ÇOK BOYUTLU TEST DESENİNİN VE KALİBRASYON YÖNTEMLERİNİN ÇOK BOYUTLU BİREYSELLEŞTİRİLMİŞ BİLGİSAYAR UYGULAMALARINA ETKİSİ

THE EFFECT OF MULTIDIMENSIONAL TEST DESIGN STRUCTURE AND CALIBRATION STRATEGIES IN MULTIDIMENSIONAL COMPUTER ADAPTIVE TESTING

Eren Halil ÖZBERK

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı İçin Öngördüğü

Doktora Tezi

olarak hazırlanmıştır.

2016

Eğitim Bilimleri Enstitüsü Müdürlüğü'ne,

Eren Halil Özberk'in hazırladığı "Çok Boyutlu Test Deseninin ve Kalibrasyon Yöntemlerinin Çok Boyutlu Bireyselleştirilmiş Bilgisayar Uygulamalarına Etkisi" başlıklı bu çalışma jürimiz tarafından **Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı'nda Doktora Tezi** olarak kabul edilmiştir.

Başkan	Prof. Dr. Şener BÜYÜKÖZTÜRK	.....
Üye (Danışman)	Prof. Dr. Selahattin GELBAL	.....
Üye	Prof. Dr. Hülya KELECİOĞLU	.....
Üye	Prof. Dr. Nuri DOĞAN	.....
Üye	Doç. Dr. Nilüfer KAHRAMAN	.....

ONAY

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından 19 / 12 / 2016 tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca ..... / ..... / ..... tarihinde kabul edilmiştir.

Prof. Dr. Ali Ekber ŞAHİN
Eğitim Bilimleri Enstitüsü Müdürü

ÇOK BOYUTLU TEST DESENİNİN VE KALİBRASYON YÖNTEMLERİNİN ÇOK BOYUTLU BİREYSELLEŞTİRİLMİŞ BİLGİSAYAR UYGULAMALARINA ETKİSİ

Eren Halil ÖZBERK

ÖZ

Testler kimi zaman bireyleri sıralamak kimi zaman da bireyler hakkında tanılayıcı değerlendirmeler yapmak gibi farklı amaçlar doğrultusunda geliştirilebilirler. Ancak son zamanlarda testlerin istenmeyen boyutları ölçtüğü veya çok boyutlu yapılarda geliştirildiği durumlarda alt boyut puanlarının rapor edilmesine yönelik problemler ortaya çıkmaktadır. Çok boyutlu bireyselleştirilmiş bilgisayarlı test (ÇBBBT) uygulamaları, çok boyutlu madde tepki kuramı (ÇBMTK) varsayımlarını kullanarak çok sayıda boyutu etkili bir şekilde ölçebilmektedir. Literatürde alt boyut ve toplam yetenek puanları kestirimlerinin kesinliklerini artırmaya yönelik çok sayıda ÇBBBT madde seçme yöntemi çalışması bulunmaktadır. Literatürde farkı koşul altında madde seçme yöntemlerinin karşılaştırılmasına ilişkin araştırmalar yer almakta fakat testin yapısına ve madde bankasının kalibrasyonlarına göre herhangi bir karşılaştırılmaya rastlanmamaktadır. Bu çalışmada, önceki çalışmalardan farklı olarak, tek ve çok boyutlu kalibrasyon yöntemleri ile çok boyutlu test yapısı (basit ve karmaşık) gerçek test koşullarına en uygun olacak şekilde karşılaştırılmıştır. Bu çalışmanın amacı ÇBBBT uygulamalarında alt boyut ve toplam yetenek puanlarını kestirirken kullanılan çok boyutlu madde seçme yöntemlerinin hata, yanlılık ve korelasyon değerlerini test desenine, kalibrasyon yöntemine ve boyut başına düşen madde sayısına göre karşılaştırmaktır.

Araştırmada test deseni, boyut başına düşen madde sayısı, kalibrasyon yöntemi ve madde seçme yöntemi olmak üzere dört koşul manipüle edilmiştir. Basit, düşük karmaşık ve yüksek karmaşık test yapıları için 1000x3 ve 1000x45 matrisleri gerçek yetenek parametreleri çok değişkenli normal dağılımdan rastgele olarak üretilmiştir. Üretilen madde ve yetenek parametreleri kullanılarak telafi edici çok boyutlu 3 parametreli lojistik modele ve belirlenen korelasyon değerlerine göre ikili puanlanan madde yanıt setleri oluşturulmuştur. 3 boyutlu madde bankası basit ve karmaşık test desenleri için simüle edilmiştir. Araştırmada boyutlar arasındaki korelasyon değerleri 0.2, 0.5 ve 0.8; kalibrasyon yöntemleri olarak tek boyutlu ayrı kalibrasyon ile iki farklı çok boyutlu (Bock and Aitkin EM ve Metropolis-Hastings Robbins-Monro algoritmaları) ile minimum açı, Kullback-Leibler (KL) ve minimum

hata varyansı madde seçme yöntemleri kullanılmıştır. Alt boyut ve toplam yetenek puanlarının performansları gerçek ve ÇBBBT sonrası kestirilen puanlar arasındaki mutlak yanlılık (ABSBIAS), hata kareleri ortalamasının karekökü (RMSE) ve korelasyona değerlerine göre karşılaştırılmıştır.

Araştırmanın sonuçları incelendiğinde kalibrasyon yöntemlerinin, test deseninin ve boyut başına düşen madde sayısının alt boyut ve toplam yetenek puanlarının kestiriminde madde seçme yöntemlerinin performansına anlamlı bir etkisinin olduğu belirlenmiştir. Çok boyutlu test modelinin karmaşıklaştığı durumda hem alt boyut hem de toplam yetenek puanları kestirimlerinde mutlak yanlılık değerlerinin anlamlı bir biçimce azaldığı görülmüştür.

Farklı test desenlerinde farklı madde seçme yöntemlerinin daha iyi performans gösterdiği belirlenmiştir. Basit yapılı test deseninde boyutlar arası korelasyonun orta düzeyde (0.5) olduğu uzun testlerde (N=45), toplam puanların kestiriminde tek boyutlu ve BAEM kalibrasyonu için minimum hata varyansı (V1) madde seçme yönteminin en düşük mutlak yanlılık değerlerine sahip olduğu belirlenmiştir. Düşük karmaşık yapılı test deseninde boyutlar arası korelasyonun yüksek düzeyde (0.8) olduğu uzun testlerde (N=45), toplam puanların kestiriminde tek boyutlu kalibrasyon için Vol madde seçme yönteminin en düşük mutlak yanlılık değerlerine sahip olduğu belirlenmiştir. BAEM kalibrasyonunda ise en düşük mutlak yanlılık değeri boyutlar arası korelasyonun düşük düzeyde (0.2) olduğu uzun testlerde (N=45) Vol madde seçme yönteminde gözlenmiştir. Yüksek karmaşık yapılı test deseninde boyutlar arası korelasyonun yüksek düzeyde (0.8) olduğu uzun testlerde (N=45), toplam puanların kestiriminde tek boyutlu kalibrasyon için V1 madde seçme yönteminin en düşük mutlak yanlılık değerlerine sahip olduğu belirlenmiştir. BAEM kalibrasyonunda ise en düşük mutlak yanlılık değeri boyutlar arası korelasyonun düşük düzeyde (0.2) olduğu uzun testlerde (N=45) KL madde seçme yönteminde gözlenmiştir.

Anahtar sözcükler: Çok boyutlu madde tepki kuramı, çok boyutlu bireyselleştirilmiş bilgisayar uygulamaları, alt boyut ve toplam yetenek puanları, madde seçme yöntemleri, tek ve çok boyutlu kalibrasyon

Danışman: Prof. Dr. Selahattin GELBAL, Hacettepe Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı

THE EFFECT OF MULTIDIMENSIONAL TEST DESIGN STRUCTURE AND CALIBRATION STRATEGIES IN MULTIDIMENSIONAL COMPUTER ADAPTIVE TESTING

Eren Halil ÖZBERK

ABSTRACT

A test can be designed for many purposes, including the ranking of people along a continuum or providing diagnostic value about examinees. However, a very common problem that often arises is the reporting diagnostic subscores when items are capable of measuring unwanted dimensions or designed for multidimensional purposes. Multidimensional computer adaptive testing (MCAT) is capable of measuring multiple dimensions efficiently by using multidimensional IRT (MIRT) applications. There have been several research studies about MCAT item selection methods to improve domain and the overall ability score estimations accuracy. According to the literature review it has been found that most studies focused on comparing item selection methods in many conditions except for the structure of test design and multidimensional calibration strategies. In contrast with the previous studies, this study employed unidimensional and multidimensional calibration approach and various test design (simple and complex) which allows the evaluation of domain and subscore ability estimations across multiple real test conditions. The purpose of this study is to compare MCAT item selection methods while estimating domain and the overall ability scores in terms of test design, number of items per dimension, calibration approaches in MCAT framework.

In this study, four factors were manipulated, namely the test design, number of items per dimension, calibration strategies and item selection methods. For each SS, CLS or CHS design 1000x3 and 1000x45 matrix of true ability parameters was randomly generated from the multivariate normal distribution. Using the generated item and ability parameters, dichotomous item responses were generated in by using M3PL compensatory multidimensional IRT model with specified correlations. A three-dimensional item bank was simulated with simple and complex structures. Dimensions correlated at ρ = 0.2, 0.5, and 0.8. Three calibration strategies, separate unidimensional and two multidimensional (Bock and Aitkin's EM and Metropolis-Hastings Robbins-Monro algorithm) calibration were examined. The

multidimensional CAT item selection procedures: minimum angle, minimize the error variance of the composite score with the optimized weight, and Kullback–Leibler (KL) information were also examined. MCAT domain and composite ability score accuracy was evaluated using absolute bias (ABSBIAS), correlation and the root mean square error (RMSE) between true and estimated ability scores.

The results suggest that the calibration approaches, multidimensional test structure and number of item per dimension have significant effect on item selection methods for both domain and the overall score estimations. As the model gets complex absolute biases had decrease significantly for both domain and overall scores.

When the test design change different item selection methods had performed better. For SS test design it was found that V1 item selection has the lowest absolute bias estimations for both SU and BAEM calibration while estimating overall scores when correlation between dimension is moderate (0.5) and test length is long (N=45). For CLS test design it was found that Vol item selection has the lowest absolute bias estimations for in SU calibration while estimating overall scores when correlation between dimension is high (0.8) and test length is long (N=45). For BAEM calibration Vol item selection has the lowest absolute bias estimations while estimating overall scores when correlation between dimension is low (0.2) and test length is long (N=45). For CHS test design it was found that V1 item selection has the lowest absolute bias estimations for in SU calibration while estimating overall scores when correlation between dimension is high (0.8) and test length is long (N=45). For BAEM calibration KL item selection has the lowest absolute bias estimations while estimating overall scores when correlation between dimension is low (0.2) and test length is long (N=45).

Keywords: Multidimensional item response theory, multidimensional computer adaptive testing, domain and overall score, item selection methods, unidimensional and multidimensional calibration

Advisor: Prof. Dr. Selahattin GELBAL, Hacettepe University, Department of Educational Science, Division of Educational Measurement and Evaluation

ETİK BEYANNAMESİ

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 tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

İmza
Eren Halil ÖZBERK

TEŞEKKÜR

Doktora eğitimim boyunca desteğini hiç esirgemeyen, yardımları ve yönlendirmeleriyle akademik gelişimimde emeği olan, desteğini her zaman hissettiğim danışmanım Prof. Dr. Selahattin GELBAL'a teşekkür ederim.

Doktora tez izleme komitesinde ve tez savunma sınavı jürimde yer alan, tezin her aşamasında önerileri ve yol göstermeleriyle desteklerini esirgemeyen hocalarım Prof. Dr. Hülya KELECİOĞLU'na ve Doç. Dr. Nilüfer KAHRAMAN'a sonsuz teşekkürlerimi sunarım.

Tez savunma sınavı jürimde yer alan Prof. Dr. Şener BÜYÜKÖZTÜRK'e ve Prof. Dr. Nuri DOĞAN'a geri bildirimleri ve yardımları için çok teşekkür ederim.

Amerika'daki eğitimim sırasında tezin ilk aşamasından son haline kadar fikirleri ve yönlendirmeleriyle ufkumu açan, hem akademik hem bireysel olarak desteğini esirgemeyen ikinci danışmanım Prof. Dr. Terry Ackerman'a ve eşi Deb Ackerman'a teşekkürlerimi borç bilirim.

Araştırmanın güncel sorunlara en yakın koşullarda olmasını adına fikirler sunan, yazılımsal, akademik ve bireysel desteğini hiç esirgemeyen, hem akademik hem de sosyal olarak örnek aldığım sayın hocam Prof. Dr. Ric Luecht'e ve eşi Victoria Luecht'e teşekkür ederim.

Ders aldığım süreler içinde aklıma takılan sorularıma bıkmadan cevap veren, tezin analizlerinde ve raporlanmasında her türlü desteği sağlayan hocalarım Dr. John Willse'ye ve Dr. Robert Henson'a teşekkür ederim.

Doktora derslerinden başlayıp, yeterlik, tez önerisi ve tez izlemelerde hep yanyana, fikir alışverişinde bulunduğum değerli dostum Arş. Gör. Dr. Akif AVCU'ya teşekkür ederim.

Yüksek lisansta başladığımız serüvenimizde hep yanımda olan, desteğini her zaman hissettiğim sevgili eşim, hayat arkadaşım Dr. Elif Bengi ÜNSAL ÖZBERK'e ve sevgili aileme şükranlarımı sunarım.

Doktora sırası araştırma bursu ile 1 yıl boyunca Amerika'da University of North Carolina-Greensboro'da eğitim almamı sağlayan TÜBİTAK'a teşekkür ederim.

İÇİNDEKİLER

ÖZ…… iii
ABSTRACT.. v
ETİK BEYANNAMESİ vii
TEŞEKKÜR viii
İÇİNDEKİLER ix
ÇİZELGELER DİZİNİ xi
ŞEKİLLER DİZİNİ xii
SİMGELER VE KISALTMALAR DİZİNİ xiii
1. GİRİŞ….. 1
1.1. Problem Durumu 1
1.2. Araştırmanın Amacı ve Önemi: 5
1.3. Problem Cümlesi: 6
1.3.1. Alt Problemler: 6
1.4. Sayıltılar: 7
1.5. Sınırlılıklar: 7
1.6. Tanımlar: 8
1.7. Araştırmanın Kuramsal Temeli 9
2. İLGİLİ ARAŞTIRMALAR 27
2.1. İlgili Araştırmalar Özet 32
3. YÖNTEM 34
3.1. Araştırmanın Türü 34
3.2. Araştırmanın Deseni 34
3.3. Verilerin Üretilmesi 34
3.3.1. Madde Parametrelerinin Üretilmesi 34
3.3.2. Test Desenlerinin Oluşturulması 35
3.3.3. Yetenek Parametrelerinin Üretilmesi 36
3.4. Kalibrasyon Yöntemleri 38
3.4.1. Ayrı Tek Boyutlu Kalibrasyon 38
3.4.2. Bock & Aitkin Expectation-Maximization (BAEM) Algoritması 38
3.4.3. The Metropolis-Hastings Robbins-Monro (MH-RM) Algoritması 40
3.5. Değerlendirme Kriteri 40
3.6. Verilerin Analizi 41
3.6.1. ÇBBBT Analizi 41
4. BULGULAR VE TARTIŞMA 43
4.1. Gerçek Yetenek Parametrelerine Ait Bulgular 43
4.2. Tek ve Çok Boyutlu Kalibrasyona Ait Bulgular 44
4.2.1. Basit Yapılı (BY) Testlerde Yetenek Puanlarının Karşılaştırılması 45
4.2.2. Düşük Karmaşık Yapılı (DKY) Testlerde Yetenek Puanlarının Karşılaştırılması 47

4.2.3. Yüksek Karmaşık Yapılı (YKY) Testlerde Yetenek Puanlarının Karşılaştırılması 52
4.3. Birinci Alt Probleme Ait Bulgular ve Yorumlar 55
4.3.1. Kalibrasyon Yönteminin Etkisinin İncelenmesi 55
4.3.2. Boyutlar Arası Korelasyonun ve Boyut Başına Düşen Madde Sayısının Etkisinin İncelenmesi 60
4.4. İkinci Alt Probleme Ait Bulgular ve Yorumlar 67
4.5. Üçüncü Alt Probleme Ait Bulgular ve Yorumlar 68
4.5.1. Kalibrasyon Yönteminin Etkisinin İncelenmesi 69
4.5.2. Boyutlar Arası Korelasyonun ve Boyut Başına Düşen Madde Sayısının Etkisinin İncelenmesi 73
4.6. Dördüncü Alt Probleme Ait Bulgular ve Yorumlar 80
4.7. Beşinci Alt Probleme Ait Bulgular ve Yorumlar 81
4.7.1. Kalibrasyon Yönteminin Etkisinin İncelenmesi 82
4.7.2. Boyutlar Arası Korelasyonun ve Boyut Başına Düşen Madde Sayısının Etkisinin İncelenmesi 85
4.8. Altıncı Alt Probleme Ait Bulgular ve Yorumlar 92

5. SONUÇ ve ÖNERİLER 95

5.1. Sonuçlar 95
5.2. Öneriler 103
5.2.1. Araştırma Sonuçlarına Dayalı Öneriler 103
5.2.2. Yapılacak Araştırma Önerileri 104

KAYNAKÇA 105

EKLER DİZİNİ 113

EK 1. ETİK KURUL ONAY BİLDİRİMİ 114
EK 2. ORJİNALLİK RAPORU 115
EK 3. TEK BOYUTLU TEST DESENİ İÇİN ÜRETİLEN MADDE PARAMETRELERİ 117
EK 4. ÇOK BOYUTLU TEST DESENLERİ İÇİN ÜRETİLEN MADDE PARAMETRELERİ 118
EK 5. YETENEK ALANLARINA GÖRE BİLGİ ORANLARI 121

ÖZGEÇMİŞ 122

ÇİZELGELER DİZİNİ

Çizelge 3.1: Araştırma Deseni 34
Çizelge 4.1: Boyutlar arası Korelasyon Değerleri Kestirimleri 43
Çizelge 4.2: Gerçek Yetenek Puanlarına Ait Ortalama ve Standart Sapma Değerleri 44
Çizelge 4.3: Gerçek ve Kalibre Edilmiş Puanlar Arası BIAS Değerleri 48
Çizelge 4.4: Gerçek ve Kalibre Edilmiş Puanlar Arası RMSE Değerleri 49
Çizelge 4.5: Gerçek ve Kalibre Edilmiş Puanlar Arası Korelasyon Değerleri 50
Çizelge 4.6: Basit Yapı Alt Boyut 1 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 56
Çizelge 4.7: Basit Yapı Alt Boyut 2 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 57
Çizelge 4.8: Basit Yapı Alt Boyut 3 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 58
Çizelge 4.9: Basit Yapı Toplam Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 59
Çizelge 4.10: Basit Yapı Test Deseni Yetenek Puanlarına Ait Mutlak Yanlılık ve Hata Değerleri ANOVA Sonuçları 68
Çizelge 4.11: Düşük Karmaşık Yapı Alt Boyut 1 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 69
Çizelge 4.12: Düşük Karmaşık Yapı Alt Boyut 2 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 70
Çizelge 4.13: Düşük Karmaşık Yapı Alt Boyut 3 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 71
Çizelge 4.14: Düşük Karmaşık Yapı Toplam Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 72
Çizelge 4.15: Düşük Karmaşık Yapı Test Deseni Yetenek Puanlarına Ait Mutlak Yanlılık ve Hata Değerleri ANOVA Sonuçları 81
Çizelge 4.16: Yüksek Karmaşık Yapı Alt Boyut 1 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 82
Çizelge 4.17: Yüksek Karmaşık Yapı Alt Boyut 2 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 83
Çizelge 4.18: Yüksek Karmaşık Yapı Alt Boyut 3 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 84
Çizelge 4.19: Yüksek Karmaşık Yapı Toplam Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri 85
Çizelge 4.20: Yüksek Karmaşık Yapı Test Deseni Yetenek Puanlarına Ait Mutlak Yanlılık ve Hata Değerleri ANOVA Sonuçları 93

ŞEKİLLER DİZİNİ

Şekil 1.1. Çok Boyutlu Test Yapıları 16
Şekil 3.1. Alt Boyut 1'e ait Yetenek Dağılımları 37
Şekil 3.2. Alt Boyut 2'ye ait Yetenek Dağılımları.....37
Şekil 3.3. Alt Boyut 3'e ait Yetenek Dağılımları 38
Şekil 4.1. Basit Yapılı Testlerin RMSE, BIAS ve Korelasyon Değerlerinin Koşullara Göre Karşılaştırması 46
Şekil 4.2. Düşük Karmaşık Yapılı Testlerin RMSE, BIAS ve Korelasyon Değerlerinin Koşullara Göre Karşılaştırması 51
Şekil 4.3. Yüksek Karmaşık Yapılı Testlerin RMSE, BIAS ve Korelasyon Değerlerinin Koşullara Göre Karşılaştırması 53
Şekil 4.4. Basit Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Yanlılık Değerleri.....62
Şekil 4.5. Basit Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Hata Değerleri 65
Şekil 4.6. Basit Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Korelasyon Değerleri.....67
Şekil 4.7. Düşük Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Yanlılık Değerleri.....75
Şekil 4.8. Düşük Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Hata Değerleri 78
Şekil 4.9. Düşük Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Korelasyon Değerleri.....79
Şekil 4.10. Yüksek Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Yanlılık Değerleri.....87
Şekil 4.11. Yüksek Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Hata Değerleri 90
Şekil 4.12. Yüksek Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Korelasyon Değerleri.....92

