed-format tests in vertical scaling* [Unpublished Doctoral Dissertation]. University of Iowa, Iowa, USA.
- Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 1-29.
- Orçan, F. (2018). Açıklayıcı ve doğrulayıcı faktör analizi: İlk hangisi kullanılmalı?. *Eğitim ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 9(4), 413-421.
- Öztürk Gübeş, N. (2014). *Testlerin boyutluluğunun, ortak madde formatının, yetenek dağılımının ve ölçek dönüştürme yöntemlerinin karma testlerin eşitlenmesine etkisi* (Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Pak, S., & Lee, W. C. (2014). An investigation of performance of equating for mixed-format tests using only multiple-choice common items. In M. J. Kolen & W. Lee (Eds.), *Mixed-format tests: Psychometric properties with a primary focus on equating (Volume 3)*. (CASMA Monograph Number 2.3) (pp. 7–23). Iowa City, IA: CASMA, The University of Iowa.
- Pallant, J. (2005). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows*. Australian Copyright

- Panidvadtana, P., Sujiva, S., & Srisuttiyakorn, S. (2021). A Comparison of the accuracy of multidimensional IRT equating methods for mixed-format tests. *Kasetsart Journal of Social Sciences*, 42, 215-220.
- Pekmezci, F. B. (2018). *İki faktör modelde (bifactor) diklik varsayımının farklı koşullar altında sınanması* (Doktora Tezi), Ankara Üniversitesi, Ankara.
- Peterson, N. S., Kolen, M. J., & Hoover, H. D. (1993). Scaling, norming, and equating. In Linn, R. L. (Ed.) *Educational measurement*. The Oryx.
- Peterson, J. L. (2014). *Multidimensional item response theory observed score equating methods for mixed-format tests* (Doctoral Dissertation), University of Iowa, USA.
- Price, L. R. (2017). *Psychometric methods: Theory into practice*. The Guilford Press.
- Puhan, G. (2011). Futility of Log-Linear Smoothing when Equating with Unrepresentative Small Samples. *Journal of Educational Measurement*, 48(3), 274-292.
- R Development Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Quinn, H. O. (2014). *Bifactor models, explained common variance (ecv), and the usefulness of scores from unidimensional item response theory analyses* (Master Dissertation), University of North Carolina, USA.
- Reckase, M. D. (2009). *Multidimensional item response theory*. Springer.
- Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, 47 (3), 361-372.
- Ross, A., & Willson, V. L. (2017). One-Way ANOVA. In R. Smith (Ed.), *Basic and Advanced Statistical Tests*. Brill.
- Sansivieri, V., Matteucci, M., & Mignani, S. (2018) Measurement of interrater agreement for the assessment of language proficiency. In A. Serafini S. Capecchi F. Di Iorio & R.

- Simone (Eds.), *ASMOD 2018: Proceedings of the advanced statistical modelling for ordinal data* (s. 187-194). Federico 2 University Press.
- Skaggs, G., & Lissitz, R. W. (1986). IRT test equating: Relevant issues and a review of recent research. *Review of Educational Research*, 56 (4), 495-529.
- Sönmez, V., & Alacapınar, F. G. (2019). *Örneklendirilmiş bilimsel araştırma yöntemleri*. Anı Yayıncılık.
- Spence, P. D. (1996). *The effect of multidimensionality on unidimensional equating with item response theory* [Unpublished Doctoral Dissertation]. University of Florida, USA.
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.
- Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of MULTILOG. *Applied Psychological Measurement*, 16, 1-16.
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52, 589-617.
- Svetina, D., & Levy, R. (2014). A framework for dimensionality assessment for multidimensional item response models. *Educational Assessment*, 19(1), 35-57.
- Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5). Pearson/Allyn & Bacon Inc.
- Tao, W., & Cao, Y. (2016). An extension of IRT-based equating to the dichotomous testlet response theory model. *Applied Measurement in Education*, 29, 108-121.
- Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement*, 27, 159–203.