SİMGELER VE KISALTMALAR DİZİNİ

BY:	Basit Yapı (Simple Structure)
DKY:	Düşük Karmaşık Yapı (Complex Low Structure)
YKY:	Yüksek Karmaşık Yapı (Complex High Structure)
MTK:	Madde Tepki Kuramı (Item Response Theory)
TBMTK:	Tek Boyutlu Madde Tepki Kuramı (Unidimensional Item Response Theory)
TBBBT:	Tek BoyutluBireyselleştirilmiş Bilgisayarlı Testler (Computer Adaptive Testing)
ÇBMTK:	Çok Boyutlu Madde Tepki Kuramı (Multidimensional Item Response Theory)
ÇBBBT:	Çok Boyutlu Bireyselleştirilmiş Bilgisayarlı Testler (Multidimensional Computer Adaptive Testing)
MLE:	En Yüksek Olabilirlik (Maximum Likelihood Estimation)
MAP:	En Yüksek Sonsal Dağılım (Maximum a Posteriori)
EAP:	Beklenen Sonsal Dağılım (Expected a Posteriori)
Vol:	Hacim (Volume) Madde Seçme Yöntemi
KL:	Kullback-Leibler Madde Seçme Yöntemi
V1:	Minimum Hata Varyansı Madde Seçme Yöntemi

1. GİRİŞ

1.1. Problem Durumu

Eğitim ve psikoloji alanında değerlendirme araçlarının başlıca amacı ölçülen özeliğin miktarını belirlemek ve elde edilen puanları kullanarak bireyleri örtük özelliklerine göre sıralamaktır. Puanlar, sıralama amacıyla kullanıldığı durumlarda önemli bir değerlendirme ölçütü olarak kullanılabilmektedirler. Özellikle başarı düzeylerinin belirlenmesinde, sertifika ve lisanslama yetkilerinin verilmesinde puanlar önemli bir sıralama ölçütüdür. Örneğin bir öğrencinin üniversiteye giriş sınavından elde ettiği puan öğrencinin bilgi ve becerisinin bir göstergesidir. Sıralama amacıyla hazırlanan testlerden farklı olarak testler bazı durumlarda tanılayıcı puanlar verecek şekilde de hazırlanmış olabilmektedir. Örneğin, İngilizce dil yeterliğini belirleyen TOEFL sınavı bireylerin genel İngilizce becerisini belirten bir toplam puan sunarken, aynı zamanda alt beceri alanları okuma, yazma, dinleme ve konuşma alanlarına ait tanılayıcı puanlar da sunabilmektedir. Bu kapsamda hazırlanmış bir testte, aynı toplam puanı alan iki bireyin alt boyutlara ait becerileri farklılık gösterebilmektedir. Bu bakımdan alt boyut puanları farklı kapsamlarda hazırlanmış alt testlere ait bireylerin performanslarını ortaya çıkarmak amacıyla kullanılır. Bu puanlar sayesinde konu ve kapsam bilgisine ait tanılayıcı geri bildirimler raporlanabilir. Öğrenciler de alt boyutlardan elde ettikleri puanlar sayesinde güçlü ve zayıf yanlarının farkına varıp çalışmalarına yön verebilmektedir. Eğitim kurumları da bu eksiklere göre eğitim programlarını geliştirme yoluna gidebilmektedir. Psikometri alanındaki yazılımların gelişmesi ile birlikte, test geliştiriciler son zamanlarda alt boyut puanlarını ya da diğer adıyla tanılayıcı puanları rapor etmenin farklı yollarını aramaktadırlar. Alt boyut puanları; boyutlara ait puanlar, tanılayıcı puanlar (diagnostic scores), alt ölçek puanları, kapsam-düzeyi puanları gibi farklı adlandırmalara sahiptirler (de la Torre ve Song, 2009; Sinharay, Puhan ve Haberman, 2010; Stone, Ye, Zhu ve Lane, 2010).

Ancak her test tanılayıcı puanlar vermede yeterli olmayabilir. Thissen ve Edwards (2005) tanılayıcı puanlamaların hesaplanmasının test geliştirme aşamasında bir takım psikometrik zorlukları olduğunu belirtmiştir. Eğer test, tek bir özelliği ölçmek için tek boyutlu olarak geliştirilmişse testten elde edilen toplam puanlar belirlenen özelliğe ait performans hakkında değerlendirmeler yapmak için kullanılabilir. Aksi

halde testi oluşturan alt boyutlara ait puanlar, doğru ölçekleme yöntemleri kullanılmadığı takdirde anlamlı olmayacaktır. Yapılan araştırmalarda da bu duruma dikkat çekilmiş; testin alt boyut puanları rapor etme amacı olmadan geliştirilmesine rağmen alt boyut puanlarını rapor ettiği durumlarda hatalı, güvenilir olmayan ve anlamsız değerlendirmelerin ortaya çıktığı vurgulanmıştır (Goodman ve Hambleton, 2004; Haberman, 2008; Sinharay ve Haberman, 2008).

İngilizce dil yeterliğini belirleyen TOEFL sınavında olduğu gibi, toplam puanlara göre değerlendirme yapmak için hazırlanan bir test, alt boyut puanlarına göre de değerlendirme yapmak amacıyla kullanılmak istenebilir. Bu durumda test birden fazla amaca hizmet etmeye başlar. Tek bir özelliği ölçen puanlar, birden fazla özelliği ölçmeye başlayacaktır ve bu durum psikometrik olarak sorunlar ortaya çıkaracaktır. Bu sorunların en başında ölçme bilgisi gelmektedir. Madde tepki kuramı (MTK) çerçevesinde değerlendirecek olursak, θ tek boyutlu varsayıldığında, testteki maddeler ölçme bilgisini θ ölçeğinin herhangi bir noktasında maksimum noktasına çıkarmaya çalışırlar. Tek boyutlu modeller doğası gereği telafi edici modellerdir ve teste ait alt boyut (yetenek) alanları eşit doğru cevaplanma olasılığına sahiptirler. Alt boyut puanlarının tek boyutlu ölçeklerde kullanılması durumunda bazı olumsuzluklar göze çarpmaktadır. Örneğin, MTK modeliyle ölçeklenmiş ikili puanlanan tek boyutlu bir test için tek boyutluluk varsayımı ele alındığında cevaplar matrisi θ dışında başka bir özelliğe katkı sağlamayacaktır. Bu durumda, tek boyutlu bir testte alt boyut puanlarının sunulması, farklı özellikleri ortaya koymak demektir ki bu durum tek boyutlu test varsayımına göre bir ikilem yaratmaktadır. Tek boyutlu madde ve test bilgileri, çok boyutlu puan profilleri oluşturmada yetersiz kalmaktadır. Var olan bu kısıtlılıktan dolayı tanılayıcı amaçlar doğrultusunda $\theta = (\theta_1, \theta_2, \dots, \theta_m)$ gibi alt boyut puanları elde edilmek istendiğinde tek boyutlu test desenleri yerine çok boyutlu test desenlerinin kullanılması gerekmektedir. Ancak bu sayede istenilen sayıda alt boyut puanı doğru ve hatasız şekilde kestirilip, yorumlanabilir (Ackerman, 1994; Haberman, 2008; Luecht, Gierl, Tan ve Huff, 2006; Wainer ve ark, 2001, Yao, 2010).

Alt boyut ve toplam yetenek puanları rapor edilmek istendiğinde, tek boyutlu test desenleri yerine kullanılan çok boyutlu test desenlerinin yapısı da kestirimlerde önemli rol oynamaktadır. Çok boyutlu madde tepki kuramı modelleri test yapıları

bakımından basit ve karmaşık yapı olarak ikiye ayrılmıştır. Literatürde bunun için birçok farklı adlandırma mevcuttur. Bazı araştırmacılar basit ve karmaşık yapılı test deseni olarak adlandırırken (Luo, 2013; Yao, 2012; Zhang, 2012) bazıları maddeler arası model (multidimensional between-item model) ve maddeler içi model (multidimensional within-item model) olarak adlandırmıştır (Adams, Wilson ve Wang, 1997; Bulut, 2013; Wang, Chen ve Cheng, 2004). Maddelerin faktöriyel karmaşık bir yapıda olması durumunda belirli maddeler birden fazla boyuta yük verebilirler ve yine birden fazla boyuta bilgi sağlayabilirler. Testlerin çok boyutlu madde içerdiği durumlarda çok boyutlu ölçme bilgisinin ortaya çıkması, alt boyut ve toplam yetenek puanlarının kestirimine etki edebilmektedir (Bulut, 2013; Liu, 2015; Luecht, Gierl, Tan ve Huff, 2006; Luo, 2013; Zhang, 2012; Finch, 2010). Yapılan araştırmalarda çok boyutlu test yapılarının çok boyutlu bilgi düzeylerini farklılaştırdığı, bu bakımdan elde edilen yetenek ve madde parametrelerini etkilediği görülmüştür. Çok boyutlu madde tepki kuramı (ÇBMTK) modellerinin verdiği avantajlar göz önüne alındığında bazı araştırmacılar gerçek test koşullarına yakın olması nedeniyle karmaşık çok boyutlu yapıların kullanılmasını önerirken (Ackerman, 1994; Reckase, 2009) diğer araştırmacılar basit yapıdaki desenler kullanmanın karmaşık yapıda desenler kullanmadan daha avantajlı olduğunu belirtmişlerdir (Luecht & Miller, 1992; Yao & Boughton, 2009).

Birey ve yetenek parametrelerini kestirmede MTK modellerinin kullanıldığı durumlarda madde kalibrasyonunun kritik bir önemi vardır. Kalibrasyon, parametre kestiriminde test verisi ile psikometrik modelin uyumunu ortaya koyar. Kalibrasyon ile birlikte aynı madde bankasında yer alan farklı ölçeklenmiş maddeler, aynı ölçek düzeyine getirilir. Madde kalibrasyonu, özellikle madde bankaları için kritik bir önem arz eder. Madde bankasındaki her bir madde kullanılmadan önce kalibre edilmek zorundadır. Araştırmalar, bireyselleştirilmiş bilgisayarlı test uygulamalarında madde bankalarındaki maddelerin güncellenmesi, yenilerinin eklenmesi ve çıkarılması konusunda kalibrasyonun yetenek puanları kestirimine doğrudan etkisi olduğunu belirtmiştir (Chuah, Drasgow, and Luecht, 2006; Folk & Golub-Smith, 1996; Glas, 2000; Levine & Williams, 1998; Samejima, 2000; Stocking, 1988; Wainer & Mislevy, 1990). Bu nedenle kalibrasyon yöntemlerinin farklılığı, alt boyut ve toplam yetenek puanlarının rapor edilmesinde farklılıklar oluşturmaktadır.

Literatürde ÇBMTK modelleri kullanılarak kalibrasyon yöntemlerinin birbirine göre avantajlarını karşılaştıran bir çok araştırma bulunmaktadır (Bahskov, 2015; Luo, 2013; Zhang, 2012, Yao, 2003). NOHARM (Fraser, 1988; Fraser & McDonald, 1988) ve TESTFACT (Wilson, Wood, & Gibbons, 1991) programları ile birlikte ÇBMTK dayalı madde ve yetenek parametreleri kestirilmeye başlanmıştır. Her iki yöntem de faktör analizi temeline dayanmaktadır ve birbirine yakın sonuçlar vermektedir (Miller, 1991). Gelişen bilgisayar programlarının yardımı ile daha karmaşık işlemler daha hızlı bir şekilde hesaplanabilir hale gelmiş, Bayes Monte Karlo simülasyonlarına dayanan BMIRT yöntemi, Marjinal Maksimum olabilirlik yöntemine dayalı BAEM ve açımlayıcı faktör analizi yöntemini Metropolis-Hastings Robbins-Monro algoritması ile birleştiren MH-RM yöntemi geliştirilmiştir (Cai, 2010a; Yao, 2003; Yao & Boughton, 2007; Yao & Schwarz, 2006). ÇBMTK modellerinde kalibrasyon yöntemlerinin performanslarının incelenmesi tek ve çok boyutlu kalibrasyonların karşılaştırılmasına dayanmaktadır. Çok boyutlu madde tepki kuramı, boyutlar arasındaki korelasyon değerlerini dikkate aldığından alt boyut puanlarının hesaplanmasında daha tutarlı sonuçlar vermiştir (Wang, Chen, & Cheng, 2004; Yao & Boughton, 2007). Ancak literatürde, çok boyutlu madde bankasında farklı kalibrasyon yöntemlerinin çok boyutlu bireyselleştirilmiş bilgisayarlı test (ÇBBBT) uygulamaları sonucu elde edilen yetenek ve madde parametre kestirimlerine etkisini inceleyen bir çalışmaya rastlanmamıştır.

Çok boyutlu bireyselleştirilmiş bilgisayarlı testler, alt boyut puanlar ya da tanılayıcı puanlar rapor etmede tek boyutlu bireyselleştirilmiş bilgisayarlı test (TBBBT) uygulamalarına göre avantajlı durumdadır. ÇBBBT uygulamalarında her bir alt yeteneğe göre puanlar daha az madde ile kestirilebilmektedir ve her boyuta ait puanlar rapor edilebildiğinden bireylerin zayıf ve güçlü yanları boyutlara göre rahatlıkla belirlenebilmektedir (Wang ve Chang, 2011). Yakın zamanda yapılan araştırmalarda TBBBT uygulamaları için geliştirilen birçok madde seçme yöntemi ÇBBBT uygulamalarına göre tekrar geliştirilmiştir (Mulder ve van der Linden, 2010; Segall, 1996; van der Linden, 1999; Veldkamp ve van der Linden, 2002). TBBBT ve ÇBBBT madde seçme yöntemleri kestirdikleri yetenek sayılarının farklı olması bakımından ayrılmaktadır. Özellikle ÇBBBT madde seçme yöntemleri, çoklu yeteneklerin kestirimlerine getirdikleri farklı teknikler ile alt boyut puanlarının hesaplanmasında daha kararlı sonuçlar elde etmeye çalışmışlardır. Örneğin,

Segall (1996) tarafından önerilen madde seçme yöntemi, genel varyansı azaltarak yeteneklere ait güven aralıklarını düşürmek isterken, van der Linden (1999) tarafından önerilen yöntem ise her bir yetenek kestirimine ait toplam varyansı azaltma yoluna gitmiştir. Literatürde alt boyut ve toplam puanları rapor etmede çok boyutlu madde seçme yöntemlerinin karşılaştırılmasına ilişkin çalışmalar olsa da (Wang ve Chang, 2011; Yao, 2012, 2013), çok boyutlu test deseninin yapısına ve kalibrasyon yöntemlerine göre nasıl performans gösterdiğine ilişkin bir çalışmaya rastlanmamıştır.

1.2. Araştırmanın Amacı ve Önemi:

Çok boyutlu bireyselleştirilmiş bilgisayarlı test (ÇBBBT) uygulamalarının tek boyutlu bireyselleştirilmiş bilgisayarlı test (TBBBT) uygulamalarına göre üstünlükleri vardır: a) ÇBBBT, TBBBT'den daha fazla bilgi sağlamaktadır. ÇBMTK test yapılarında boyutlar arasında ilişkiler mevcuttur ve bir madde birden çok boyuta yük verebilir. Bu sayede her bir maddenin bilgi vericiliği ÇBBBT uygulamalarında daha dikkatle ele alınır ve TBBBT'ye göre daha kararlı sonuçlar elde edilir (Segall, 1996; Wang, Chang ve Boughton, 2011). Maddelerden elde edilen bilgiler arttığından test uzunluğu da düşmektedir (Segall, 1996). b) ÇBMTK desenleri doğal olarak bir kapsam dengeleme avantajı sunar. TBBBT uygulamalarında yaşanan sadece belirli bir kapsamdan soruların seçilmesi sorunu, geliştirilen çok boyutlu madde seçme yöntemleri yardımı ile sorun olmaktan çıkmaktadır. TBBBT uygulamalarında kullanılan kapsam dengeleme kısıtlamaları sonucu bazı alt boyutlar bireyin genel yeteneğine daha az katkı sağlamaktadır. ÇBBBT uygulamaları, kapsam alanlarını korelasyon değerlerini de göz önüne alarak ayrı ayrı ele alır ve farklı kapsamlardan elde edilen bilgiyi bütün boyutlarla beraber işleme koyar (Segall, 1996; Wang ve Chang, 2011). İncelenen araştırmalar madde sayısının ve maddeler arası korelasyon değerlerinin alt boyut ve toplam puanları hesaplamada değişebildiğini göstermektedir. Bu sebeple boyutlar arası korelasyon ve boyut başına düşen madde sayısının madde seçme yöntemleri üzerindeki etkisinin nasıl değişeceğinin belirlenmesinin uygulayıcılara önemli bilgiler sağlayacağı düşünülmektedir.

Bu çalışmada ele alınan bir başka durum ise ÇBBBT uygulamalarında yetenek parametrelerini kestirmede çok boyutlu test desenine bağlı olarak kalibrasyon yöntemlerinin etkisinin incelenmesidir. Alan yazında basit ve karmaşık yapılı çok

boyutlu test desenlerinde tek boyutlu kalibrasyon ile çok boyutlu kalibrasyon yöntemleri karşılaştırılmıştır. Ancak çok boyutlu test yapılarının farklılaşması durumunda BAEM, MHRM ve tek boyutlu kalibrasyonları karşılaştıran araştırma bulunmamaktadır. Belirlenen kalibrasyon yöntemlerinin farklı koşullarda nasıl performans göstereceğine ilişkin elde edilecek bulguların ÇBBBT uygulamaları açısından önemli olacağı düşünülmektedir.

Sonuç olarak, ÇBBBT uygulamalarında alt boyut ve toplam puanları hesaplamada madde seçme yöntemlerinin farklı kestirim değerleri sunduğu belirlenmiştir (Yao, 2012, 2013). Bu sebepten madde seçme yöntemlerinin karşılaştırılması ve alt boyut ve toplam puanları kestirmede farklı koşullar için en az hata veren yöntemlerin belirlenmesi gerekmektedir. Geniş ölçekli sınavlarda tanılayıcı puanların rapor edilmesinin her geçen gün arttığı, sınavı alan bireylere bu puanlar doğrultusunda geri bildirimler verildiği dikkate alındığında ÇBBBT uygulamalarında tanılayıcı ve toplam puanları rapor etmede en az hata içeren koşulların belirlenmesi önem taşımaktadır.

1.3. Problem Cümlesi:

Basit, düşük ve yüksek karmaşık yapıdaki testlerde kalibrasyon yöntemlerinin, boyutlar arası korelasyonun ve boyut başına düşen madde sayısının madde seçme yöntemlerinin hata, mutlak yanlılık ve korelasyon değerlerine etkisi nasıldır?

1.3.1. Alt Problemler:

1. Basit Yapılı (BY) test deseninde çok boyutlu madde seçme yöntemlerinin hatası, mutlak yanlılığı ve korelasyon değerleri
    a. Kalibrasyon yöntemine göre,
    b. Boyutlar arası korelasyona göre ve boyut başına düşen madde sayısına göre alt boyut ve toplam yetenek puanları için nasıl değişmektedir?
2. Basit Yapı test deseninde yetenek puanlarına ait mutlak yanlılık ve hata değerleri arasında koşullara göre anlamlı farklılıklar var mıdır?
3. Düşük Karmaşık Yapı (DKY) test deseninde çok boyutlu madde seçme yöntemlerinin hatası, mutlak yanlılığı ve korelasyon değerleri

a. Kalibrasyon yöntemine göre,

b. Boyutlar arası korelasyona göre ve boyut başına düşen madde sayısına göre alt boyut ve toplam yetenek puanları için nasıl değişmektedir?

4. Düşük Karmaşık Yapı test deseninde yetenek puanlarına ait mutlak yanlılık ve hata değerleri arasında koşullara göre anlamlı farklılıklar var mıdır?

5. Yüksek Karmaşık Yapı (YKY) test deseninde çok boyutlu madde seçme yöntemlerinin hatası, mutlak yanlılığı ve korelasyon değerleri

   a. Kalibrasyon yöntemine göre,

   b. Boyutlar arası korelasyona göre ve boyut başına düşen madde sayısına göre alt boyut ve toplam yetenek puanları için nasıl değişmektedir?

6. Yüksek Karmaşık Yapı test deseninde yetenek puanlarına ait mutlak yanlılık ve hata değerleri arasında koşullara göre anlamlı farklılıklar var mıdır?

1.4. Sayıltılar:

Araştırmada kullanılan madde parametreleri gerçek koşullara yakın olması nedeniyle PISA 2012 Türkiye verisinden elde edilmiştir. ÇBBBT koşulları oluşturulurken gerçek koşullara mümkün olduğunca benzer veriler türetilmeye çalışılmıştır. Bu bakımdan türetilen verilerin gerçek test koşullarını yansıttığı varsayılmıştır.

1.5. Sınırlılıklar:

1. Araştırmada 3 boyutlu bir test üzerinden yürütülmüştür.

2. Araştırmada parametre kestirimleri M3PLM (Çok boyutlu Üç Parametreli Lojistik Model) ile sınırlandırılmıştır.

3. Yetenek parametrelerinin kestiriminde MAP (Maximum a posteriori) yöntemi kullanılmıştır.

4. Araştırmada boyutlar arası korelasyonlar literatürdeki çalışmalar dikkate alınarak 0.2, 0.5 ve 0.8 olarak belirlenmiştir.

5. Araştırma PISA 2012 testinin Türkiye verisi üzerinden yürütülmüştür.

6. ÇBMTK ve ÇBBBT analizlerinde *c*-parametreleri sabit tutulmuştur.

1.6. Tanımlar:

Tek Boyutlu Test: Tek bir örtük özelliği ölçen test

Çok Boyutlu Test: Birden fazla örtük özelliği ölçen test

Basit Yapı Test Deseni: Her bir maddenin sadece bir boyuta yük verdiği çok boyutlu test deseni

Düşük Karmaşık Yapı Test Deseni: Her bir maddenin esas boyut dışında diğer boyutlara düşük oranda yük verdiği test deseni

Yüksek Karmaşık Yapı Test Deseni: Her bir maddenin esas boyut dışında diğer boyutlara yüksek oranda yük verdiği test deseni.

Gerçek Puan: Her bir test koşulu için önceden belirlenen madde parametrelerinden üretilen alt boyut ve yetenek puanları

Kalibre Edilmiş Puan: Madde bankalarının tek ve çok boyutlu kalibrasyon yöntemleri ile kalibrasyonu sonucunda kestirilen alt boyut ve yetenek puanları

Kestirilen Puan: Çok boyutlu bireyselleştirilmiş bilgisayarlı test uygulamaları sonucunda kestirilen nihai alt boyut ve yetenek puanları

$\mathbf{RMSE_{Kal}}$: Kalibre edilmiş puanlar ile gerçek puanlar arasındaki hata değeri

RMSE: Kestirilen puanlar ile gerçek puanlar arasındaki hata değeri

$\mathbf{BIAS_{Kal}}$: Kalibre edilmiş puanlar ile gerçek puanlar arasındaki yanlılık değeri

BIAS: Kestirilen puanlar ile gerçek puanlar arasındaki yanlılık değeri

$\mathbf{ABSBIAS_{Kal}}$: Kalibre edilmiş puanlar ile gerçek puanlar arasındaki mutlak yanlılık değeri

ABSBIAS: Kestirilen puanlar ile gerçek puanlar arasındaki mutlak yanlılık değeri

$\mathbf{Korelasyon_{Kal}}$: Kalibre edilmiş puanlar ile gerçek puanlar arasındaki korelasyon değeri

Korelasyon: Kestirilen puanlar ile gerçek puanlar arasındaki korelasyon değeri

1.7. Araştırmanın Kuramsal Temeli

Tek boyutlu Madde Tepki Kuramı

Madde tepki kuramı teste verilen cevaplardan yararlanarak bireylerin yetenek, ilgi, tutum ve bilgilerini ortaya koyan psikometrik bir yapıdır. Temelde teste verilen cevaplar ile maddelerin etkileşimleri kullanılarak bireylere ait yetenekler ile madde parametresi değerlerinin aynı ölçek üzerinde ifade edilmesi, bu sayede sürekli bir ölçek üzerinden karşılaştırmalar yapılması amaçlanmıştır (Hambleton ve Swaminathan, 1985). Test maddelerinin ölçtüğü yetenek/özellik/performans (θ) örtük bir özelliktir. Test maddelerinin tek bir özelliği ölçtüğü durumlarda tek boyutlu MTK (TBMTK) modelleri kullanılır. TBMTK modellerinde maddeye doğru cevap verme olasılıkları yeteneğin bir fonksiyonu olarak ifade edilmektedir. Madde tepki fonksiyonu olarak adlandırılan bu fonksiyon aynı zamanda maddeye ait güçlük ve ayırt edicilik değerlerini de içermektedir.

Madde tepki kuramının bir diğer amacı ise ölçülen yetenek ya da özelliğe göre bireyleri puanlamaktır (DeMars, 2010). TBMTK modellerinde ölçülmek istenen özellik bazı durumlarda tek bir örtük özelliği ifade edebilirken, bazen de birden fazla özelliğin birleşimi olabilir. Örneğin bir matematik testi, sadece öğrencilerin aritmetik becerileri ya da aritmetik, cebir ve geometri konularını kapsayarak öğrencinin genel matematik becerilerini ölçmek için hazırlanmış olabilir.

Maddelerin puan kategorilerine göre MTK modelleri farklılık göstermektedir. İki kategorili olarak puanlanan bir madde için doğru cevap verme olasılığı 3 parametreli lojistik modele (3PLM) için denklem 1.1' e göre hesaplanmaktadır.

$$P(u_{ij}) = 1|\theta_j, a_i, b_i, c_i) = c_i + \frac{1-c_i}{1+e^{Da_i(\theta_j - b_i)}} \qquad (1.1)$$

denkleminde u_{ij} *j*'inci bireyin *i*'inci soruya doğru (1) veya yanlış (0) cevap verme olasılığı, θ_j *j*'inci bireyin θ sürekli ölçeğindeki yeteneği, a_i *i* maddesinin ayırt ediciliği, b_i *i* maddesinin güçlüğü ve c_i *i* maddesinin düşük asimptot değerini gösterir. Modelde *c*-parametresinin sıfır olması durumunda 3PL model 2PL modele, *a*-parametresinin aynı değerlerde olması durumunda 2PL model 1PL modele dönüşmektedir.

Testi alan bireylerin maddeye veya testteki performansları belirlenen özellik (θ_j) dışında başka bir beklenmeyen/gözlenmeyen özellik tarafından etkilendiği

durumda test artık tek boyutluluk özelliğini sağlamayacaktır. Bu da MTK modellerindeki iki temel varsayımdan birincisi olan tek boyutluluk varsayımının ihlali demektir. Tek boyutluluk varsayımının sağlanmadığı durumda tek boyutluluğu ihlal eden maddeler yerel bağımlı maddeler olarak adlandırılır (Yen, 1993).

$$L\left(U_{ij}\middle|\theta_j, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\right) = \prod_{j=1}^{N} P(u_{ij} = 1|\theta_j, a_i, b_i, c_i)^{u_{ij}}\, P(u_{ij} = 1|\theta_j, a_i, b_i, c_i)^{1-u_{ij}} \qquad (1.2)$$

Maddeler yerel bağımlılık gösterdiği durumda cevaplar, ölçülmek istenen yeteneğe (θ_j) göre değil, beklenmeyen/gözlenmeyen özelliğe göre belirmektedir. Bu durum denklem 1.2'deki MTK olabilirlik fonksiyonu hesaplamaları da olumsuz etkileyecektir ve testten/maddeden elde edilen bilgi ölçülmek istenen bilgiyi yansıtmayacaktır. Yerel bağımlı maddelerin varlığı çok boyutluluğun bir göstergesi olarak ele alınmakta ve araştırmacılar bu tür durumlarda ÇBMTK tekniklerinin kullanılmasını önermektedir (Ackerman, 1994; Reckase, 2009; van der Linden, 2005).

Testin tek boyutluluk özelliğini yansıtmaması istenen bir durum olmamasına rağmen bazen bu tür durumlar oluşabilmektedir. Testi alan bireylerin hazırbulunuşlukları, önceki öğrenmeleri ve becerileri bu durumun ortaya çıkmasında etkili olabilir. Testi geliştiriciler her ne kadar en belirgin boyutu ortaya çıkarıcı sorular sorsalar da deneysel olarak göz ardı edilebilir boyutlar ortaya çıkabilmektedir. İhmal edilen boyutların testin toplam puanlarının güvenirliğini etkilememesi gerekmektedir. Farklı özelliklerin istendik bir şekilde teste dâhil edilebildiği durumlar da vardır. Farklı konu kapsamlarından oluşan alt testlerden elde edilen toplam puanların yanında, her bir konu kapsamının alt testine ait alt boyut puanları tanılayıcı bilgiler içermektedir. Bu tür puanlar testin toplam test puanının anlamlılığı destekleyici nitelikte olup, testi alan bireylerin belirlenen konu kapsamlarında güçlü ve zayıf yönlerini ortaya çıkarmak için kullanılmaktadır.

Tek boyutlu olarak geliştirilen testlerde alt boyut puanlarının da kullanılmasına ilişkin çalışmalarda toplam puanların güvenilir ve doğru bir şekilde kestirildiği ancak alt boyut puanlarının tutarsız, güvenilir olmayan ve tanılayıcı olarak zayıf bilgiler ürettiği belirlenmiştir (Luecht, Gierl, Tan ve Huff, 2006; Wainer ve ark., 2001). Haberman (2008) araştırmasında alt test puanlarını düşük hata ve yüksek güvenirlikle kestirmek için gerçek puanlar ile kestirilen puanlar arasında ortalama

hata karelerinin oransal azalımı (PRMSE-protortional reduction of mean square error) kullanarak deneysel karşılaştırmalar yapmış, ancak çoğu alt boyut puanlarının tanılayıcı olarak kullanılmasının uygun olmadığını belirtmiştir. Yapılan araştırmalar incelendiğinde, boyutlara bağlı yeteneklerin puanlanmasının toplam testin puanlanmasından farklı olduğu, buna bağlı olarak yapılan değerlendirmelerin de farklılık gösterdiği belirlenmiştir. Bu sebeple boyutlar arasında bilgiler toplanırken ÇBMTK modellerinin uygulanmasının daha uygun olacağı sonucuna varılmıştır (Luecht, 1996; Reckase, Ackerman ve Carlson, 1988; Segall, 1996, 2010).

Çok Boyutlu Madde Tepki Kuramı

Reckase (1997) ÇBMT kuramını tek boyutlu madde tepki kuramının çok boyutlu verilere uygulanan genişletilmiş bir hali ya da doğrulayıcı faktör analizinin özel bir durumu olarak tanımlamaktadır. Eğitimde ve psikolojide kullanılan birçok test aslında çok boyutludur ve bu testler birden fazla yapıyı veya boyutu ölçmektedir (Ackerman, Gierl ve Walker, 2003). Yapılan araştırmalarda tek boyutlu olarak varsayılan testlere TBMTK modelleri uygulanmış, değerlendirmeler ve puanlamalar bu modele göre yapılmış ancak gerçek test koşullarına ulaştığında tek boyutluluk varsayımının sıklıkla ihlal edildiği ve boyut sayısının olması gerekenden az olduğu belirlenmiştir (Ackerman, 1994; Nandakumar, 1994; Reckase, Hirsh, 1991; Reckase, 2009).

ÇBMTK, teste ait boyutlar arasındaki ilişkileri göz önüne alması bakımından TBMTK kuramından ayrılmaktadır. Üniversitelerde de sıklıkla yapılan dil yeterliğini ölçen bir sınavı ele aldığımızı varsayalım. Bu tür sınavlar dinleme, okuma, yazma ve konuşma becerilerini kullanarak bireylerin dil yeterliklerini belirlemektedir. Bu dört beceri dil yeteneği ile doğrudan ilgili olmasına rağmen hiçbirinin tek başına dil yeterliğini yordadığı söylenemez. Ya da bu becerilerin birindeki başarı veya başarısızlık dil yeterliğinde başarı ya da başarısızlığın tek nedeni olmayabilir. Bu noktada dil yeterliğinin TBMTK ve ÇBMTK modelleri ile kestirilmesi arasında farklar oluşmaktadır. ÇBMTK modelleri boyutlar arasındaki ilişkileri dikkate aldığından TBMTK modellerine göre daha doğru kestirimler sağlar. Walker ve Beretvas (2003) yaptıkları araştırmada birden fazla yeteneği ölçen bir testte yetenek ve madde parametrelerinin kestirimlerini TBMTK modelleri ile incelemişlerdir. Araştırmanın sonucuna bakıldığında madde ve yetenek

parametresi kestirimlerini TBMTK varsayımlarına göre yapmanın hataları arttırdığı belirlenmiş, alt yeteneklerde öğrenci yeterlikleri hakkında hatalı kestirimlere rastlanmıştır.

Tek boyutlu olarak geliştirilen testlerin katı tek boyutluluk özelliği göstermediği durumlar da gerçek test koşullarında karşımıza çıkmaktadır. Matematik becerisini ölçen bir test, aynı zamanda okuma becerilerini de ölçüyor olabilir. Bu durumda testin temel boyutu matematik becerisi iken, okuma becerisi boyutu beklenmeyen/artık (nuisance) yetenek olarak adlandırılır (van der Linden, 2005). Yapılan araştırmalar incelendiğinde beklenmeyen yetenek içeren bir teste TBMTK modelleri uygulandığında, madde parametreleri kestirimlerinin yanlılığının fazla, yetenek parametrelerine ait hata kestirimlerinin beklenenin aksine yanıltıcı bir şekilde az hata ile kestirildiği görülmüştür (Drasgow, Parson, 1983; Reckase, 1979). Bu bakımdan ÇBMTK modellerinin, istenen ve istenmeyen çok boyutluluğu değerlendirmede daha güçlü olduğu düşünülmektedir.

Çok Boyutlu Madde Tepki Kuramı Modelleri

Reckase (2009) ÇBMTK modellerini telafi edici (compensatory) ve telafi edici olmayan (noncompensatory) olmak üzere iki kategoride gruplamaktadır. Kategoriler, maddenin cevaplama olasılıklarını belirlemek amacıyla θ-koordinatındaki vektörden elde edilen bilginin madde karakteristikleriyle birlikte ele alınma şekline göre değişmektedir. Telafi edici model θ-koordinatlarının doğrusal kombinasyonuna dayanmaktadır. Testi alan bireylerin bir boyuttan elde ettiği θ_1 düşük olsa bile, diğer boyuttan elde ettiği θ_2 değeri ile toplam yeteneğini telafi edebilecektir (Reckase, 1997; Yao, Boughton, 2007). Modelde cevap verme olasılıkları belirlemede lojistik ya da normal ogive modeller kullanılabilmektedir.

Telafi edici olmayan modeller her bir maddeye ait bilişsel beceriler birbirinden ayrılmakta ve her bir beceri tek boyutlu olarak ele alınmaktadır. Maddenin doğru cevaplanma olasılığı da birbirinden ayrılan becerilerin olasılıklarının çarpımı ile hesaplanmaktadır. Matematiksel olarak olasılıkların çarpımı doğrusal olmayan bir model ortaya çıkarmaktadır. Bu bakımdan telafi edici modelin aksine bir boyuttan elde edilen θ_1 düşük olduğunda, diğer boyuttan elde edilen yüksek θ_2 değeri toplam yeteneğini telafi etmeyecektir. Telafi edici olmayan modellerde yüksek

puan elde etmek için testi alan bireylerin her iki yetenekten de yüksek puan alması gerekmektedir.

Literatürde hem iki kategorili hem de çok kategorili maddeler için ÇBMTK modelleri önerilmiştir (Ackerman, Gierl, Walker, 2003; Reckase, 2009). Bu araştırmada iki kategorili maddeler kullanıldığından iki kategorili ÇBMTK modelleri tanıtılmıştır.

Çok Boyutlu İki Parametreli Lojistik Model (Multidimensional Two-Parameter Logistic Model-M2PL)

Çok boyutlu iki parametreli lojistik modelde *i* maddesine ait doğru cevap verme olasılığını denklem 1.3 ile hesaplanmaktadır:

$$P\left(U_{ij} = 1|\boldsymbol{\theta_j}, \boldsymbol{a_i}, d_i\right) = \frac{\exp(\sum_{p=1}^{P} \boldsymbol{a_i}\boldsymbol{\theta_j} + d_i)}{1+\exp(\sum_{p=1}^{P} \boldsymbol{a_i}\boldsymbol{\theta_j} + d_i)} \qquad (1.3)$$

Denklemi tek boyutlu 2PL model denkleminden ayıran en temel nokta ayırt edicilik ve yetenek parametrelerinin tek bir değer değil (skaler), vektörel olmasıdır. Denklemde U_{ij} *j* bireyinin *i* maddesine verdiği yanıt (doğru için 1, yanlış için 0); a_i *i* maddesine ait ayırt edicilik değerleri vektörünü; d_i *i* maddesine ait güçlük değerini (skaler) ve θ_j, k boyuttan oluşan yetenek parametrelerinin vektörel değerini göstermektedir. Denklemde hem θ hem de a değerleri k boyut sayısı olmak üzere, $1\ x\ k$ vektörleri şeklindedir ($a_i = (a_{i1}, a_{i2}, \dots, a_{ik})$; $\theta_j = \left(\theta_{j1}, \theta_{j2}, \dots, \theta_{jk}\right)$).

Çok Boyutlu Üç Parametreli Lojistik Model (Multidimensional Three-Parameter Logistic Model-M3PL)

Reckase (1985) tek boyutlu 3PL modelin çok boyutlu verilere uyarlanmış halini formülleştirmiştir. M2PL modele ek olarak düşük asimptot parametresi de hesaplamaya dahil edilmiştir. Çok boyutlu üç parametreli lojistik modelde *i* maddesine ait doğru cevap verme olasılığını denklem 1.4 ile hesaplanmaktadır:

$$P\left(U_{ij} = 1|\boldsymbol{\theta_j}, \boldsymbol{a_i}, d_i, c_i\right) = c_i + (1 - c_i)\frac{\exp(\sum_{p=1}^{P} \boldsymbol{a_i}\boldsymbol{\theta_j} + d_i)}{1+\exp(\sum_{p=1}^{P} \boldsymbol{a_i}\boldsymbol{\theta_j} + d_i)} \qquad (1.4)$$

Denklem 1.4'te U_{ij} *j* bireyinin *i* maddesine verdiği yanıt (doğru için 1, yanlış için 0); a_i *i* maddesine ait ayırt edicilik değerleri vektörünü; d_i *i* maddesine ait güçlük değerini, c_i *i* maddesine ait düşük asimptot değeri ve θ_j *k* boyuttan oluşan yetenek parametrelerinin vektörel değerini göstermektedir. Yukarıda belirtilen her iki ÇBMTK modeli incelendiğinde toplam doğru cevap verme olasılığı, her bir θ

değerinin doğru cevap verme olasılıklarının toplamı alınarak elde edilir. Bu bakımdan denklemler telafi edici modellerde doğru cevap verme olasılıklarını hesaplamaktadır. Telafi edici olmayan modellerde toplam olasılık değerlerinin bulunması farklılık göstermektedir. Telafi edici olmayan modellerde toplam olasılık her bir θ değerinin doğru cevap verme olasılıklarının toplamı yerine çarpımı ile hesaplanmaktadır. Modeldeki her bir boyutun ayrı olasılık değeri vardır ve bu değerler çarpımı toplam doğru cevap verme olasılığının belirtir (Reckase, McKinley, 1982). Sympson (1978), telafi edici olmayan modellerde ikili puanlanan maddelerde doğru cevap verme olasılıklarını denklem 1.5 ile hesaplamıştır.

$$P\left(U_{ij} = 1|\boldsymbol{\theta_j}, \boldsymbol{a_i}, \boldsymbol{b_i}, c_i\right) = c_i + (1 - c_i) \prod_{k=1}^{k} \frac{exp\left[\boldsymbol{a_i}(\boldsymbol{\theta_j} - \boldsymbol{b_i})\right]}{1 + exp\left[\boldsymbol{a_i}(\boldsymbol{\theta_j} - \boldsymbol{b_i})\right]} \qquad (1.5)$$

Telafi edici modellerin aksine telafi edici olmayan modelde doğru cevap verme olasılığı hesaplama denklemindeki ilk farklılık her bir boyuta ait ayrı güçlük parametrelerinin varlığıdır. Denklemde bulunan b_i parametresi de artık θ_j ve a_i gibi $1\ x\ k$ boyutunda bir vektördür ($b_i = (b_{i1}, b_{i2}, \dots, b_{ik})$). Telafi edici olmayan model bir boyuttaki yüksek yetenek değerinin diğer boyuttaki düşük yetenek değerini telafi etmesine izin vermemektedir.

ÇBMTK modellerinde de TBMTK modellerinde olduğu madde karakteristiklerini belirlemek amacıyla güçlük, ayırt edicilik ve bilgi fonksiyonu gibi bir takım istatistikler geliştirilmiştir (Reckase, 1985; Reckase, McKinley, 1991). Madde ayırt ediciliği, madde cevap vektörünün uzunluğu olarak tanımlanmıştır ve bu uzunluk ayırt ediciliğin miktarını belirlemektedir (Reckase, McKinley, 1991). i maddesine ait çok boyutlu ayırt edicilik (multidimensional discrimination-MDISC)

$$MDISC = \sqrt{\sum_{k=1}^{k} {a_{ik}}^2} = \sqrt{{a_{i1}}^2 + {a_{i2}}^2 + \cdots + {a_{ik}}^2} \qquad (1.6)$$

ile hesaplanmaktadır. Denklem 1.6'da a_{i1}, a_{i2} ifadeleri i maddesinin $1, 2, \dots, k$ boyutlarındaki ayırt edicilik parametrelerini belirtmektedir. MDISC değeri daima pozitif değerdedir ve TBMTK modelindeki *a*-parametresinin ÇBMTK modelindeki karşılığıdır.

Çok boyutlu güçlük (multidimensional difficulty-MDIFF), madde cevap vektörünün koordinat eksenindeki yeri olarak tanımlanmıştır ve D ile gösterilmektedir ve denklem 1.7 ile hesaplanmaktadır.

$$MDIFF = D = \frac{-d_i}{\sqrt{\sum_{k=1}^{k} a_{ik}{}^2}} = \frac{-d_i}{MDISC} \qquad (1.7)$$

Denklem 1.7'deki d_i i maddesinin yer parametresini belirtmektedir. Negatif D değerine sahip maddeler kolay, pozitif D değerine sahip maddeleri ise zor maddeleri ifade etmektedirler. MDIFF TBMTK modelindeki *b*-parametresinin ÇBMTK modelindeki karşılığıdır.

Maddelerin koordinat düzlemindeki yerini belirlemede madde cevap vektörünün herhangi bir θ ekseni ile arasındaki geometrik açı denklem 1.8 ile hesaplanmaktadır.

$$cosa_{ik} = \frac{a_{ik}}{\sqrt{\sum_{k=1}^{k} a_{ik}{}^2}} = \frac{a_{ik}}{MDISC} \qquad (1.8)$$

Örneğin θ_1 ve θ_2 gibi iki eksenden oluşan bir koordinat düzleminde 45° derecelik bir açıya sahip madde için her iki boyuta da eşit oranda ölçüyor demektir.