- Thompson, T., Nering, M., & Davey, T. (1997). Multidimensional IRT scale linking. The Psychometric Society Kongresi, Gatlinburg, Amerika.
- Tindal, G., & Haladyna, T. M. (2002). *Large scale assessment programs for all students: validity, technical adequacy and implementation*. Lawrence Erlbaum Associates.
- Tinsley, H. E. A., & Brown, S. D. (Eds.). (2000). *Handbook of applied multivariate statistics and mathematical modelin*. Elsevier Science & Technology Books.
- Uğurlu, S. (2020). *Comparison of equating methods for multidimensional test which contain items contain items with differential item functioning* (Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Wang, S. & Liu, H. (2018). Minimum sample size needed for equipercentile equating under the random groups design. In M. J. Kolen ve W. Lee (Eds.), *Mixed-format tests: Psychometric properties with a primary focus on equating* (vol 2.5, s. 107-126) içinde. Center for Advanced Studies in Measurement and Assessment.
- Wang, T. (2006). *Standard errors of equating for equipercentile equating with log-linear pre-smoothing using the delta method* (CASMA Research Report, No. 14). Center for Advanced Studies in Measurement and Assessment, Iowa
- Yao, L., & Boughton, K. (2009). Multidimensional linking for tests with mixed item types. *Journal of Educational Measurement*, 46(2), 177–197.
- Zhang, J. (1996). *Some fundamental issues in item response theory with applications* [Unpublished Doctoral Dissertation]. Illinois Üniversitesi.
- Zhang, O. (2012). *Observed score and true score equating form multidimensional response theory under nonequivalent group anchor test design* (Doctoral Dissertation). University of Florida, USA.
- Zor, Y. M. (2023). Investigation of multidimensional scale transformation methods applied to multidimensional test according to various conditions. *Adiyaman University Journal of Educational Sciences*, 13(1),41-53.

**EK-A: Veri Setleri (Koşullar)**

<b>Koşul</b>	<b>Örneklem Büüklüğü</b>	<b>Madde Oranı</b>	<b>İlişki Düzeyi</b>
1	500	20	0
2	500	35	0
3	500	50	0
4	1500	20	0
5	1500	35	0
6	1500	50	0
7	3000	20	0
8	3000	35	0
9	3000	50	0
10	500	20	30
11	500	35	30
12	500	50	30
13	1500	20	30
14	1500	35	30
15	1500	50	30
16	3000	20	30
17	3000	35	30
18	3000	50	30
19	500	20	60
20	500	35	60
21	500	50	60
22	1500	20	60
23	1500	35	60
24	1500	50	60
25	3000	20	60
26	3000	35	60
27	3000	50	60



### EK-B: Tek Yönlü ANOVA Sonuçları

a) Birinci alt probleme ait Tek Yönlü ANOVA testi sonuçları

	Ortalama Farkı			F	p
	BF-MIRT MOSE	BF-MIRT AOSE	MOSE AOSE		
Eş Zamanlı Kalibrasyon	0,0467	0,1975	0,2442	0,003	0,997
Ayrı Kalibrasyon	1,0850	2,4192	1,3342	0,0359	0,699

b) Üçüncü alt probleme ait Tek Yönlü ANOVA testi sonuçları

	Ortalama Farkı			F	p
	BF-MIRT MOSE	BF-MIRT AOSE	MOSE AOSE		
Eş Zamanlı Kalibrasyon	0,7779	0,8411	1,6190	0,109	0,897
Ayrı Kalibrasyon	0,7562	2,1404	1,3842	0,3170	0,729

## EK-C: Post-Hoc Tukey Testi Sonuçları

a) Birinci alt probleme ait Tukey fark testi sonuçları

Gruplar	Eş Zamanlı		Ayrı	
	Fark	p	Fark	p
MOSE:A-BF-MIRT:A	0,4535	1	-4,6793	0,9672
AOSE:A-BF-MIRT:A	-0,3208	1	-9,1584	0,3907
BF-MIRT:O-BF-MIRT:A	25,4087	0*	19,0159	0*
MOSE:O-BF-MIRT:A	25,5767	0*	17,8729	0*
AOSE:O-BF-MIRT:A	26,0631	0*	16,3655	0*
BF-MIRT:U-BF-MIRT:A	51,1744	0*	37,1809	0*
MOSE:U-BF-MIRT:A	50,3172	0*	39,9280	0*
AOSE:U-BF-MIRT:A	47,6076	0*	42,4487	0*
AOSE:A-MOSE:A	-0,7743	1	-4,4792	1
BF-MIRT:O-MOSE:A	24,9553	0*	23,6952	0*
MOSE:O-MOSE:A	25,1232	0*	22,5521	0*
AOSE:O-MOSE:A	25,6096	0*	21,0448	0*
BF-MIRT:U-MOSE:A	50,7209	0*	41,8602	0*
MOSE:U-MOSE:A	49,8637	0*	44,6073	0*
AOSE:U-MOSE:A	47,1541	0*	47,1280	0*
BF-MIRT:O-AOSE:A	25,7295	0*	28,1744	0*
MOSE:O-AOSE:A	25,8975	0*	27,0313	0*
AOSE:O-AOSE:A	26,3839	0*	25,5240	0*
BF-MIRT:U-AOSE:A	51,4952	0*	46,3393	0*
MOSE:U-AOSE:A	50,6380	0*	49,0864	0*
AOSE:U-AOSE:A	47,9284	0*	51,6072	0*
MOSE:O-BF-MIRT:O	0,1679	1	-1,1431	0,9997
AOSE:O-BF-MIRT:O	0,6544	1	-2,6504	0,9281
BF-MIRT:U-BF-MIRT:O	25,7656	0*	18,1650	0*
MOSE:U-BF-MIRT:O	24,9085	0*	20,9121	0*
AOSE:U-BF-MIRT:O	22,1989	0*	23,4328	0*
AOSE:O-MOSE:O	0,4864	1	-1,5073	0,9981
BF-MIRT:U-MOSE:O	25,5977	0*	19,3080	0*
MOSE:U-MOSE:O	24,7405	0*	22,0551	0*
AOSE:U-MOSE:O	22,0309	0*	24,5759	0*
BF-MIRT:U-AOSE:O	25,1113	0*	20,8154	0*
MOSE:U-AOSE:O	24,2541	0*	23,5625	0*
AOSE:U-AOSE:O	21,5445	0*	26,0832	0*
MOSE:U-BF-MIRT:U	-0,8572	1	2,7471	0,9991
AOSE:U-BF-MIRT:U	-3,5668	0,9973	5,2678	0,9351
AOSE:U-MOSE:U	-2,7096	0,9996	2,5207	0,9995