MTK'da ölçmenin hassasiyeti (measurement precision) bilgi fonksiyonunu yardımı ile belirlenmektedir. Bilgi fonksiyonu ne kadar fazlaysa asimptotik varyans o kadar düşük olacaktır ve ölçme hassasiyeti artacaktır (Ackerman, Gierl, Walker, 2003). Çok boyutlu bilgi (multidimensional information-MINF) belirtilen ölçme hassasiyetinin bir göstergesidir ve denklem 1.9 ile hesaplanmaktadır.

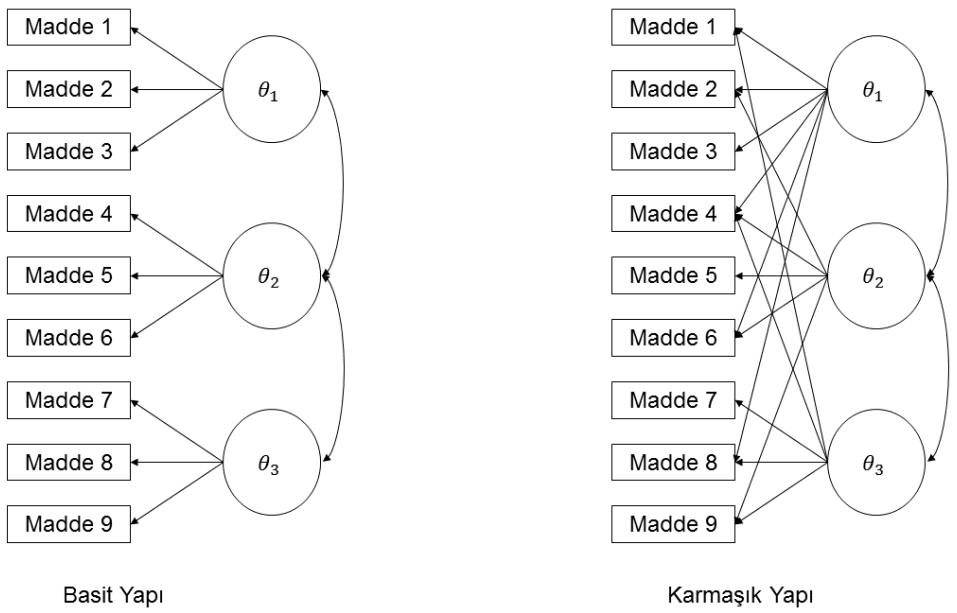
$$MINF = I_a(\theta) = P_i(\theta)[1 - P_i(\theta)](\sum_{k=1}^{k} a_{ik} cosa_k)^2 \qquad (1.9)$$

$$= P_i(\theta)Q_i(\theta)(a_{i1}cosa_{i1} + a_{i2}cosa_{i2} + \cdots + a_{ik}cosa_{ik})^2$$

MINF hem madde düzeyinde hem de test düzeyinde hesaplanabilmektedir. Test düzeyinde MINF değeri maddelere ait MINF değerlerinin toplamına eşittir.

Çok Boyutlu Test Yapıları

MTK analizlerindeki tek boyutluluk varsayımı maddelerin aynı örtük özelliği ölçtüğü ilkesine dayanmaktadır. Birden çok tek boyutlu test birlikte kalibre edilirken testler ayrı ayrı ele alınmakta ve örtük özellikler arasındaki korelasyonlar dikkate alınmamaktadır. Çok boyutlu madde tepki kuramı modelleri test yapıları bakımından basit ve karmaşık yapı olarak ikiye ayrılmıştır ve Şekil 1.1'de gösterilmiştir.



Şekil 1.1. Çok Boyutlu Test Yapıları

Bazı maddeler faktöriyel olarak basit veya karmaşık yapıda olabilmektedir. Faktöriyel olarak karmaşık yapıda olan madde birçok faktöre bilgi sağlayan maddeler olarak adlandırılır (Luecht, Gierl, Tan ve Huff, 2006). Boyutların birbirleri ile etkileşimin olduğu ve maddelerin sadece belirli bir faktöre yük verdiği test yapıları basit (simple) test yapıları ya da maddeler-arası test yapıları (between-item) olarak adlandırılmaktadır. Boyutların birbirleri ile etkileşimin olduğu ve maddelerin birden fazla faktöre yük verdiği test yapıları ise karmaşık (complex) test yapıları ya da maddeler-içi test yapıları (within-item) olarak adlandırılmaktadır (Bulut, 2013; Wang, Chen ve Cheng, 2004; Yao, 2012; Zhang, 2012). Basit ve karmaşık yapıdaki test desenleri maddelerin faktörlerle etkileşimini doğrudan etkilediğinden maddelerin bilgi vericiliklerini de etkilemektedir. Bu durumda faktörlere/örtük özelliklere ait bireylerin alt boyut ve toplam yetenek puanlarının hesaplanması da farklılaşmaktadır.

Çok Boyutlu Madde Kalibrasyon Yöntemleri

Madde parametrelerinin kestirmede MTK modellerinin kullanıldığı durumlarda madde kalibrasyonunun kritik bir önemi vardır. Kalibrasyon, parametre kestiriminde test verisi ile psikometrik modelin uyumunu ortaya koyar ve özellikle madde bankaları için kritik bir önemi arz eder. ÇBMTK modelleri için farklı

kalibrasyon yöntemleri ortaya atılmıştır. İlk kalibrasyon yöntemleri temelde faktör analizi üzerinden yürütülmüş ve hesaplamalardaki gelişmeler yazılımlar ile beraber değişmiştir. İlk olarak açımlayıcı faktör analizi yöntemi TESTFACT programı (Wilson, Wood ve Gibbons, 1998) ile birlikte kullanılmaya başlanmış, madde parametrelerinin kestiriminde marjinal en yüksek olabilirlik (marginal maximum likelihood-MML) ile EM algoritması (Bock, Aitkin, 1981) birlikte kullanmıştır. NOHARM (Fraser, 1988) TESTFACT'den farklı olarak doğrulayıcı faktör analizi yöntemlerini kullanmıştır. Reckase (2009) her iki programın kalibrasyon performansların tutarsız olduğunu, birbirine göre üstünlüklerinin olmadığını belirtmiştir. ÇBMTK modellerinde madde kalibrasyonlarına ait çalışmalar incelendiğinde faktör analizi yöntemlerinin tutarlı sonuçlar vermediği, boyutları belirlemede rotasyon yöntemlerine göre çok fazla değişiklik gösterdiği belirtilmiştir (McLeod, Swygert, Thissen, 2001).

Faktör analizi yöntemlerine farklı bir alternatif Luecht ve Miller (1992) tarafından getirilmiştir. Araştırmacılar ÇBMTK modellerini kalibre etmede faktör analizi ve hiyerarşik kümeleme analizi yöntemini beraber kullanmışlardır. İlk aşamada maddeler bağımsız kümeler (independent cluster) oluşturacak şekilde tek boyutlu alt testlere ayrılmakta, ikinci aşamada ise madde parametreleri kestirilmektedir. Wu, Adams, Wilson (1998), ÇBMTK modellerinde maddelerin kalibrasyonlarında örtük regresyon modelini kullanmışlar ve ConQuest programını geliştirmişlerdir. Program ile hem basit hem de karmaşık yapıdaki testlere ait madde parametreleri kestirilebilmektedir. Son zamanlarda bilgisayar yazılımlarının gelişmesi ile birlikte bir çok farklı kalibrasyon yöntemi kullanılmaya başlanmış yöntemlerin birbirine üstünlükleri üzerine çalışmalar yapılmıştır. Han, Peak (2014) halihazırda kullanılan tüm ÇBMTK madde kalibrasyon yöntemlerini [BAEM (Bock, Aitkin, 1981), MHRM (Cai, 2010b), Uyarlanabilen Quadrature Yaklaşımı (Adaptive Quadrature Approach-ADQ) (Schilling, Bock, 2005), MCMC] programları karşılaştırarak performanslarını hem basit hem de karmaşık yapıda karşılaştırmıştır.

Alt boyut puanlarının kestiriminde ÇBMTK modellerinin birçok avantajı vardır. ÇBMTK modellerine göre alt boyut puanlarının kestiriminde testlerin basit yapıda olması zorunluluğu yoktur. Testler ister basit ister karmaşık yapıda olsun alt boyut puanları kolayca kestirilebilir. Ayrıca ÇBMTK modelleri ile alt boyut puanları kestirirken boyutlar arası korelasyonlar dikkate alınmakta ve tek bir kestirim süreci

kullanılmaktadır. Literatürde alt boyut puanlarının kestirimlerine ilişkin birçok farklı yaklaşım kullanılmıştır (Beguin, Glas, 2001; de la Torre, Patz, 2005; Yao, Boughton, 2007).

Bireyselleştirilmiş Bilgisayarlı Testler

Bireyselleştirilmiş testler, testleri oluşturma, testi bireylere uygulama ve bireyleri puanlama işlemlerinin bilgisayar ortamında yapılmış halidir (Reckase, 2009). Kapsam ve madde sayısı bakımından, bir ölçme aracında yer alan maddelerin tümü her bireye uygulanırsa, ölçme aracı "sabit ve değişmez" olarak tanımlanırken; madde sayısı değişkenlik gösterdiğinde ve/veya maddeler her bir birey için farklılık gösterdiğinde, "bireye uyarlanmış/bireyselleştirilmiş/uyarlamalı (adaptive)" kategorisinde değerlendirilir. Bireye uyarlanmış test ile benzer olarak kullanılan diğer terimler; adapte edilen test yöntemi/bireye göre ayarlanan test yöntemi (tailored testing), bölümlendirilmiş test yöntemi (branched testing), programlanmış test yöntemi (programmed testing), kişiselleştirilmiş test yöntemi (individualized testing)'dir. Ölçme aracı bireye göre uyarlandığı için, bireyselleştirilmiş (adaptive) terimi daha yaygın olarak kullanılmaktadır.

BBT, 1905 yılında Binet ile başlayan bireyselleştirilmiş test uygulamalarının bir serisidir. Bireyselleştirilmiş testler, iteratif bir süreç olup; madde bankalarındaki maddeler kullanılarak gerçekleştirilmektedir. Maddeler, bireyin yetenek düzeyi ile eşleşecek şekilde seçilmektedir. Bu süreçte ilk olarak, bankadaki maddelerden orta derecede zorluğa sahip olan biri seçilmekte ve birey bu soruya yanıt vermektedir. Verilen yanıta göre bireyin yetenek düzeyine ilişkin tahmin elde edildikten sonra, bankadaki hangi maddenin bireyin yetenek tahmini için en fazla bilgi sağlayacağına karar verilmektedir. Bir sonraki adımda, maddelere verilen yanıtlara dayalı olarak, yeni bir yetenek tahmini hesaplanmaktadır. Bireyin yetenek düzeyi optimum düzeyde kestirildiği durumda test sona ermektedir (Weiss, Kingsbury, 1984).

BBT uygulamaları madde bankalarının kalibrasyonu, ilk maddenin/başlama kuralının belirlenmesi, madde seçme yöntemleri, yetenek kestirim yöntemleri ve durdurma kuralları bölümlerini içermektedir (Reckase, 2009).

Madde Bankası Kalibrasyonu

Madde bankası, bireyselleştirilmiş test uygulamasında kullanılacak maddelerden oluşmaktadır. Temel veya işevuruk olmak üzere iki çeşidi bulunmaktadır (van der Linden, 2005). BBT uygulamalarında her bireye farklı sayıda sorular sorulmaktadır. Bu yüzden madde bankalarının kalitesi BBT uygulamaları sonucu elde edilen performansı/yeteneği doğrudan etkilemektedir. Madde bankası yüzlerce maddeden oluşabilmekte fakat bankadaki maddeler aynı ölçek üzerinde ifade edilemeyecek durumda olabilmektedir. Böyle bir durumda maddeler ölçülmek istenen örtük özelliği doğru bir şekilde yansıtmayabilmektedir. Bu bakımdan madde bankasındaki tüm maddelerin kalibrasyonları önem kazanmaktadır. Birçok MTK modeli yardımıyla kalibre edilmiş madde parametreleri oluşturmak mümkün olmaktadır (Hambleton, Swaminathan, Rogers, 1991).

İlk Madde / Başlama Kuralı

BBT uygulamalarında giriş maddesinin uygulama öncesinde belirlenmesi gerekmektedir. Çoğunlukla, BBT uygulamalarında bireyin yeterliliği hakkında ön bilgi bulunmadığı durumda ilk madde $\theta = 0$ olacak şekilde seçilir ve genel olarak evren ortalamasını yansıtmaktadır (Mills ve Stocking, 1996). Araştırmacılar, pratikte ilk maddenin seçilmesi ile ilgili bir takım durumlar göz önüne alınmasını önermektedir. Düşük ya da yüksek yetenek düzeyine sahip gruplarda seviyelerinin altında ya da üstünde maddeler verildiğinde bireylerin test puanları bu durumdan olumsuz etkilenebilmektedir (Thissen ve Mislevy, 2000).

Madde Seçme Kuralı

BBT uygulamalarındaki en krtik bölümlerden birisi de madde seçme yöntemleridir. Literatürde en sık kullanılan iki yöntemi vardır: maksimum bilgi yöntemi ve Bayes yöntemi (Thissen ve Mislevy, 2000).

Maksimum bilgi yönteminde bireyin o anki yetenek düzeyinde madde bilgisi en fazla olan maddeler seçilmektedir. Seçilen her madde ile bilgi artırıldığından, daha az madde ile θ kestirimlerinin doğruluğu artmaktadır. Bayes yöntemine dayanan diğer madde seçme yöntemi aynı zamanda en yüksek sonsal doğruluk kestirimi (maximum posterior precision selection) olarak da adlandırılmaktadır (Owen, 1975). Bu madde seçme yönteminde her bir madde sonsal yetenek dağılımını maksimize edecek şekilde seçilmektedir. Bu yöntemin en büyük avantajı, ilk

başlarda oluşan yetenek kestirimdeki büyük miktardaki hata değerleri sorununun üstesinden gelmesidir. Yöntem, sonsal bir dağılım kullanıldığından tek bir nokta kestirimine göre daha az hata üretmektedir. Bu yöntemde hesaplamaların fazla zaman alması en büyük dezavantaj olarak belirtilmiştir (Thissen ve Mislevy, 2000). Literatürde bu iki yöntemin yanında en yüksek global bilgi kriteri (Chang, Ying, 1996), Olabilirlik ağırlıklandırmalı bilgi kriteri (Veerkamp, Berger, 1997) gibi farklı yöntemler de geliştirilmiştir (van der Linden, Pashley, 2010).

Yetenek Kestirim Yöntemi

BBT uygulamalarında bireyin cevapladığı her sorudan sonra yeteneği kestirilmektedir. Bireyin soruları cevapladıktan sonra elde ettiği geçici (interim) θ değeri, bir sonraki maddenin seçimi için önemlidir. Bireyin toplam yetenek kestirimi testten aldığı tüm sorulara verdiği yanıtlara göre belirlenmektedir. Son yetenek kestirimleri farklı yetenek kestirimleri kullanılarak aynı metrik üzeride ifade edilebilir (Thissen, Mislevy, 2000). BBT yetenek kestirimleri iki farklı şekilde yapılmaktadır: En yüksek olabilirlik kestirimi ve Bayes kestirimi. En yüksek olabilirlik kestiriminde madde parametrelerinin bilindiği durumda bireylerin θ düzeyleri olabilirlik fonksiyonundan kestirilebilir. Olabilirlik fonksiyonu $L(\theta)$, madde tepki fonksiyonlarının çarpımına eşittir. Olabilirlik fonksiyonunun belirlenen θ değerindeki en yüksek değeri, olabilirlik fonksiyonunun doğal logaritma değerinin birinci türevinin sıfır olduğu noktaya eşitlenerek bulunabilir. Mükemmel cevapların (tüm soruların doğru ya da tüm soruların yanlış) olduğu durumlarda en yüksek olabilirlik yetenek kestirimi hatalı kestirimler yapmaktadır. Bu durumda olabilirlik fonksiyonunun önsel dağılıma ait olasılık fonksiyonuna oranı üzerinden kestirimler yapan Bayes yöntemleri kullanılabilmektedir. θ önsel olasılık dağılımı standart normal dağılım olarak varsayılmaktadır. θ önsel olasılık dağılımının en yüksek değerini bulması durumunda en yüksek sonsal dağılım (Maximum a Posteriori-MAP), ortalamasını bulması durumda ise beklenen sonsal dağılım (Expected a Posteriori-EAP)yöntemleri kullanılarak yetenekler kestirilmektedir (Owen, 1975).

Durdurma Kuralı

BBT uygulamaları belirli sayıda madde uygulandıktan sonra veya bireylerin ölçme kestirimleri (measurement precision) istenilen noktaya ulaştığı durumda durdurulmaktadır. Sonlandırma durumları testin amacına, madde bankasının

özelliğine ve içerik kısıtlamalarına göre çeşitli farklılık göstermektedir (Segall, 2010). Sonlandırma kuralları iki ana kategoriye ayrılmaktadır. Sabit ve değişen test uzunluğu. Sabit uzunluktaki BBT uygulamalarında önceden belirlenen madde sayısı kadar madde bireye uygulanmaktadır. Bu sayede her birey aynı sayıda soruya cevap vermektedir. Sabit uzunluklu test kullanımı, madde bankası kullanım oranlarını artırması ve kolay uygulanması bakımından uygulayıcılara avantaj sağlamaktadır. Değişen test uzunluğuna sonlandırma kuralının uygulandığı durumda cevap örüntülerine bağlı olarak bireylere farklı sayıda madde uygulanmaktadır. Bu yöntemin kullanıldığı BBT uygulamalarında bireylerin yetenekleri daha az madde ile ölçülmekte, uygulama zamanı azalmakta ve madde bankasının etkililiği artmaktadır (Weiss ve Kingsbury, 1984).

Çok Boyutlu Bireyselleştirilmiş Bilgisayarlı Testler

Bireyselleştirilmiş testlerde çok boyutlulukla ilgili ilk çalışmalar Bloxom ve Vale (1987), Fan ve Shu (1996), Luecht (1996), Segall (1996) ile başlamış daha sonraları van der Linden (1999; 2005), Mudler ve van der Linden (2009) ile devam etmiştir. Yapılan ilk çalışmalar yetenek kestirimleri ve madde seçme yöntemleri üzerine yoğunlaşmıştır. Daha sonraları Wang ve Chen (2004), çok boyutlu test yapılarının yetenek kestirimleri üzerindeki etkilerini incelemişledir. Yakın zamanda yapılan çalışmalar, madde seçme yöntemlerinin yetenek kestirimlerine etkisi ile (Wang ve Chang, 2011; Yao, 2012, 2014), ÇBBBT uygulamalarında farklı durdurma kurallarının etkisi üzerinde yoğunlaşmaktadır (Wang, Chang ve Boughton, 2011; Yao, Pommerich, Segall, 2014).

TBBBT ve ÇBBBT madde seçme yöntemleri kestirdikleri yetenek sayılarının farklı olması bakımından ayrılmaktadır. Özellikle ÇBBBT madde seçme yöntemleri, çoklu yeteneklerin kestirimlerine getirdikleri farklı teknikler ile alt boyut puanlarının hesaplanmasında daha kararlı sonuçlar elde etmeye çalışmışlardır. Segall (1996) gerçek madde bankalarında ÇBBBT uygulamalarının üçte bir daha az madde kullanarak daha doğru kestirimler yaptığını belirtmiştir. Luecht (1996), Segall'in yaklaşımını geliştirerek, birden fazla zorunlu kısıtlamanın kullanıldığı sertifika testlerinde ÇBBBT uygulaması sonucu daha az madde ile daha güvenilir alt boyut puanları elde etmiştir. Her iki araştırma da ÇBBBT uygulamalarının tek boyutlu ve çok sayıda maddeden oluşan TBBBT uygulamalarından daha etkili olduğunu belirtmişlerdir. Ancak araştırmada her bir alt test ayrı olarak kalibre edilmesi ve

basit yapıda ele alınması bir takım kısıtlamalara yol açmaktadır. Testler ayrı olarak kalibre edildiğinde çok boyutlu yetenek dağılımlarının birlikte kestirimleri mümkün olmamaktadır ve çok boyutlu örtük özellikler arasındaki korelasyon matrisleri Lord ve Novick (1968) tarafından önerilen klasik hataları içeren (disattenuation) formül kullanılarak hesaplanmaktadır.

ÇBBBT, alt boyut puanlar ya da tanılayıcı puanlar rapor etmede TBBBT uygulamalarına göre daha avantajlıdır. ÇBBBT uygulamalarında her bir alt yeteneğe/kapsama göre puanlar daha az madde ile kestirilebilmektedir ve her boyuta ait puanlar rapor edilebildiğinden dolayı bireylerin zayıf ve güçlü yanları boyutlara göre rahatlıkla belirlenebilmektedir (Segall, 2010; Wang ve Chang, 2011).

Hem TBBBT hem de ÇBBBT uygulamalarında alt boyut ve toplam puanlarını kestirmede yetenek kestirim yöntemleri doğrudan sonuçları etkilese de, uygun maddelerin seçilmemesi durumunda hiçbir yetenek kestirim yöntemi fonksiyonel olmayacaktır (Reckase, 2009). Maddelerin çok zor, çok kolay ve düşük bilgi verici maddeler arasından seçilmesi yetenek kestirimlerini etkilemektedir. Bunun yanında madde bankasında bulunan maddelerin psikometrik özellikleri de yetenek kestirimlerine doğrudan etki etmektedir (Veldkamp, van der Linden, 2010).

Maddelerin bankadan seçme işlemindeki kurallar, ÇBBBT uygulamalarında önem kazanmaktadır. Literatürdeki madde seçme yöntemleri θ kestirimde kullanılacak bazı kritik değerleri maksimize ya da minimize etme ilkesine dayanmaktadır. Madde seçme yöntemleri de kritik değerleri tanımlama konusunda birbirinden farklılaşmaktadır.

ÇBBBT Madde Seçme Yöntemleri

Fisher Bilgi Matrisinin Determinantının Artırtılması-Hacim Yöntemi (Maximizing the Determinant of the Fisher Information Matrix-Volume or D-Optimally)

Segall (1996) önceki araştırmalarda (Bloxom & Vale, 1987; Tam, 1992) çok boyutlu madde tepki kuramı çerçevesinde yeteneğin ortak dağılımına ait önsel bilgilerin kullanılmadığı için sonuçların geçerli sayılamayacağını belirtmiştir. Segall'e göre ÇBBBT her bir alt boyuttan belirli sayıda madde seçme yerine her bir alt boyutun özelliğini etkili şekilde ortaya çıkaracak madde seçme prosedürleri

sağlayabilmektedir. ÇBBBT uygulamalarında ayrıca boyutlar arasındaki ilişkiler dikkate alındığından madde seçme prosedürlerinin etkililiği daha da artırılabilmektedir. Segall (1996) Bayes modellemesine dayalı, yeteneğin ortak dağılımına ait önsel bilgileri de dikkate alan bir madde seçme yöntemi önermiştir. Yöntem, bireylerin alt boyut yeteneklerini ÇBMTK modelleri yardımıyla seçilen maddelerden $(k-1)$ kestirmektedir. Bu maddelerden elde edilen bilgi $(I_{k-1}(\theta^{k-1})$ önsel dağılım olarak kullanılmakta ve bir sonraki maddenin (k) seçiminde kullanılmaktadır. Bu sayede yetenek kestirimlerinin (θ^{k-1}) doğruluğunun artırıldığı belirtilmektedir. Sonsal bilgi dağılımının determinantını maksimize eden eşitlik denklem 1.10'da gösterilmiştir.

$$W = |I_{k-1}(\theta^{k-1}) + I_k(\theta^{k-1}) + \sum^{-1}| \qquad (1.10)$$

Madde bankasındaki her i maddesi için, hacim ya da bilgi fonksiyonunun determinantı denklem 1.11'deki eşitlik ile hesaplanabilmektedir (Yao, 2012).

$$W_m = \left| I_{k-1}(\theta^{k-1}) + \frac{(P_{i1}-\beta_{3i})^2(1-P_{i1})}{P_{i1}(1-\beta_{ik})^2}\beta_{2i}x\hat{\beta}_{2i} + \sum^{-1} \right| \qquad (1.11)$$

Kullback-Leibler

Kullback-Leibler (KL) bilgisinin TBBBT uygulamalarında ilk olarak Chang ve Ying (1996) tarafından kullanılmıştır. Veldkamp ve van der Linden (2002) KL bilgisini gölge test yöntemi (shadow test method) kullanarak çok boyutlu yapıya uyarlamışlardır. KL madde seçme yöntemi gerçek yetenek (θ_0) ile kestirilen yetenek (θ) arasındaki iki olabilirlik arasındaki uzaklığı ölçmektedir. M3PL model için KL bilgisi denklem 1.12 ile gösterilmiştir.

$$K_i(\theta,\theta_0) = P_i(\theta_0)Ln\left[\frac{P_i(\theta_0)}{P_i(\theta)}\right] + [1-P_i(\theta_0)]Ln\left[\frac{1-P_i(\theta_0)}{1-P_i(\theta)}\right] \qquad (1.12)$$

Denklem 1.12'de $i = (1,2,\dots,N)$ madde bankasindaki N sayıdaki maddeyi belirtmektedir. θ_o değeri, θ_o bilinmediğinde ve θ tanımlanmadığı durumda, sonsal beklenen KL bilgisine göre kestirilmektedir. θ sonsal dağılımının yoğunluğu $f(\theta\,|u_{i_1},\dots,u_{i_{k-1}})$ ile tanımlanmaktadır ve uygulanan $k-1$ sayıdaki maddenin bir fonksiyonudur. Sonsal beklenen KL bilgisi kullanılarak $\hat{\theta}^{k-1}$ kestirimi denklem 1.13 kullanılarak hesaplanmaktadır.

$$K_i^B(\hat{\theta}^{k-1}) \equiv \int_\theta K_i\left(\theta,\hat{\theta}^{k-1}\right)f(\theta\,|u_{i_1},\dots,u_{i_{k-1}})\partial\theta \qquad (1.12)$$

Çok boyutluluk açısından bakıldığında KL denklemindeki θ ve θ_0 değerleri ÇBMTK modelinde skaler yerine vektörel olarak ifade edilmektedir.

ÇBBBT uygulamalarında KL madde seçme yönteminin kullanılmasının iki temel nedeni bulunmaktadır. İlk olarak, tek boyutlu KL madde seçme yönteminde gerçek θ değerlerinin kestiriminde Fisher bilgisinden daha başarılıdır. Ayrıca KL bilgisi önsel dağılımları kullandığından gerçek ve kestirilen yetenekleri geçerli bir şekilde ayırt etmektedir. KL yöntemi, Fisher yönteminin aksine θ ve θ_0 değerlerinin birer fonksiyonu olarak ifade edilebilir ve yetenek seviyelerinin birbirine yakın olmasını gerektirmez (Chang ve Ying, 1996; Veldkamp ve van der Lİnden, 2002).

Minimum Hata Varyansı Kriteri

Hata varyanslarının doğrusal birleşimlerinin minimize eden (V1) madde seçme yöntemi, eşit ağırlıklandırılmış boyutlardan elde edilen toplam puanlara en düşük hata varyansı veren maddeyi seçmektedir. Bu yöntem toplam puanların doğruluğunu artırmak için van der Linden (1999) tarafından ortaya atılmıştır.

van der Linden (1999) çok değişkenli bilgi matrisinin yerine asimptotik varyans-kovariyans matrisinin kullanılmasındaki amacın madde seçme yöntemini çok boyutlu MTK'ya göre uyarlamak olduğunu belirtmiştir. Bireyselleştirilmiş test uygulamalarında kestirilen yetenek, her bir madde seçiminden sonra elde edilen yeteneklerin $(\lambda(\theta_1, \dots, \theta_m);\ \lambda = \lambda_j = (\lambda_1, \lambda_2, \dots, \lambda_m);\ \lambda_j \geq 0)$ doğrusal kombinasyonlarına eşittir. Ağırlıklandırmanın (λ) değeri testin amacına göre değişmekte ve bu değer, BBT prosedürlerini ve yetenek kestirimlerini değiştirebilmektedir.

İlk olarak prosedür MAP yöntemi kullanarak yetenek kestirimi ile başlamaktadır. Yerel bağımsızlık varsayımından dolayı $\widehat{\lambda'\theta} = \lambda'\theta$ olarak ifade edilebilir ve θ olabilirlik fonksiyonu denklem 1.14 yardımı ile hesaplanabilmektedir.

$$g\left(\theta \middle| u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}\right) = \frac{L\left(\theta \middle| u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}\right) g(\theta)}{\int L\left(\theta \middle| u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}\right) g(\theta) d\theta} \qquad (1.14)$$

Bireyselleştirilmiş test algoritmasında herhangi iki yetenek değişkenine ait madde seçme prosedürü için senaryo belirtilen şekildedir: $k-1$ madde seçilmiş olsun. $S_k = (i_1, i_2, \dots, i_{k-1})$ seçilen maddeleri; $R_k = (1,2, \dots, i)/S_k$ ise reddedilen maddeleri

göstermektedir. Bireyselleştirilmiş test algoritması kullanılarak $k-1$ sayıda madde uygulandıktan sonra k maddesi denklem 1.15'teki kritere göre seçilmektedir.

$$min_{R_k}\left[Var(\lambda\hat{\theta}_1^k + (1-\lambda)\hat{\theta}_2^k | \hat{\theta}_1^{k-1}, \hat{\theta}_2^{k-1}\right] \quad (1.15)$$

Denklem 1.15'te $\hat{\theta}_1^k$ ve $\hat{\theta}_2^k$, θ_1 ve θ_2 değerlerinin kestirimlerini göstermektedir. Denklem incelendiğinde madde $\lambda\hat{\theta}_1^k + (1-\lambda)\hat{\theta}_2^k$ varyansını minimize edecek şekilde seçilmektedir. R_k kümesindeki maddeleri seçmek için, final denklemi madde parametrelerini $(a_{i1}, a_{i2}\ ve\ d_i)$ ve olasılık $[P_i(\theta_1, \theta_2)]$ değerlerini içermektedir. Özetlenecek olursa, algoritma ilk olarak testin amacını yansıttığı düşünülen ağırlıklandırılmış değerlerini (λ) seçmektedir. Ağırlıklandırılmış deneysel veya seçilen problemden elde edilmiş olabilmektedir. En son adımda ise yetenek parametreleri MAP eşitlikleri kullanılarak kestirilmektedir.

V1 yöntemini kullanmanın bir takım avantajları vardır (van der Linden, 1999; Yao, 2012). Bazı durumlarda madde bankası birden çok yeteneği ölçecek şekilde desenlenmiştir. Bu yetenekler arasında ilişkiler olabileceği gibi herhangi bir ilişkiye rastlanmıyor olabilir. Bu durumda bireysel yetenekler ağırlıklandırma seçimini etkileyebilmektedir. Bu bakımdan V1 yöntemi alt boyut puanlarının doğrusal kombinasyonlarını düzenleyerek toplam puanların daha doğru kestirilmesini sağlamaktadır.

ÇBBBT Yetenek Kestirim Yöntemleri

ÇBBBT uygulamalarında en yüksek olabilirlik (Maximum Likelihood Estimation-MLE) yetenek kestiriminin kullanımına ilk olarak Segall'in (1996, 2010) çalışmalarında rastlanmıştır. MLE kestiriminde var olan θ aralıklarında olabilirlik fonksiyonunu maksimize eden (en yüksek noktasını bulan) nokta belirlenmektedir.

$$\hat{\theta}_{u_{i_1},\dots u_{i_k}}^{MLE} \equiv argmax_\theta\left\{L\left(\theta \middle| u_{i_1}, \dots, u_{i_k}\right) : \theta \in (-\infty, \infty)\right\} \quad (1.16)$$

$$L\left(\theta \middle| u_{i_1}, \dots, u_{i_k}\right) = \prod_{i=1}^{n-1} P(\theta|u_i)^{u_i}\, Q(\theta|u_i)^{1-u_i} \quad (1.17)$$

Bayes yaklaşımında ise denklem 1.14'te gösterildiği gibi sonsal dağılıma dayalı olarak noktasal şekilde belirlenebilmektedir. Bayes yöntemi ile bilinmeyen yetenek parametresine $[g(\theta)]$ ait önsel bir dağılım olduğu varsayılmaktadır. Olabilirlik fonksiyonu ve önsel dağılımın oluşturacağı θ sonsal dağılımı denklem 1.14 ile hesaplanabilmektedir. Genellikle elde edilen sonsal dağılım tek biçimlilik özelliği göstermektedir. Eğer dağılım testi alan bireylerden elde edilmiş ise, yetenek

dağılımının evrendeki bireylerden elde edilen deneysel kestirimi olarak da varsayılabilmektedir. van der Linden ve Pashley (2010), evren dağılımının genellikle normal dağılım özelliği gösterdiğini belirtmiştir. Bayes yaklaşımında θ sonsal dağılımını maksimize eden iki yöntem bulunmaktadır. MAP yöntemi dağılımdaki mod (en yüksek) değeri, EAP ise ortalama değeri belirtmektedir.

$$\hat{\theta}^{MAP}_{u_{i_1},\dots u_{i_k}} \equiv argmax_{\theta}\{g(\theta|u_{i_1},\dots,u_{i_k}): \theta \in (-\infty,\infty)\} \quad (1.18)$$

$$\hat{\theta}^{EAP}_{u_{i_1},\dots u_{i_k}} \equiv \int \theta \; g(\theta|u_{i_1},\dots,u_{i_k}) \, d\theta \quad (1.19)$$

ÇBBBT Durdurma Kuralları

Madde bankasındaki maddelerin kullanımlarını etkililiği artırmak ve daha az soru ile daha kesin yetenek kestirimleri yapmak adına TBBBT uygulamalarında farklı durdurma kuralları uygulanmaktadır. Madde bankasının ortalama güçlük seviyesinin üstünde yeteneğe sahip bireylerin yeteneklerinin kestiriminde sabit uzunluklu testlerin daha az kesin sonuçlar verdiği belirtilmiştir (Wang, Chang ve Boughton, 2011). Bu bakımdan son zamanlarda yapılan araştırmalarda sabit uzunluk durdurma kuralı yerine daha fazla bilgi verdiği belirtilen minimum determinant (K-Kuralı), minimum özdeğer (E-Kuralı), minimum iz (T-Kuralı) ve Kullback-Leibler farklılık kuralı (KL-Kuralı) gibi durdurma kriterleri üzerine çalışmalar yapılmaya başlanmıştır. Reckase (2009), ÇBBBT araştırmalarında farklı durdurma kurallarının incelenmesinin önemini vurgulamıştır. Literatürde ÇBBBT uygulamalarında farklı durdurma kurallarının etkisi üzerinde karşılaştırmalı araştırmalar yapılmaktadır (Wang, Chang ve Boughton, 2011; Yao, Pommerich, Segall, 2014).

2. İLGİLİ ARAŞTIRMALAR

Alan yazınına bakıldığında ÇBBBT uygulamalarına ilişkin farklı sayıda birçok araştırma bulunmasına karşın ülkemizde bu konuya ait çalışmalara bir çalışma dışında rastlanmamıştır. Bu bölümde ÇBBBT uygulamalarının son 10 yıldaki gelişimine üzerine yapılan araştırmalar ele alınmaktadır. İncelenen araştırmalarda ÇBBBT madde seçme yöntemleri, yetenek kestirimleri, durdurma kuralları ve çok boyutlu test yapıları üzerinde yapılan çalışmalara ilişkin bilgi verilmiştir.

Diao (2009) ÇBBBT uygulamalarında D-Optimality, A-Optimality, Fisher ve KL madde seçme yöntemleri ile MLE ve Bayes yetenek kestirim yöntemlerini farklı koşullar altında karşılaştırmıştır. Araştırmanın yönteminde ise test uzunluğu (uzun ve kısa) ve Bayes dağılımları için önsel dağılıma koşulları (güçlü, orta ve gerçek önsel dağılım) manipüle edilmiştir. Araştırmada Michigan Eğitim Değerlendirme Programına (MEDP) ait testle elde edilen gerçek veriler kullanılmıştır. MEDP testi 50 maddeden oluşmaktadır ve 3 boyutlu bir yapıya sahiptir. Araştırmanın sonuçları incelendiğinde özellikle kısa testlerde Bayes yetenek kestirim yönteminin MLE yetenek kestirim yönteminden daha iyi performans gösterdiği belirlenmiştir. Madde seçme yöntemlerine bakıldığında ise KL yönteminin diğer yöntemlerden daha iyi sonuçlar verdiği ancak hesaplamalarının daha yavaş olduğu belirtilmiştir.

Song (2010) araştırmasında TBMTK modelinin çok boyutlu veriye uyumunu farklı kapsam dengeleme prosedürleri kullanarak BBT uygulamalarında incelemiştir. Araştırmada maksimum bilgi madde seçme yöntemi kullanılarak kısıtlama olmadan, kısıtlamalı BBT (constarined CAT-CCAT), modifiye edilmiş model (modified multinominal model-MMM) ve modifiye edilmiş kısıtlamalı BBT (modified constrained CAT-MCCAT) olmak üzere dört farklı kapsam dengelemesi yöntemi karşılaştırılmıştır. Karşılaştırmalar ölçme kesinliği, madde bankası kullanımı ve madde kullanım sıklığı üzerinden yapılmıştır. Araştırmanın yönteminde çok boyutlu madde yapısı, yetenek dağılımı ve konu kapsamlarının zorluk düzeyleri manipüle edilmiştir. Araştırma bulguları incelendiğinde konu kapsamı zorluk düzeylerinin dengesiz dağıldığı durumda kapsam dengeleme yöntemlerinin maksimum bilgi yönteminden daha iyi performans gösterdiği belirlenmiştir. Madde bankası kullanımı ve madde sıklığı bakımından kapsam dengeleme yöntemleri arasında bir

farklılık gözlenmemiştir. Genel olarak bakıldığında ise MMM yönteminin madde bankasını kullanmada en etkili yöntem olduğu belirtilmiştir.

Wang ve Chang (2011) çalışmalarında dört farklı madde seçme yönteminin teorik bağlantılarını ortaya çıkararak, simülasyon çalışmaları ile performanslarını karşılaştırmışlardır. Araştırmada iki boyutlu test yapısında M3PL modeli kullanılmış; D-optimality, KL Bilgi İndeksi, Sürekli Entropi Yöntemi ve Ortak Bilgi madde seçme yöntemleri karşılaştırılmıştır. Bireylere ait yetenekler EAP yöntemi ile kestirilmiştir. Araştırmada 450 maddelik bankadan 25 maddelik testler oluşturulacak şekilde ÇBBBT uygulaması yapılmıştır. Ortak bilgi madde seçme yönteminde alt boyutlara ait ortalama hata karelerinin en az şekilde kestirildiği belirlenmiştir.

Yao (2012) çalışmasında alt boyut ve toplam puanların kestiriminde beş farklı madde seçme yöntemini madde bankalarının yapısına, bireylerin dağılımına, seçilen madde sayısına ve konu kapsamına göre karşılaştırmıştır. Araştırmada en minimum açı (Ag), Hacim (Vol), minimum hata varyansı (V1), ağırlıklandırılmış minimum hata varyansı (V2) ve Kullback-Leibler (KL) madde seçme yöntemleri karşılaştırılmıştır. Araştırma ile madde bankasından en yüksek oranda madde seçerek alt boyut ve toplam yetenek puanlarını en iyi şekilde kestiren madde seçme yönteminin belirlenmesi amaçlanmaktadır. Yöntemler mutlak yanlılık, korelasyon, test güvenirliği madde kullanımı ve test süresine göre karşılaştırılmıştır. Dört farklı alt boyuttan ve toplam 912 maddeden oluşan madde bankasından 18, 36 ve 55 madde uzunluğunda olacak şekilde sabit uzunluklu ÇBBBBT uygulaması gerçekleştirilmiştir. Araştırmada üç farklı kapsam dengelemesi (C0, C1 ve C2) kullanılmıştır. Madde seçerken kapsam dengelemesinin kullanılmadığı (C0) durumda Vol ve Ag yöntemlerinin daha çok madde seçtiği ve alt boyut ve yetenekleri puanlarını daha iyi kestirdiği belirlenmiştir. Her bir boyuttan belirli sayıda maddenin seçilmesi gereken durumda (C1) ise Vol ve Ag'nin performansının düştüğü, V1, V2 ve KL yöntemlerinin hem alt boyut hem de toplam puanları kestirmede daha iyi sonuçlar verdiği görülmüştür. Toplam puanların kestiriminde KL ve V2 yönteminin V1 ve Vol yöntemlerinden daha iyi sonuçlar verdiği belirlenmiştir.

Wang, Chang ve Boughton (2013) araştırmalarında ÇBBBT uygulamalarında sabit uzunlukta durdurma kuralları yerine ölçme doğruluğunu artıracak şekilde durdurma

kuralları üzerinde çalışmışlardır. Araştırmada minimum determinant (D-Kuralı), minimum özdeğer (E-Kuralı), minimum iz (T-Kuralı) ve Kullback-Leibler farklılık kuralı (KL-Kuralı) olmak üzere dört farklı durdurma kuralı karşılaştırılmıştır. Araştırma üç farklı simülasyon deseni üzerinden yürütülmüştür. İlk simülasyon deseninde 450 maddeden oluşan 2 boyutlu madde bankası hazırlanmıştır. Bankadaki madde parametreleri, her bir boyut önceden belirlenmiş olmak koşuluyla çok değişkenli normal dağılıma göre üretilmiştir. İkinci simülasyon deseninde madde bankası, matematik testinden elde edilen 200 maddeden oluşmaktadır. Bankadaki maddeler doğrulayıcı faktör analizi ve M3PL modeli kullanılarak parametreleri kestirilmiş; dış ankor eş zamanlı kalibrasyon yöntemi ile de kalibre edilmiştir. Üçüncü koşulda ise 480 maddeden oluşan üç boyutlu madde bankası geliştirilmiş ancak diğer simülasyonların aksine üçüncü boyut istenen bir boyut olarak tasarlanmamıştır (nuisance). Sonuçlar incelendiğinde bireyin gerçek yeteneği ne olursa olsun dört durdurma kuralı da belirlenen hata aralığında testi başarılı bir şekilde durdurmuşlardır. D, E ve T kuralı kullanıldığında uç yetenekteki bireylere, orta yetenekteki bireylerden iki kat daha fazla soru sorulduğu belirlenmiştir. Ancak KL Kuralında test uzunluğunun pek fazla değişmediği gözlenmiştir.

Yao (2013) yaptığı araştırmada farklı durdurma kurallarına ve değişen test uzunluklarına göre ÇBBBT madde seçme yöntemlerini karşılaştırmıştır. Araştırmada sabit madde kullanım sıklığı kontrol prosedürü ile kapsam dengeleme için öncelik indeksi (priority index) kullanılmıştır. Araştırmacı ÇBBBT uygulamalarında kullanılmak üzere iki farklı durdurma kuralı sunmuştur. Standart hata ve tahmini standart hata azaltması durdurma kurallarına göre minimum açı (Ag), Hacim (Vol), minimum hata varyansı (V1), ağırlıklandırılmış minimum hata varyansı (V2) ve Kullback-Leibler (KL) madde seçme yöntemlerinin performansı incelenmiştir. ÇBBT uygulaması sonucu elde edilen sonuçlar alt boyut ve toplam yetenek puanlarının kesinliği, test uzunlukları ve test güvenirlikleri açısından karşılaştırılmıştır. Araştırma sonuçları incelendiğinde KL yönteminin en az sayıda madde kullanmasına rağmen en iyi performansı verdiği belirtilmiştir. En çok madde ise Ag yönteminde kullanılmıştır. Durdurma kuralları incelendiğinde en iyi performansı standart hata yönteminin verdiği belirtilmiş, madde seçme yöntemleri arasında tüm koşullarda KL yönteminin kullanılması önerilmiştir.