Not: A: alt, O: orta, U: üst, \* istatistiksel olarak anlamlı.

## b) Üçüncü alt probleme ait Tukey fark testi sonuçları

Gruplar	Eş Zamanlı		Ayrı	
	Fark	p	Fark	p
MOSE:A-BF-MIRT:A	-1,2370	1	-4,0399	0,9787
AOSE:A-BF-MIRT:A	-1,1843	1	-9,2596	0,2725
BF-MIRT:O-BF-MIRT:A	26,7147	0*	18,0123	0*
MOSE:O-BF-MIRT:A	27,5769	0*	17,1633	0*
AOSE:O-BF-MIRT:A	25,9690	0*	15,7318	0*
BF-MIRT:U-BF-MIRT:A	52,9105	0*	35,6779	0*
MOSE:U-BF-MIRT:A	55,3574	0*	38,5861	0*
AOSE:U-BF-MIRT:A	52,0213	0*	41,2310	0*
AOSE:A-MOSE:A	0,0527	1	-5,2197	0,9068
BF-MIRT:O-MOSE:A	27,9517	0*	22,0522	0*
MOSE:O-MOSE:A	28,8139	0*	21,2033	0*
AOSE:O-MOSE:A	27,2060	0*	19,7718	0*
BF-MIRT:U-MOSE:A	54,1475	0*	39,7179	0*
MOSE:U-MOSE:A	56,5944	0*	42,6260	0*
AOSE:U-MOSE:A	53,2583	0*	45,2710	0*
BF-MIRT:O-AOSE:A	27,8990	0*	27,2719	0*
MOSE:O-AOSE:A	28,7612	0*	26,4229	0*
AOSE:O-AOSE:A	27,1533	0*	24,9914	0*
BF-MIRT:U-AOSE:A	54,0948	0*	44,9376	0*
MOSE:U-AOSE:A	56,5418	0*	47,8457	0*
AOSE:U-AOSE:A	53,2056	0*	50,4907	0*
MOSE:O-BF-MIRT:O	0,8622	1	-0,8490	1
AOSE:O-BF-MIRT:O	-0,7457	1	-2,2805	0,9526
BF-MIRT:U-BF-MIRT:O	26,1958	0*	17,6656	0*
MOSE:U-BF-MIRT:O	28,6428	0*	20,5738	0*
AOSE:U-BF-MIRT:O	25,3067	0*	23,2187	0*
AOSE:O-MOSE:O	-1,6079	0,9993	-1,4315	0,9977
BF-MIRT:U-MOSE:O	25,3336	0*	18,5146	0*
MOSE:U-MOSE:O	27,7806	0*	21,4228	0*
AOSE:U-MOSE:O	24,4444	0*	24,0677	0*
BF-MIRT:U-AOSE:O	26,9415	0*	19,9461	0*
MOSE:U-AOSE:O	29,3884	0*	22,8543	0*
AOSE:U-AOSE:O	26,0523	0*	25,4992	0*
MOSE:U-BF-MIRT:U	2,4469	0,9999	2,9082	0,9976
AOSE:U-BF-MIRT:U	-0,8892	1	5,5531	0,8723
AOSE:U-MOSE:U	-3,3361	0,9991	2,6449	0,9988

Not: A: alt, O: orta, U: üst, \* istatistiksel olarak anlamlı.