Yao (2014) madde kullanım sıklığının kontrol edildiği durumlarda, madde bankasının kullanım sıklığını artıran ve en doğru şekilde alt boyut ve toplam yetenek puanlarını en iyi derecede kestiren madde seçme yöntemini ÇBBBT uygulamaları kullanarak belirlemeye çalışmıştır. Araştırmada en düşük açı, hacim, minimum hata varyansı, ağırlıklandırılmış minimum hata varyansı ve Kullback-Leibler madde seçme yöntemleri karşılaştırılmıştır. Araştırmanın yöntemi incelendiğinde madde kullanım sıklığına belirlemede Sypmson-Hetter ve eşit oran prosedürü, kapsam dengelemesi için ise maksimum öncelik indeksi (maximum priatory index) kullanılmıştır. Araştırmada dört alt boyuttan oluşan ve basit yapıda olan Askeri Servisler Mesleki Kabiliyet Testinin (Armed Services Vocational Aptitude Battery-ASVAB) testinden elde edilen veriler kullanılmıştır. Madde bankası 912 maddeden oluşmaktadır. Araştırmada alt boyutlar arası korelasyon 0.6 olarak belirlenmiş; test uzunlukları 18, 36 ve 55 olacak şekilde uygulanmış ve yetenek kestirim olarak MAP yöntemi kullanılmıştır. Araştırmada değerlendirme kriteri olarak gerçek yetenek ve kestirilen yetenekler arası korelsayon, yanlılık (BIAS) ve mutlak yanlılık (ABSBIAS) değerleri kullanılmıştır. Araştırma sonuçları incelendiğinde madde kullanım sıklığı oranları artırıldığında ABSBIAS değerlerinin her koşulda azaldığı belirlenmiştir. Özellikle sabit oranlı madde kullanım sıklığı yönteminde KL madde seçme yönteminin en iyi performansı gösterdiği belirlenmiştir. Madde kullanım sıklıkları içinde en iyi performans Sypmson-Hetter yönteminde görülmüştür. Ancak bu yöntemde madde seçme yöntemlerinin birbirlerine göre üstünlükleri bulunamamış, hepsi benzer performans sergilemiştir. Madde sayısı arttığında korelasyonların arttığı ve mutlak yanlılık değerlerinin ise azaldığı belirlenmiştir. Toplam yetenek puanlarına ait mutlak yanlılık değeri ise tüm koşullarda alt boyut yetenek puanları mutlak yanlılık değerlerinden düşük olarak kestirilmiştir.

Yao, Pommerich ve Segall (2014) yaptıkları araştırmalarda tek boyutlu olarak uygulanan Askeri Servisler Mesleki Kabiliyet Testinin (Armed Services Vocational Aptitude Battery-ASVAB) çok boyutlu BBT şeklinde uygulandığında puan kestirimlerinde kayıplar yaşamadan test süresine nasıl etki ettiğini incelemişlerdir. Bu bakımdan 3 farklı madde bankasında test durdurma kuralı, madde seçme yöntemleri, kapsam dengelemesi, zaman kısıtlaması ve evren dağılımı koşulları manipüle edilerek sonuçlar ÇBBBT uygulaması çerçevesinde değerlendirilmiştir.

Araştırma sonuçları değerlendirildiğinde ÇBBBT uygulamasında kapsam dengelemesi ve cevaplama zamanını aynı anda dengeleyebildiği belirtilmiş ve ÇBBT uygulamasının dikkate değer sonuçlar ortaya koyduğu vurgulanmıştır. ÇBBBT yetenek kestirimlerinde boyutlar arasındaki korelasyonlar arttığında dört alt boyutun ve toplam yeteneğin gerçek ve kestirilen yetenekleri arasındaki korelasyonların da arttığı belirtilmiştir. Test cevaplama süresinin öncelik olarak ele alınması durumunda ÇBBBT koşulunda boyut başına düşen madde sayısının en az 3 olması gerektiği belirtilmiştir.

Lee (2014) araştırmasında yüksek düzeyli ve hiyerarşik MTK modellerinin farklı koşullarda ÇBBBT algoritmaları üzerindeki tutarlığını incelemiştir. Bu amaçla 6 farklı durum ve 192 çapraz koşul oluşturulmuştur. Araştırmada 4 farklı ÇBMTK test deseni (bifaktör, iki-katmanlı (two-tier), tek birincil faktörlü yüksek düzeyli ve iki birincil faktörlü yüksek düzeyli), iki birincil faktörlü düzeye ilişkin 3 farklı korelasyon değeri (düşük, orta, yüksek), birincil faktöre düşen grup faktör sayısı (iki ve dört), 3 farklı test uzunluğu (40, 80 ve 160), 2 çok boyutlu madde seçme yöntemi (en yüksek Fisher indeksi ve Bayes) ile 2 farklı yetenek kestirim yöntemi (MLE ve EAP) koşulları çaprazlanmıştır. Araştırma sonuçları incelendiğinde test uzunluğu arttığında korelasyonların arttığı, RMSE değerlerinin ise azaldığı belirtilmiştir. MTK modelleri içinde ise yüksek düzeyli modeller en düşük performansı göstermiştir. Tüm koşullarda Bayes madde seçme yönteminin en düşük hata değerlerine sahip olduğu; kısa testlerde EAP yönteminin MLE yönteminden daha iyi performans gösterdiği belirtilmiştir.

Seo ve Weiss (2015) tarafından yapılan çalışmada üç farklı çok boyutlu bifaktör modelinde dört madde seçme yöntemi ÇBBBT uygulamaları çerçevesinde karşılaştırılmıştır. ÇBBBT uygulamasında sabit uzunluklu durdurma kuralı kullanılmıştır. Araştırmada 2 farklı çok boyutlu MTK modeli (M2PL, M3PL), 3 farklı bi-faktör örüntüsü, 4 madde seçme yöntemi (D-Optimality, D_s Optimality, A-Optimality, E-Optimality) ve 2 farklı yetenek kestirim yöntemi (MAP ve EAP) olmak üzere 72 koşul manipüle edilmiştir. Araştırmada bifaktör modellerinin belirlenmesinde her bir boyuta düşen *a*-parametrelerinin oranları manipüle edilmiştir. *a*-parametrelerinin boyutlara verdiği yükler belirlenirken MDISC değerleri dikkate alınmıştır. Araştırmanın bulguları incelendiğinde genel faktörü ölçmede D_s Optimality madde seçme yönteminin her üç bifaktör yapısında da daha az hata ile

kestirimler yaptığı belirlenmiştir. Ancak grup faktörlerinde her üç bifaktör modeli için de A-Optimality ve D-Optimality yöntemlerinin en az hata içerdiği ve birbirine benzer sonuçlar verdiği belirlenmiştir. Yetenek kestirim yöntemleri incelendiğinde MAP yönteminin tüm koşullarda EAP kestiriminden daha düşük hata içerdiği belirlenmiştir.

Su (2016) ÇBBBT uygulamalarında dört farklı madde seçme yöntemini basit ve karmaşık test yapılarında karşılaştırmıştır. Yao tarafından geliştirilen çok boyutlu öncelik indeksin (multidimensional priority index-MPI) basit yapıda (between-item) çok boyutlu BBT uygulaması üzerinden geliştirdiğini, ancak testlerin her zaman bu yapıda olmadığını ve karmaşık yapıda çok boyutlu BBT uygulamalarının geliştirilmesi gerektiğini belirtmiştir. Bu amaçla modifiye edilmiş çok (modified multidimensional priority index-MMPI) boyutlu öncelik indeksi geliştirilerek madde seçme yöntemleri karşılaştırılmıştır. Araştırmada basit ve karmaşık yapıdaki madde bankası için 2 boyuttan oluşan 1000 madde simüle edilmiştir. Boyutlar arasındaki korelasyonlar ise 0.4 ve 0.8 olarak belirlenmiştir. Sonuçlar ölçme kesinliği, madde kullanım sıklığı ve kapsam dengeleme bakımından karşılaştırılmıştır. Araştırma sonuçları incelendiğinde tüm madde seçme yöntemlerinde testin karmaşıklığı arttığında yanlılık ve hata değerlerinin azaldığı görülmüştür. Ayrıca boyutlar arası korelasyonlar arttığında da yanlılık ve hata değerlerinin düştüğü belirlenmiştir.

Yurt içinde yapılan çalışmalar genelde TBBBT uygulamaları üzerine yoğunlaşmıştır. ÇBBBT üzerine yapılan çalışmada Özdemir (2015) basit ve karmaşık yapılarda bireylerin yabancı dil yeterliklerini farklı yetenek kestirim, madde seçme ve sonlandırma kurallarını manipüle ederek karşılaştırmıştır. A-optimality madde seçme yöntemi kullanıldığı durumda test uzunluğu ve hata değerlerinin azaldığı, her bir boyuta ait test güvenirliğinin değerlerinin ise arttığı belirlenmiştir. Yetenek kestirimi yöntemlerinde ise en iyi performansı MAP yönteminin gösterdiği belirtilmiştir.

2.1. İlgili Araştırmalar Özet

ÇBBBT üzerine yapılan çalışmalar genelde madde seçme yöntemlerinin karşılaştırılması ve performanslarının incelenmesi üzerine yoğunlaşmaktadır. (Diao, 2009; Wang ve Chang, 2011; Yao, 2012, 2013; 2014). Yapılan çalışmalara bakıldığında farklı amaçlara göre geliştirilen çok boyutlu madde seçme

yöntemlerinin alt boyut ve toplam puanları kestirmedeki performansları, madde bankası ve madde kullanımındaki oranları sıklıkla karşılaştırılmıştır.

ÇBBBT uygulamalarında son zamanlarda artan bir başka ilgi alanı da farklı durdurma kurallarının performanslarının incelenmesidir. Wang, Chang ve Boughton (2013), Yao (2013) ve Yao, Pommerich ve Segall (2014) yaptıkları araştırmalarda sabit uzunluklu madde kullanılmasının yanında değişen uzunlukta madde kullanılmasının etkilerinin araştırmışlardır.

Lee (2014), Seo ve Weiss (2015) ve Su (2016) çalışmalarında ise tek boyutlu test yapılarından öte farklı çok boyutlu yapıların ÇBBBT madde seçme yöntemleri üzerindeki etkilerini incelemişlerdir. Yapılan araştırmalarda çok boyutlu madde bankalarının oluşturulmasında kalibrasyon yöntemlerinin etkisini inceleyen, basit ve karmaşık çok boyutlu test yapılarını karşılaştıran bir çalışmaya rastlanmamıştır.

3. YÖNTEM

3.1. Araştırmanın Türü

Bu araştırmada, çok boyutlu test yapılarında alt test puanları ve toplam puanları belirlemede kullanılan madde seçme yöntemlerinin farklı koşullar altındaki performansları ortaya konması amaçlanmıştır. Araştırmada var olan yöntem ve tekniklerin gerçek veri üzerinden sınanması amaçlandığından araştırma temel araştırma niteliğindedir (Karasar, 2010).

3.2. Araştırmanın Deseni

Araştırmada test deseni, boyut başına düşen madde sayısı, kalibrasyon yöntemi ve madde seçme yöntemleri olmak üzere dört farklı durum manipüle edilmiş ve Çizelge 3.1'de gösterilmiştir. Manipülasyonların sonucunda 9x2x3x3 olmak üzere toplam 162 koşul çapraz olarak test edilmiştir.

Çizelge 3.1. Araştırma Deseni

Test Deseni	*Boyut Başına Düşen Madde Sayısı*	*Kalibrasyon Yöntemi*	*Madde Seçme Yöntemi*
Basit Yapı (BY) ρ=0.2 Basit Yapı (BY) ρ=0.5 Basit Yapı (BY) ρ=0.8 Düşük Karmaşık Yapı (DKY) ρ=0.2 Düşük Karmaşık Yapı (DKY)) ρ=0.5 Düşük Karmaşık Yapı (DKY) ρ=0.8 Yüksek Karmaşık Yapı (YKY) ρ=0.2 Yüksek Karmaşık Yapı (YKY) ρ=0.5 Yüksek Karmaşık Yapı (YKY) ρ=0.8	10, 15	Ayrı Tek Boyutlu Kalibrasyon (ATB) BAEM Kalibrasyonu MHRM Kalibrasyonu	Kullback-Leibler (KL) Hacim (Vol) Minimum Hata Varyansı (V1)

3.3. Verilerin Üretilmesi

3.3.1. Madde Parametrelerinin Üretilmesi

Araştırmada kullanılan madde parametrelerinin üretilmesinde PISA 2012 Türkiye verisinden yararlanılmıştır. PISA 2012 Türkiye verisinden kestirilen parametre değerleri doğrultusunda eğim, güçlük ve düşük asimptot parametreleri *a, b ve c*, 3 parametreli lojistik model kullanılarak *a ~ LN{0, 0.2), b ~ N(0, 1), ve c ~ Beta{6,16)* koşullarını sağlayacak şekilde üretilmiştir. Tek boyutlu test deseni için üretilen madde parametreleri Ek 3'te verilmiştir. *LN{0, 0.2)* koşulunda elde edilen *a*-parametreleri log-normal dağılıma sahip, ortalaması 0, standart sapması 0.2 olan dağılımı belirtirken, *b*-parametresi ortalaması 0, standart sapması 1 olan normal

dağılımı, *c*-parametresi ise 6 ve 16 değerlerine ait *Beta* dağılımını belirtmektedir. Ampirik olarak elde edilen madde parametresi değerlerine bağlı kalmak koşuluyla a ve b parametreleri [0.5, 1.5] ve [-2, 2] arasında değerler alıncaya kadar yeniden üretilmiştir. Düşük asimptot değeri olan c-parametresi 0.15 değerine sabitlenmiştir.

3.3.2. Test Desenlerinin Oluşturulması

Araştırmada çok boyutlu test desenleri, tek boyutlu olarak üretilen a-parametrelerinden yararlanılarak ve

$$a_j = MDISC = \sqrt{a_1^2 + a_2^2 + a_3^2} \qquad (3.1)$$

formülü kullanılarak belirlenmiştir. Yukarıda belirlenen bileşke a-parametresi değeri bir maddenin çok boyutlu ayırt ediciliğini (ÇBAE-MDISC) göstermektedir ve ayırt ediciliğin en yüksek değerini belirtir ($\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_3}$ ayırt edicilik parametreleridir) (Ackerman, Gierl ve Walker, 2003). MDISC değeri çok boyutlu ayırt ediciliğin tek boyutlu halidir.

Araştırmada çok boyutlu test yapılarından basit yapı, düşük karmaşık yapı ve yüksek karmaşık yapı olmak üzere üç farklı test deseni çok boyutlu ayırt edicilik değerlerinin sabit tutulması koşuluyla *a*-parametresi değerlerinin boyutlara dağıtılmasıyla belirlenmiştir. Belirlenen test desenleri Ek 4'te gösterilmiştir.

Çok boyutlu test desenleri belirlenirken bazı adımlar izlenmiştir. İlk olarak, basit yapılı test deseninde her bir maddeye ait *a*-parametresi değerleri sadece belirli bir boyuta yük verirken diğer boyutlara hiç yük vermemiştir. Bu bakımdan madde birincil boyut için ayırt edici iken diğer boyutlar için ayırt edici değildir. Literatürde bu durum bağımsız kümeleme (independent cluster) (McDonald, 2000) olarak isimlendirilmiş olsa da diğer araştırmacılar (Gierl, Leighton & Tan, 2006; Zhang, 1996) bu yapıyı basit olarak adlandırmada bir sorun görmemişlerdir. Düşük karmaşık ve yüksek karmaşık yapılara ait kesin kurallar olmamakla birlikte genellikle birincil boyutun ağırlık olarak 0.1 ve 0.4'ü belirlenebilmektedir. Bu kapsamda düşük karmaşık yapılı test deseni oluşturulurken birincil faktöre verilen yük toplam yük değerinin %70 civarına sabitlenip, diğer iki faktöre verilen yük ise %21 ve %9 olarak belirlenmiştir. Yüksek yapılı test deseninde ise birincil faktöre verilen yük %50 civarına sabitlenmiş, diğer iki faktöre verilen yük ise %35 ve %15 olarak belirlenmiştir.

Bu tür karmaşık bir yapının oluşturulmasındaki ikinci amaç ise ÇBMTK veri setleri üretilirken kullanılan telafi edici modelin etkisinin test edilmek istenmesidir. Telafi edici modelde bir boyuttaki yüksek yetenek diğer boyutlardaki düşük yetenekleri telafi edebilir (Ackerman, 1987; 1994).

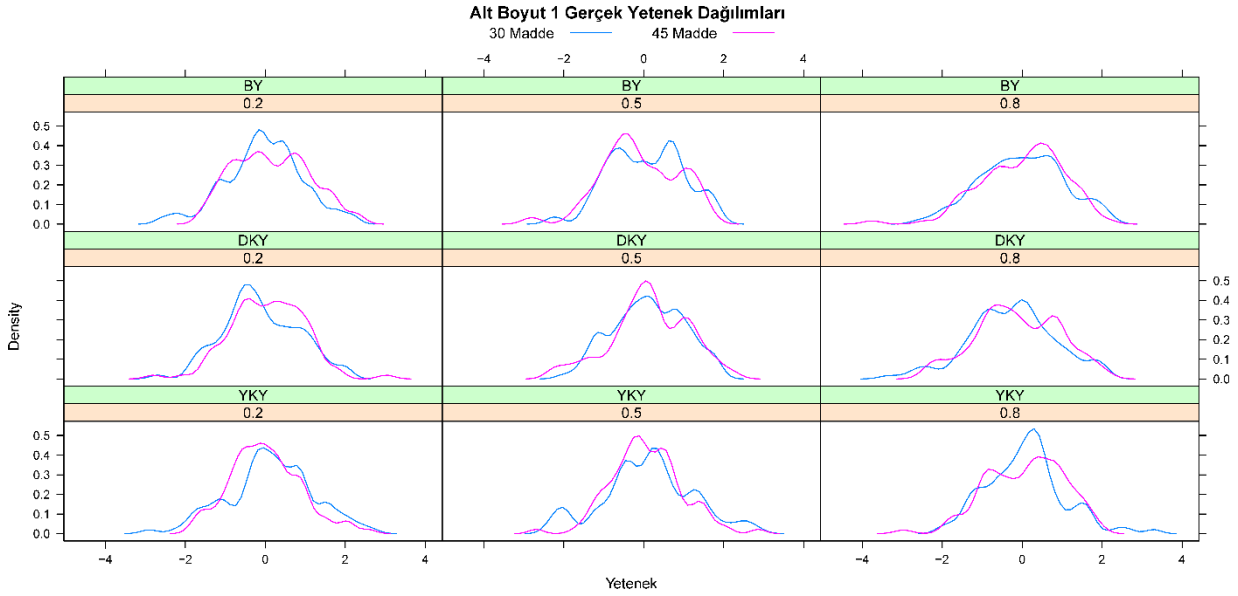
3.3.3. Yetenek Parametrelerinin Üretilmesi

BY, DKY ve YKY test desenleri için 1000 birey ve 3 alt boyuttan oluşan 1000x3 gerçek yetenek parametreleri matrisi, çok değişkenli normal dağılıma göre ($\boldsymbol{\theta}_{\mathrm{i}} = \boldsymbol{MVN}(\mathbf{0}, \sum)$) aşağıda belirtilen ortalama vektörleri ve varyans-kovariyans matrisleri kullanılarak üretilmiştir.

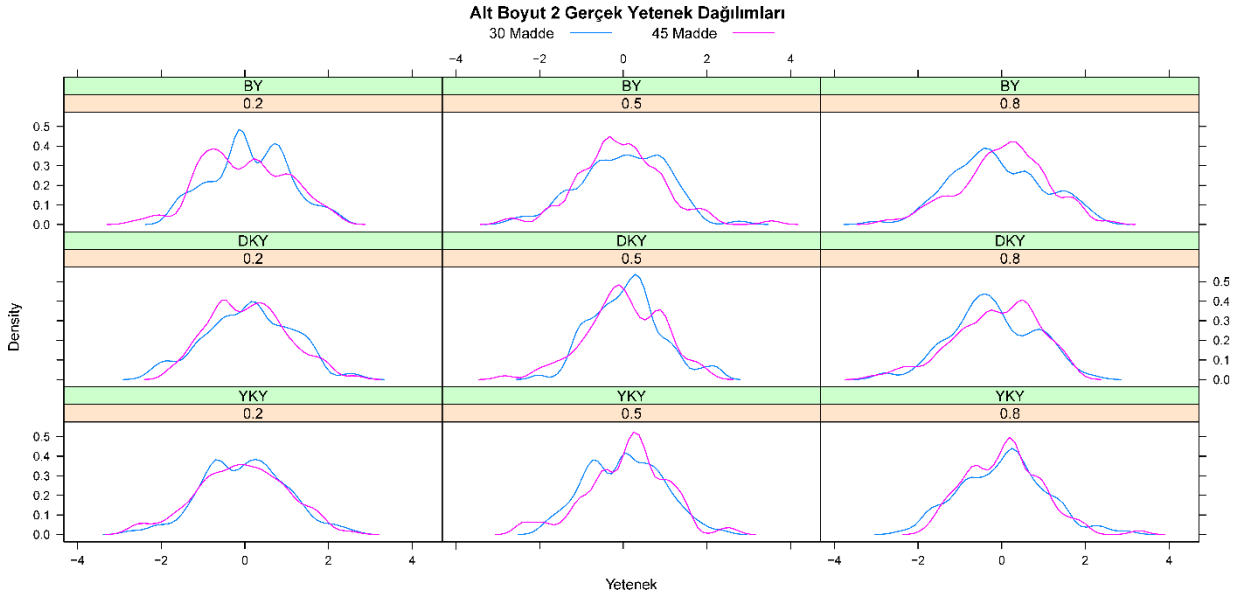
$$\mu = [0 \quad 0 \quad 0 \quad 0] \qquad \mu = [0 \quad 0 \quad 0 \quad 0] \qquad \mu = [0 \quad 0 \quad 0 \quad 0]$$

$$\sigma^2 = \begin{bmatrix} 1 & & \\ 0.2 & 1 & \\ 0.2 & 0.2 & 1 \end{bmatrix} \qquad \sigma^2 = \begin{bmatrix} 1 & & \\ 0.5 & 1 & \\ 0.5 & 0.5 & 1 \end{bmatrix} \qquad \sigma^2 = \begin{bmatrix} 1 & & \\ 0.8 & 1 & \\ 0.8 & 0.8 & 1 \end{bmatrix}$$

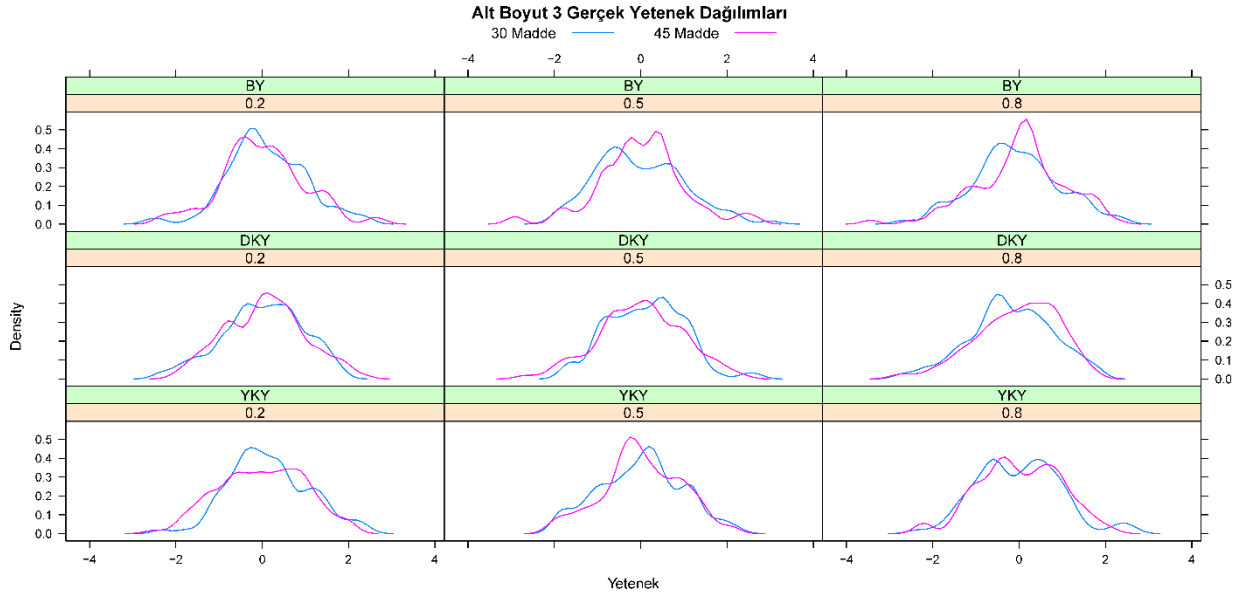
Simülasyon sonucu elde edilen madde ve yetenek parametreleri kullanılarak, boyutlar arası korelasyon değerleri ile birlikte, telafi edici çok boyutlu MTK modeline göre cevap matrisleri MIRTGEN 3.0 (Luecht, 2004) programı kullanılarak üretilmiştir. Cevap matrisleri üretilirken çok boyutlu 3 parametreli lojistik model kullanılmıştır (M3PL). Her bir boyuta düşen madde sayılarının manipüle edildiği durumda, üç farklı boyut arası korelasyon değerleri ve üç farklı test deseni için 1000x30 ve 1000x45 olmak üzere toplamda 18 cevap matrisi üretilmiştir. Üretilen cevap matrislerine ait yoğunluk grafikleri her bir alt boyut için Şekil 3.1-3.3'te gösterilmiştir.



Şekil 3.1. Alt Boyut 1'e ait Yetenek Dağılımları



Şekil 3.2. Alt Boyut 2'ye ait Yetenek Dağılımları



Şekil 3.3. Alt Boyut 3'e ait Yetenek Dağılımları

3.4. Kalibrasyon Yöntemleri

3.4.1. Ayrı Tek Boyutlu Kalibrasyon

Ayrı Tek Boyutlu (ATB) Kalibrasyon her bir alt boyutu bağımsız olarak ele alır ve boyutlar arasındaki ilişkiyi yok sayar. Her bir alt boyuta ait madde ve yetenek parametreleri 3 PL modele göre kestirilmiştir. Kalibrasyon sonucunda her bir madde için tek a-parametresi ve d-parametresi, her bir boyut için bir yetenek parametresi hesaplanmıştır. Örneğin, Alt Boyut 1'e ait parametreler, sadece o boyutun özelliklerini belirtmektedir ve diğer boyutlarla ya da testin genel yapısıyla ilgisi bulunmamaktadır. Tek boyutlu yetenek parametreleri -5.0 ve +5.0 aralığında 49 quadrature nokta belirlenerek MAP yetenek kestirim yöntemi kullanılarak hesaplanmıştır.

3.4.2. Bock & Aitkin Expectation-Maximization (BAEM) Algoritması

EM (Expectation-Maximization) tam verinin olabilirlik değerlerini en yüksek olabilirlik (MLE) yardımıyla hesaplayarak madde parametrelerini kestiren genel bir yöntemdir. EM algoritması iki adımdan oluşur: E (expectation)-Adımı ve M (maximization)-Adımı. EM algoritmasındaki temel amaç olabilirlik değerini olabildiğince artırmaktır. EM algoritmasının kullanılmasındaki en büyük avantaj, karmaşık yapıdaki en yüksek olabilirlik problemlerine çözümler bulmasıdır. Gözlenmeyen herhangi değişkenleri kayıp veri olarak ele alır. MTK tarafından bakılacak olursa yetenek parametresi ($\boldsymbol{\theta}$) hem gözlenebilen hem de

gözlenemeyen değişken olarak ele alınabilir. Bu durumda θ gözlenemeyen (tam) veriyi, $\boldsymbol{Y}$ ise gözlenen (tam olmayan) veriyi ifade eder. E-adımında beklenen log-olabilirlik değerleri bulunur. M-adımında ise E-adımında elde edilen log-olabilirlik değerlerini maksimize eden parametreler hesaplanır. Örneğin $\boldsymbol{U}$ gözlenen değerlerden oluşan veri matrisimiz olsun. $\boldsymbol{U_{ij}}$ matrisinde bireyler $\boldsymbol{i} = (\boldsymbol{1} \dots \boldsymbol{N})$ maddeler $\boldsymbol{j} = (\boldsymbol{1} \dots \boldsymbol{J})$ olarak verilmiş olsun. $\boldsymbol{\theta}$ gözlenmeyen örtük özellikler matrisini ifade eder ve $\boldsymbol{\theta_i}$ ise i sırasındaki bireye ait yetenek değerini gösterir. $\boldsymbol{\theta}$ değeri burada sürekli değil kesikli veri olarak varsayılmıştır.ψ ise kestirmeye çalıştığımız madde parametreleri matrisi olsun. Parametrelerin olabilirlik değerleri bireyler ve maddeler arasındaki olasılık fonksiyonlarının çarpımı olarak ifade edilebilir:

$$L(\psi) = \prod_{i=1}^{N} \prod_{j=1}^{J} P_j(\theta_i)^{u_{ij}} Q_j(\theta_i)^{(1-u_{ij})} \qquad (3.2)$$

E-adımında gözlenen ve kayıp verinin log-olabilirlik değerlerini buluruz.

$$q(\psi|\psi_n) = \varepsilon_{\theta|\psi_n}\{\ln P(U, \theta|\psi_n)\} \qquad (3.3)$$

M-adımında ise $q(\psi|\psi_n)$ noktasını maksimize eden ψ noktalarını hesaplanır. Tüm bunlar hesaplansa bile a_j, b_j ve c_j parametreleri henüz hesaplanmamış haldedir. Bu noktada da Newton-Raphson prosedürleri kullanılarak parametreler kestirilebilir.

Bock ve Aitkin (1981) madde parametrelerinin kestiriminde farklı bir MML yöntemi kullanmışlardır. Yukarıda belirtilen yöntemde her bir madde parametresinin log-olabilirlik değerlerinin türevleri tüm madde parametrelerine bağlı olarak hesaplanmaktadır. Bu durumun üstesinden gelmek için Bock ve Aitkin iterarif bir yöntem izlemişlerdir. İlk adımda yetenek parametrelerinin dağılımını hesaplamak için ilk başlarda kestirilen madde parametreleri kestirimleri ve ona bağlı cevaplar matrisi (U_{ij}) ve parametre matrisi (ψ_n) kullanılmıştır. İkinci adımda ise elde edilen bu dağılımdan yararlanılarak yeni madde parametreleri kestirilmiştir. İkinci adımda her bir madde parametresinin log-olabilirlik değerlerinin türevleri sadece o madde parametrelerine bağlı olarak hesaplanmaktadır.

ÇBMTK modellerinde marjinal en yüksek olabilirlik kullanılırken (MML) çok sayıda Gaussian quadrature noktalarının integralinin alınması gerekmektedir. Bu durum da hesaplama işlemlerini yoğunlaştırmakta ve fazla zaman almaktadır (Cai, 2010a;

Cai, 2010b). Örneğin, her bir boyut için 8 quadrature nokta belirlendiği durumda 2 ve 3 boyutlu testler için hesaplaması yapılacak toplam qaudrature nokta sayısı $8^2 = 64$ ve $8^3 = 512$ olacaktır. Hâlihazırdaki bilgisayar programlarında quadrature noktalarda kısıtlamaya gitmek gibi bir durum olmamasına rağmen hesaplamalar uzun süreler alabilmektedir. Quadrature noktaların daha az seçilmesi durumunda da yakınsama (converge) problemleri ortaya çıkmaktadır. Bu durum "boyutluluğun laneti" olarak adlandırılmaktadır (Baskhov, 2015). BAEM kalibrasyonu -5.0 ve +5.0 aralığında her bir boyuta 21 quadrature nokta belirlenerek MAP yetenek kestirim yöntemi kullanılarak hesaplanmıştır.

3.4.3. The Metropolis-Hastings Robbins-Monro (MH-RM) Algoritması

Son zamanlarda Cai (2008) tarafından boyutluluğun lanetini ortadan kaldırmak amacıyla Metropolis-Hastings Robbins-Monro algoritması önerilmiştir. MHRM yaklaşımında BAEM kalibrasyonlarında olduğu gibi integrasyon ve türev alma işlemleri bulunmamaktadır. Temelde modele ait parametreler gözlenen bir dağılımdan kestirilmektedir. Belirtilen gözlenen dağılım, Markov Zinciri Monte Karlo yöntemleri kullanılarak oluşturulmuş simülasyondan oluşmuş bir dağılımdır; gerçekten veriden oluşturulmuş gözlenen bir dağılımı belirtmez.

MHRM algoritması, Fisher'in özdeşlik (Identity) yöntemini kullanarak Markov Zincirleri Monte Karlo yöntemi ile Robbins-Monro stokastik yakınsama algoritmasının birlikte kullanılmasına dayanır. Fisher'in özdeşlik yöntemi en yüksek olabilirlik parametre kestirimi olarak da adlandırılır. MHRM koşullu beklenen sonsal dağılımı denklemine doğrudan dahil etmediğinden çok-parametreli optimizasyon problemlerine neden olmamaktadır. Aksine, çok sayıda madde ve çoklu boyutlar ile başa çıkabilmek için yakınsama işlemini hızlandırmak adına eğrisel bilgiler (curvature information) kullanmaktadır (Cai, 2008). MHRM kalibrasyonu -5.0 ve +5.0 aralığında her bir boyuta 21 quadrature nokta belirlenerek MAP yetenek kestirim yöntemi kullanılarak hesaplanmıştır.

3.5. Değerlendirme Kriteri

Her bir ÇBBBT koşulu için, alt boyutların ve genel yetenek parametrelerinin kesinliğini belirlemede dört farklı kriter kullanılmıştır: Pearson momentler çarpımı korelasyon katsayısı, hata kareleri ortalamasının karekökü (root mean square error-RMSE), yanlılık (BIAS) ve mutlak yanlılık (ABSBIAS). Pearson Momentler

Çarpımı Korelasyon Katsayısı denklem 3.4'te verilen formül yardımıyla hesaplanmıştır:

$$r = \frac{\sigma_{\hat{\theta}\theta}}{\sigma_{\hat{\theta}}\sigma_{\theta}} \qquad (3.4)$$

Denklem 3.4'te $\sigma_{\hat{\theta}\theta}$ gerçek ve kestirilen puanlar arasındaki kovariyansı $\sigma_{\hat{\theta}}$ ve σ_{θ} ise gerçek ve kestirilen puanların standart sapmalarını belirtmektedir. RMSE, BIAS ve ABSBIAS değerleri ise aşağıda belirtilen formüller ile hesaplanmıştır:

$$RMSE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(\hat{\theta}_i - \theta_i)^2}$$

$$BIAS = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta_i) \qquad (3.5)$$

$$ABSBIAS = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{\theta}_i - \theta_i\right|$$

Denklem 3.5'deki N, her bir koşuldaki birey sayısını; $\hat{\theta}_i$ i'inci sıradaki bireyin kestirilen puanını ve θ_i i'inci sıradaki bireyin gerçek puanını belirtmektedir. Mutlak yanlılık (Yao, 2012) yanlılık değerlerinin mutlak değer içindeki ifadesidir ve pozitif bir sayıdır ve ÇBBBT uygulamalarında alt boyut ve toplam yetenek puanlarının kesinliğini hesaplamada referans olarak kullanılması önerilmiştir.

3.6. Verilerin Analizi

Araştırmada madde bankalarının tek boyutlu ve çok boyutlu kalibrasyon işlemleri yazılımdan kaynaklı farklılıkların oluşmasını önlemek adına flexMIRT v3.0 (Cai, 2013) programı kullanılarak yapılmıştır. Kalibrasyon sonuçları ve kalibrasyondan elde edilen madde parametreleri kullanılarak, ÇBBBT alt boyut ve toplam yetenek puanlarının kestirimi SimuMCAT (Yao, 2011) programı kullanılarak hesaplanmıştır. SimuMCAT programı çok boyutlu BBT uygulamalarına elverişli Java temelli bir yazılımdır.

3.6.1. ÇBBBT Analizi

ÇBBBT analizlerinde yetenek kestirimleri ve madde seçme yöntemleri her boyuta aynı anda uygulanmıştır. Başlangıç maddesi $\boldsymbol{\theta}_{\boldsymbol{başlangıç}} = \{\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \boldsymbol{\theta_3}\} = \{\mathbf{0}, \mathbf{0}, \mathbf{0}\}$ koşulunu sağlayacak şekilde seçilmiştir. Test uzunlukları madde bankalarının uzunluklarına eşit olacak şekilde her bir boyut için toplamda 30 ve 45 olarak

belirlenmiştir ve sabit uzunluklu sonlandırma kuralı uygulanmıştır. Araştırmada ÇBBBT alt boyut ve toplam yetenek puanları MAP yöntemi kullanılarak 100 iterasyonun ortalaması hesaplanarak kestirilmiştir.

4. BULGULAR VE TARTIŞMA

Bu bölümde elde edilen bulgular üç aşamada ele alınmıştır. İlk iki bölümde araştırmanın alt problemlerinin daha iyi anlaşılması için gerekli olan ön bulgulara ait sonuçlar verilmiştir. İlk bölümde üretilen gerçek yetenek parametrelerine ait bulgular her bir koşul için değerlendirilmiştir. İkinci bölümde, tek boyutlu ve çok boyutlu kalibrasyon yöntemleri sonucu elde edilen alt boyut ve toplam yetenek puanlarına ait hata, yanlılık ve korelasyon değerleri test desenlerine, boyut başına düşen madde sayılarına ve boyutlar arası korelasyon koşullarına göre karşılaştırılmıştır. En son aşamada ise her bir alt probleme ait mutlak yanlılık, hata ve korelasyon değerleri tüm test koşulları için çok boyutlu madde seçme yöntemlerine göre karşılaştırılmıştır.

4.1. Gerçek Yetenek Parametrelerine Ait Bulgular

Gerçek yetenek parametreleri basit, düşük karmaşık ve yüksek karmaşık test desenlerine, boyutlar arasındaki korelasyona ve her bir boyuta düşen soru sayısına göre çok boyutlu normal dağılım kullanılarak üretilmiştir. Üretilen parametrelerin kesinliğini kontrol etmek amacıyla boyutlar arası korelasyon değerleri, ortama ve standart sapma değerleri her bir koşul için hesaplanmıştır. Çizelge 4.1 ve Çizelge 4.2 MIRTGEN (Luecht, 2004) programı kullanılarak üretilen gerçek yetenek parametrelerinin sonuçlarını özetlemektedir.

Çizelge 4.1. Boyutlar arası Korelasyon Değerleri Kestirimleri

Boyut Başına Düşen Madde Sayısı	*Test Deseni*	*Korelasyon*	ρ'_{12}	ρ'_{13}	ρ'_{23}
10 Madde	**BY**	$\boldsymbol{\rho = 0.2}$	0.191	0.229	0.247
		$\boldsymbol{\rho = 0.5}$	0.462	0.473	0.443
		$\boldsymbol{\rho = 0.8}$	0.814	0.827	0.806
	DKY	$\boldsymbol{\rho = 0.2}$	0.268	0.192	0.276
		$\boldsymbol{\rho = 0.5}$	0.508	0.548	0.528
		$\boldsymbol{\rho = 0.8}$	0.809	0.803	0.792
	YKY	$\boldsymbol{\rho = 0.2}$	0.148	0.221	0.227
		$\boldsymbol{\rho = 0.5}$	0.543	0.539	0.501
		$\boldsymbol{\rho = 0.8}$	0.803	0.802	0.781
15 Madde	**BY**	$\boldsymbol{\rho = 0.2}$	0.185	0.185	0.231
		$\boldsymbol{\rho = 0.5}$	0.526	0.490	0.495
		$\boldsymbol{\rho = 0.8}$	0.798	0.789	0.814
	DKY	$\boldsymbol{\rho = 0.2}$	0.209	0.173	0.194
		$\boldsymbol{\rho = 0.5}$	0.505	0.517	0.502
		$\boldsymbol{\rho = 0.8}$	0.798	0.813	0.816
	YKY	$\boldsymbol{\rho = 0.2}$	0.170	0.228	0.196
		$\boldsymbol{\rho = 0.5}$	0.492	0.516	0.498
		$\boldsymbol{\rho = 0.8}$	0.793	0.794	0.802

ρ: Veri üretimi sırasında kullanılan alt boyutlar arasındaki teorik korelasyon değeri, ρ': Kalibrasyon sonucu elde edilen boyutlar arası korelasyon değerleri

Çizelge 4.2. Gerçek Yetenek Puanlarına Ait Ortalama ve Standart Sapma Değerleri

			Altboyut 1		*Altboyut 2*		*Altboyut 3*	
Boyut Başına Düşen Madde Sayısı	***Test Deseni***	***Korelasyon***	***Ort***	***Ss***	***Ort***	***Ss***	***Ort***	***Ss***
10 Madde	**BY**	$\boldsymbol{\rho = 0.2}$	.006	.99	-.067	.99	-.059	.99
		$\boldsymbol{\rho = 0.5}$	.032	.97	.047	.96	-.012	.98
		$\boldsymbol{\rho = 0.8}$	-.048	1.02	-.058	1.01	-.058	1.02
	DKY	$\boldsymbol{\rho = 0.2}$	.013	.97	-.034	1.01	-.057	.99
		$\boldsymbol{\rho = 0.5}$	.000	.96	.028	1.02	-.014	.98
		$\boldsymbol{\rho = 0.8}$	-.005	.98	-.018	.98	.005	.98
	YKY	$\boldsymbol{\rho = 0.2}$	.084	1.01	.032	1.01	.051	.98
		$\boldsymbol{\rho = 0.5}$	.035	1.06	.034	1.01	-.001	.99
		$\boldsymbol{\rho = 0.8}$	.034	1.00	.044	1.01	.005	1.01
15 Madde	**BY**	$\boldsymbol{\rho = 0.2}$	.013	.99	-.009	1.01	.009	.99
		$\boldsymbol{\rho = 0.5}$	.006	1.01	-.026	1.04	.012	1.02
		$\boldsymbol{\rho = 0.8}$	-.004	.99	-.021	1.03	.002	1.01
	DKY	$\boldsymbol{\rho = 0.2}$	-.008	.98	.000	1.01	-.014	.94
		$\boldsymbol{\rho = 0.5}$	-.035	.96	-.012	1.01	-.048	1.04
		$\boldsymbol{\rho = 0.8}$	-.020	1.03	-.013	1.02	.009	1.03
	YKY	$\boldsymbol{\rho = 0.2}$	-.024	.99	-.034	1.01	-.052	.99
		$\boldsymbol{\rho = 0.5}$	-.061	.98	-.027	.99	-.001	.96
		$\boldsymbol{\rho = 0.8}$	-.015	1.01	-.021	.97	-.018	1.01

ρ: Veri üretimi sırasında kullanılan alt boyutlar arasındaki teorik korelasyon değeri

Çizelge 4.1 incelendiğinde üretilen yetenek parametrelerine ait korelasyon değerleri varsayılan (hipotetik) korelasyon değerlerine benzer olarak elde edilmiştir. Her bir alt boyuta ait üretilen gerçek yetenek puanlarına ait ortalamalar -0.067 ile 0.084 arasında, standart sapma değerleri ise 0.94 ile 1.06 arasında değişmektedir. Alt boyutlara ait gerçek puanlar, ortalaması 1, standart sapması 0 olan çok boyutlu normal dağılıma yakın değerlerde elde edilmiştir.