## EK-Ç: Araştırma Etik Komisyon İzin Muafiyeti Formu

	<b>Hacettepe Üniversitesi</b> <b>Eğitim Bilimleri Enstitüsü</b> <b>Tez Çalışması/Araştırma Etik Komisyon İzin Muafiyeti Formu</b>	<b>F46</b>
03 / 04 / 2023		
Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitimde Bilimleri Ana Bilim Dalı Başkanlığına		
<b>Tez/Araştırma Başlığı</b>	<b>ÇOK BOYUTLU MADDE TEPKİ KURAMI TEST EŞİTLEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI</b>	
Yukarıda başlığı/konusu verilen tez/araştırma çalışmam,		
<ol style="list-style-type: none"> <li>1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.</li> <li>2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.</li> <li>3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir.</li> <li>4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırmalar niteliğinde değildir.</li> <li>5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.</li> </ol>		
Çalışmada kullanacağım veriler: <input type="checkbox"/> Kamusal erişime açık (buraya yazınız): ..... <input type="checkbox"/> Özel izin ve onaya tabi (buraya yazınız): ..... <input checked="" type="checkbox"/> Üretilmiş veri (buraya yazınız): Çalışmada simülasyon koşullarına göre veri üretimi gerçekleştirilecektir. <input type="checkbox"/> Diğer (buraya yazınız): .....		
Yükseköğretim Kurumları Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.		
Gereğini saygılarımla arz ederim.		
<b>Burcu DEMİRÖZ</b> <small>(Araştırmacı Adı Soyadı, İmzası)</small>		
<b>Araştırmacı Bilgileri</b>		
<b>Adı Soyadı</b>	<b>Burcu DEMİRÖZ</b>	
<b>Öğrenci ise No</b>	N20146525	
<b>Ana Bilim Dalı</b>	Eğitim Bilimleri	
<b>Programı</b>	Eğitimde Ölçme ve Değerlendirme	
<b>Statüsü</b>	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr. <input type="checkbox"/> Diğer	
<b>Danışman Görüşü ve Onayı*</b>		
<b>Prof. Dr. Nuri DOĞAN</b> <small>(İmza)</small> <small>(Danışmanın Ünvanı, Adı ve Soyadı)</small>		
<small>*Tez ve tezden üretilen yayınlarda gerekli</small>		
<small>Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü. Beytepe Yerleşkesi, 06800, Çankaya / ANKARA          Telefon: 0(312) 297 85 72 Belgegeçer: 0(312) 297 85 66 e-Ağ: <a href="http://ebe.hacettepe.edu.tr/">http://ebe.hacettepe.edu.tr/</a> e-Posta: <a href="mailto:ebe@hacettepe.edu.tr">ebe@hacettepe.edu.tr</a></small>		

**EK-D: Etik Beyanı**

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında,

- \* tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- \* görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- \* başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- \* atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- \* kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- \* bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

25/07/2024

Burcu DEMİRÖZ

**EK-E: Doktora Tez Çalışması Orijinallik Raporu**

16/07/2024

HACETTEPE ÜNİVERSİTESİ  
Eğitim Bilimleri Enstitüsü  
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı: Çok Boyutlu Madde Tepki Kuramı Test Eşitleme Yöntemlerinin Karşılaştırılması

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
27/05/2024	140	174107	25/06/2024.	%7	2389175254

Uygulanan filtreler:

- Kaynaklar hariç
- Alıntılar dâhil
- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esaslarını inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

**Ad Soyadı:** Burcu DEMİRÖZ

**Öğrenci No.:** N20146525

**Ana Bilim Dalı:** Eğitim Bilimleri

**Programı:** Eğitimde Ölçme ve Değerlendirme

**Statüsü:**  Y.Lisans  Doktora  Bütünleşik Dr.

**DANIŞMAN ONAYI**

UYGUNDUR.

Prof. Dr. Nuri DOĞAN

## EK-F: Thesis/Dissertation Originality Report

16/07/2024

HACETTEPE UNIVERSITY  
Graduate School of Educational Sciences  
To The Department of Educational Sciences

Thesis Title: Comprasion of Multidimensional Item Response Theory Test Equating Methods

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
27/05/2024	140	174107	25/06/2024.	%7	2389175254

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

**Name Lastname:** Burcu DEMİRÖZ  
**Student No.:** N20146525  
**Department:** Educational Sciences  
**Program:** Measurement and Evaluation in Education  
**Status:**  Masters  Ph.D.  Integrated Ph.D.

### ADVISOR APPROVAL

APPROVED

Prof. Dr. Nuri DOĞAN

## EK-G: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. <sup>(1)</sup>
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren ... ay ertelenmiştir. <sup>(2)</sup>
- Tezimle ilgili gizlilik kararı verilmiştir. <sup>(3)</sup>

..... / ..... / .....

Burcu DEMİRÖZ

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezinerişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6.2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internette paylaşılması durumunda 3 şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir\*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.  
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir  
\*Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar ver