4.2. Tek ve Çok Boyutlu Kalibrasyona Ait Bulgular

Bu bölümde kalibrasyon yöntemlerinin alt boyut puanları ve toplam puanlar üzerindeki performansı, tek boyutlu kalibrasyon ve çok boyutlu kalibrasyon yöntemi kullanılarak, üç boyutlu veri seti üzerinden karşılaştırılmıştır. Ayrı tek boyutlu kalibrasyon (ATB) geleneksel tek boyutlu kalibrasyon yöntemidir ve her bir alt test için madde ve yetenek parametreleri ayrı olarak hesaplanır. BAEM yaklaşımı Bock ve Aitkin (1981) tarafından ortaya sunulan çok boyutlu kalibrasyon yöntemidir ve özünde EM algoritması kullanarak olabilirliğin en yüksek değerini bulmayı amaçlamaktadır. MHRM yaklaşımı ise en yeni yöntemlerden birisidir. Cai (2008) tarafından geliştirilen bu yöntem, Metropolis-Hasting örnekleminden rastgele impute edilen Robbins-Monro stokastik yakınsama algoritmasının birlikte kullanılmasına dayanır. Gerçek parametrelerin üretilmesinden elde edilen 18 veri

seti (*Test Deseni(3) x Boyutlar arası korelasyon (3) x Boyut başına düşen soru sayısı (2)*) için alt boyut ve toplam yetenek parametreleri üç kalibrasyon yöntemi için flexMIRT v3.0 (Cai, 2013) kullanılarak hesaplanmıştır. ATB, BAEM ve MHRM kalibrasyonları sonucunda elde edilen üç alt boyut ile bir toplam yetenek kestirimleri $RMSE_{Kal}$, yanlılık ($BIAS_{Kal}$) ve korelasyon ($Korelasyon_{Kal}$) değerleri hesaplanarak gerçek yetenek değerleri ile karşılaştırılmıştır.

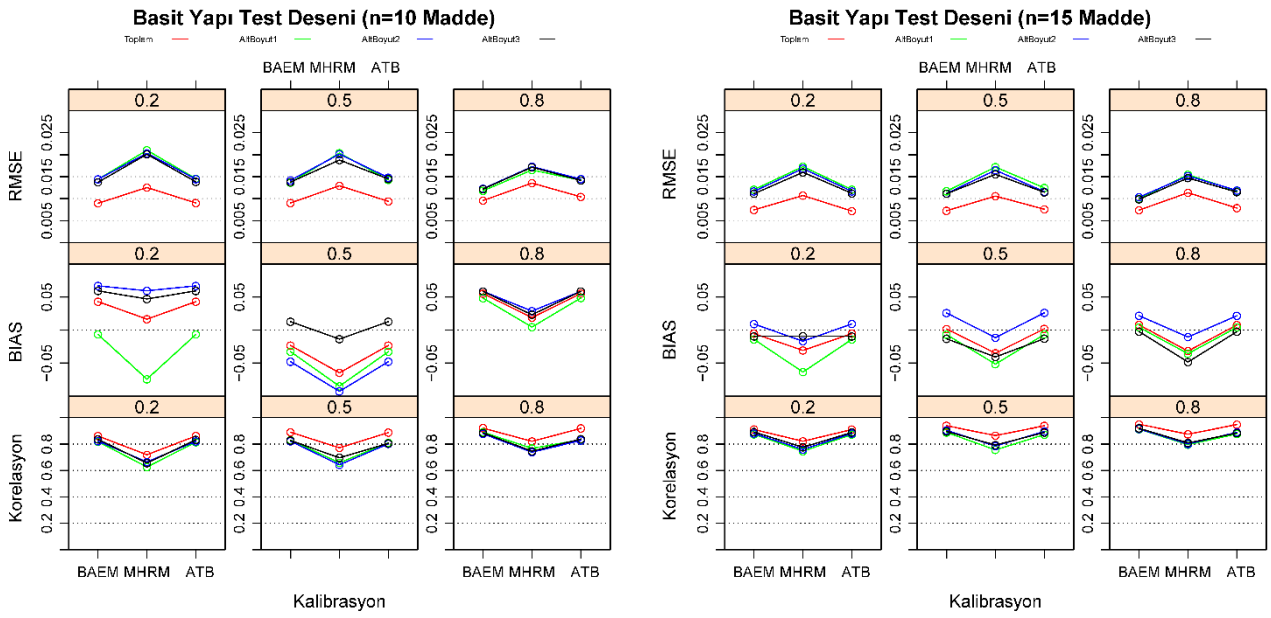
4.2.1. Basit Yapılı (BY) Testlerde Yetenek Puanlarının Karşılaştırılması

Çizelge 4.4 incelendiğinde basit yapılı (BY) test deseninde tüm yetenek puanları kestiriminde en yüksek $RMSE_{Kal}$ değerleri MHRM yönteminde gözlenmiştir. BAEM ve ATB kalibrasyonları, MHRM yöntemine göre daha düşük $RMSE_{Kal}$ ile kestirilmiştir. Çizelge 4.3 ‘te gösterilen kestirilen puanlara ait $BIAS_{Kal}$ değerlerine bakıldığında ise durum farklılık göstermemektedir. En yüksek $BIAS_{Kal}$ değerleri MHRM kalibrasyonunda görülmüştür. Bu durumda hem alt boyut hem de toplam yetenek puanları kestirmede MHRM kalibrasyonun yanlı ve hatası yüksek kestirimler yaptığı sonucuna varılabilir. Çizelge 4.5’te gösterilen gerçek puanlar ile kalibre edilmiş puanlar arasındaki korelasyon ($Korelasyon_{Kal}$) değerlerine bakıldığında yine var olan tüm koşullarda MHRM kalibrasyonu diğer iki kalibrasyon yönteminden daha düşük değerler üretmiştir.

Şekil 4.1 inclendiğinde alt boyut yetenek puanlarının kestiriminde boyut başına düşen madde sayısı arttıkça $RMSE_{Kal}$ ve $BIAS_{Kal}$ değerlerinde azalma, $Korelasyon_{Kal}$ değerlerinde ise bir artma olduğu belirlenmiştir. Örneğin *Alt Boyut 1* için, maddeler arası korelasyonların 0.2 olduğu koşulda BAEM, MHRM ve ATB kalibrasyonları için madde sayısı arttıkça $RMSE_{Kal}$ ve $BIAS_{Kal}$ değerleri azalmakta, $Korelasyon_{Kal}$ değerleri artmaktadır. Bu artış içinde yine en yüksek $RMSE_{Kal}$ ve $BIAS_{Kal}$ değeri ile en düşük $Korelasyon_{Kal}$ değeri MHRM kalibrasyonunda görülmüştür. Toplam yetenek puanlarında ise durum alt boyut puanları ile benzerdir.

Testin çok boyutluluğu azaldıkça özellikle alt boyut yetenek puanlarının kestirimlerine ait $RMSE_{Kal}$ değerleri azalmaktadır; ancak $BIAS_{Kal}$ değerleri aynı örüntüyü izlememiştir. Özellikle madde sayısının düşük (n=10) ve testlerin çok boyutluluğun en zayıf olduğu korelasyon değerinde (p=0.8) yanlılık en düşük MHRM değerlerinde gözlenmiştir. Çok boyutluluğun artmasına bağlı olarak

$Korelasyon_{Kal}$ da ters orantılı olarak değişmektedir. Test ne kadar katı çok boyutluluk özelliği gösteriyorsa, $Korelasyon_{Kal}$ değerleri o kadar düşüktür demektir. Çok boyutluluğun ihlal edildiği durumlarda $Korelasyon_{Kal}$ değerleri artmaktadır. Bu durum madde sayısının arttığı koşullarda da geçerlidir.



Şekil 4.1. Basit Yapılı Testlerin $RMSE_{Kal}$, $BIAS_{Kal}$ ve $Korelasyon_{Kal}$ Değerlerinin Koşullara Göre Karşılaştırması

Çizelge 4.4 incelendiğinde basit yapılı bir testte toplam puanlara ait en düşük $RMSE_{Kal}$ değeri boyutlar arası korelasyonun en düşük olduğu ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM ve ATB kalibrasyonları için eşit olacak şekilde 0.07 olarak kestirilmiştir. Aynı yapıdaki testte alt boyut puanlarına ait en düşük $RMSE_{Kal}$ değeri boyutlar arası korelasyonun en yüksek olduğu ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM kalibrasyonunda 0.10 olarak kestirilmiştir. BY testteki toplam puanlara ait en yüksek $Korelasyon_{Kal}$ değeri boyutlar arası korelasyonun en yüksek olduğu ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM ve ATB kalibrasyonları için eşit olacak şekilde 0.95 olarak kestirilmiştir. Ayrıca elde edilen değerlere bakıldığında alt boyut yetenek puanlarını tüm koşullarda toplam yetenek puanlarından daha yüksek $RMSE_{Kal}$ ve $BIAS_{Kal}$, daha düşük $Korelasyon_{Kal}$ değerleri ürettiği gözlenmiştir.

4.2.2. Düşük Karmaşık Yapılı (DKY) Testlerde Yetenek Puanlarının Karşılaştırılması

Çizelge 4.4 inclendiğinde düşük karmaşık yapılı (DKY) test deseninde, BY desenine benzer olarak, tüm yetenek puanları kestiriminde en yüksek $RMSE_{Kal}$ değerleri MHRM yönteminde gözlenmiştir. BAEM ve ATB kalibrasyonları, MHRM yöntemine göre daha düşük $RMSE_{Kal}$ ile kestirilmiştir. Kestirilen puanlara ait $BIAS_{Kal}$ değerlerinde de benzer durum görülmeketdir. Çizelge 4.3 inclendiğinde en yüksek $BIAS_{Kal}$ değerleri MHRM kalibrasyonunda görülmüştür. Bu durumda hem alt boyut hem de toplam yetenek puanları kestirmede MHRM kalibrasyonun yanlı ve hatası yüksek kestirimler yaptığı sonucuna varılabilir. Çizelge 4.5'teki $Korelasyon_{Kal}$ değerlerine bakıldığında yine var olan tüm koşullarda MHRM kalibrasyonunun diğer iki kalibrasyon yönteminden daha düşük değerler ürettiği görülmüştür.

Şekil 4.2 incelendiğinde alt boyut yetenek puanlarının kestiriminde boyut başına düşen madde sayısı arttıkça $RMSE_{Kal}$ ve $BIAS_{Kal}$ değerlerinde azalma, $Korelasyon_{Kal}$ değerlerinde ise bir artma olduğu belirlenmiştir. Örneğin *Alt Boyut 1* için, maddeler arası korelasyonların 0.2 olduğu koşulda BAEM, MHRM ve ATB kalibrasyonları için madde sayısı arttıkça $RMSE_{Kal}$ ve $BIAS_{Kal}$ değerleri azalmakta, $Korelasyon_{Kal}$ değerleri artmaktadır. Bu artış içinde yine en yüksek $RMSE_{Kal}$ ve $BIAS_{Kal}$ değeri ile en düşük korelasyon değeri MHRM kalibrasyonunda görülmüştür. Toplam yetenek puanlarında ise durum alt boyut puanları ile benzerdir.

Çizelge 4.3. Gerçek ve Kalibre Edilmiş Puanlar Arası BIAS Değerleri

			Altboyut 1			*Altboyut 2*			*Altboyut 3*			*Toplam Puan*		
Boyut Başına Düşen Madde Sayısı	***Test Deseni***	***Korelasyon***	***BAEM***	***MHRM***	***ATB***	***BAEM***	***MHRM***	***ATB***	***BAEM***	***MHRM***	***ATB***	***BAEM***	***MHRM***	***ATB***
		$\rho = 0.2$	-0.01	-0.07	-0.01	0.07	0.06	0.07	0.06	0.05	0.06	0.04	0.02	0.04
	BY	$\rho = 0.5$	-0.03	-0.09	-0.03	-0.05	-0.09	-0.05	0.01	-0.01	0.01	-0.02	-0.06	-0.02
		$\rho = 0.8$	0.05	0.00	0.05	0.06	0.03	0.06	0.06	0.02	0.06	0.02	0.06	0.06
		$\rho = 0.2$	-0.01	-0.08	-0.01	0.03	0.02	0.03	0.06	0.04	0.06	0.03	-0.01	0.03
10 Madde	DKY	$\rho = 0.5$	0.00	-0.07	0.00	-0.03	-0.07	-0.03	0.01	-0.04	0.01	-0.01	-0.06	-0.01
		$\rho = 0.8$	0.00	-0.08	0.00	0.02	-0.06	0.02	-0.01	-0.06	-0.06	0.01	-0.07	-0.01
		$\rho = 0.2$	-0.08	-0.14	-0.08	-0.03	-0.04	-0.03	-0.05	-0.09	-0.05	-0.05	-0.08	-0.05
	YKY	$\rho = 0.5$	-0.04	-0.11	-0.04	-0.03	-0.07	-0.03	0.00	-0.04	0.00	-0.02	-0.07	-0.02
		$\rho = 0.8$	-0.03	-0.08	-0.04	-0.04	-0.06	-0.04	-0.01	-0.11	-0.01	-0.03	-0.09	-0.03
		$\rho = 0.2$	-0.01	-0.06	-0.01	0.01	-0.02	0.01	-0.01	-0.01	-0.01	-0.01	-0.03	-0.01
	BY	$\rho = 0.5$	-0.01	-0.05	-0.01	0.03	-0.01	0.03	-0.01	-0.04	-0.01	0.00	-0.03	0.00
		$\rho = 0.8$	0.00	-0.04	0.00	0.02	-0.01	0.02	0.00	-0.05	0.00	0.01	-0.03	0.01
		$\rho = 0.2$	0.01	-0.07	0.01	0.00	-0.10	0.00	0.01	-0.05	0.01	0.01	-0.07	0.01
15 Madde	DKY	$\rho = 0.5$	0.04	-0.10	0.04	0.01	-0.08	0.01	0.05	-0.05	0.05	0.03	-0.08	0.03
		$\rho = 0.8$	0.02	-0.13	0.02	0.01	-0.14	0.01	-0.01	-0.15	-0.01	0.01	-0.14	0.01
		$\rho = 0.2$	0.02	-0.07	0.02	0.03	-0.03	0.03	0.05	-0.03	0.05	0.04	-0.04	0.04
	YKY	$\rho = 0.5$	0.06	-0.07	0.06	0.03	-0.08	0.03	0.00	-0.15	0.00	0.03	-0.10	0.03
		$\rho = 0.8$	0.02	-0.14	0.02	0.02	-0.13	0.02	0.02	-0.12	0.02	0.02	-0.13	0.02

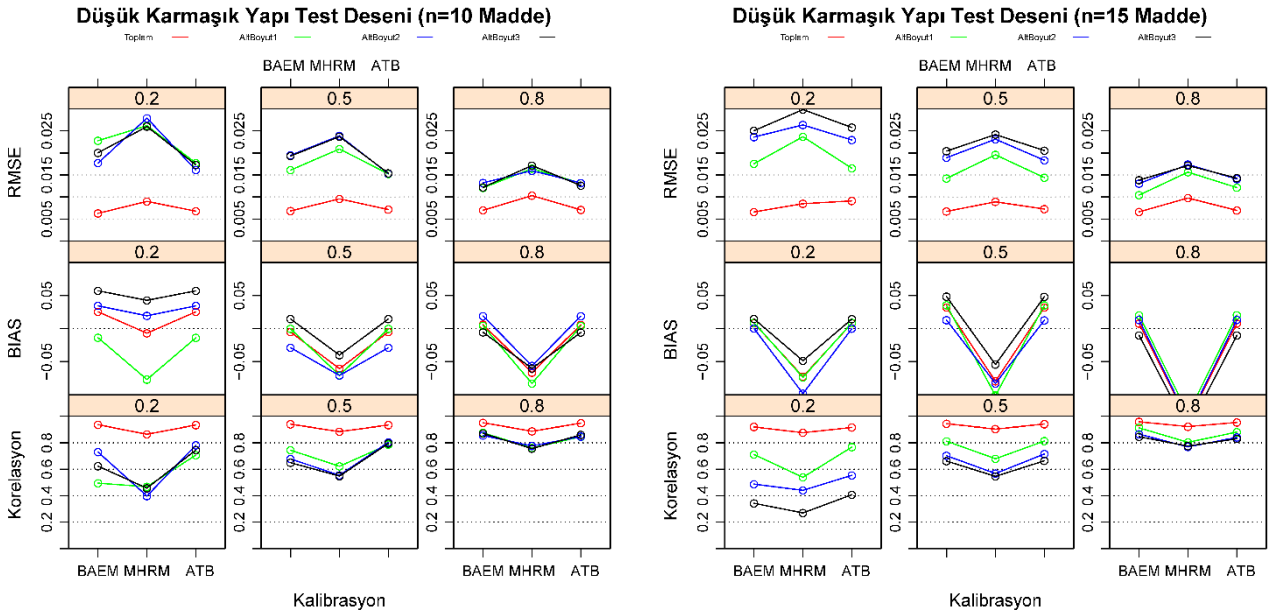
Çizelge 4.4. Gerçek ve Kalibre Edilmiş Puanlar Arası RMSE Değerleri

			Altboyut 1			*Altboyut 2*			*Altboyut 3*			*Toplam Puan*		
Boyut Başına Düşen Madde Sayısı	***Test Deseni***	***Korelasyon***	***BAEM***	***MHRM***	***ATB***	***BAEM***	***MHRM***	***ATB***	***BAEM***	***MHRM***	***ATB***	***BAEM***	***MHRM***	***ATB***
		$\rho = 0.2$	.014	.021	.015	.014	.020	.014	.014	.020	.014	.009	.013	.009
	BY	$\rho = 0.5$	.014	.020	.014	.014	.020	.015	.014	.019	.014	.009	.013	.009
		$\rho = 0.8$	.012	.017	.014	.012	.017	.014	.012	.017	.014	.010	.014	.010
		$\rho = 0.2$	.023	.026	.018	.018	.028	.016	.020	.026	.017	.006	.009	.007
10 Madde	DKY	$\rho = 0.5$	.016	.021	.015	.019	.024	.015	.019	.024	.015	.007	.010	.007
		$\rho = 0.8$	.012	.016	.013	.013	.016	.013	.012	.017	.013	.007	.010	.007
		$\rho = 0.2$	.021	.027	.020	.024	.027	.020	.022	.026	.019	.007	.009	.008
	YKY	$\rho = 0.5$	.018	.023	.017	.018	.022	.016	.018	.022	.016	.006	.009	.007
		$\rho = 0.8$	.013	.017	.013	.013	.017	.014	.012	.017	.013	.007	.010	.007
		$\rho = 0.2$	.012	.017	.012	.012	.017	.012	.011	.016	.011	.007	.011	.007
	BY	$\rho = 0.5$	.012	.017	.012	.011	.016	.012	.011	.016	.011	.007	.011	.008
		$\rho = 0.8$	.010	.015	.012	.010	.015	.012	.010	.015	.012	.007	.011	.008
		$\rho = 0.2$	.018	.024	.017	.024	.026	.023	.025	.030	.026	.007	.008	.009
15 Madde	DKY	$\rho = 0.5$	.014	.020	.014	.019	.023	.018	.020	.024	.021	.007	.009	.007
		$\rho = 0.8$	.010	.016	.012	.013	.017	.014	.014	.017	.014	.007	.010	.007
		$\rho = 0.2$	.019	.026	.019	.022	.027	.021	.024	.027	.024	.006	.008	.008
	YKY	$\rho = 0.5$	.016	.020	.016	.018	.022	.018	.019	.023	.019	.006	.008	.007
		$\rho = 0.8$	.012	.017	.013	.012	.016	.013	.013	.016	.014	.006	.009	.006

Çizelge 4.5. Gerçek ve Kalibre Edilmiş Puanlar Arası Korelasyon Değerleri

			Altboyut 1			*Altboyut 2*			*Altboyut 3*			*Toplam Puan*		
Boyut Başına Düşen Madde Sayısı	***Test Deseni***	***Korelasyon***	***BAEM***	***MHRM***	***ATB***	***BAEM***	***MHRM***	***ATB***	***BAEM***	***MHRM***	***ATB***	***BAEM***	***MHRM***	***ATB***
		$\rho = 0.2$	.82	.63	.81	.82	.66	.82	.84	.66	.83	.86	.72	.86
	BY	$\rho = 0.5$	.83	.66	.81	.82	.64	.80	.83	.70	.81	.89	.77	.89
		$\rho = 0.8$	.89	.77	.83	.88	.74	.83	.88	.75	.84	.92	.82	.92
		$\rho = 0.2$	.49	.47	.71	.73	.40	.78	.62	.46	.75	.94	.86	.93
10 Madde	DKY	$\rho = 0.5$	.74	.62	.79	.68	.55	.80	.65	.55	.80	.94	.88	.93
		$\rho = 0.8$	.88	.76	.85	.86	.78	.85	.87	.76	.86	.95	.89	.95
		$\rho = 0.2$	.58	.36	.65	.46	.38	.65	.53	.44	.68	.92	.86	.92
	YKY	$\rho = 0.5$	.74	.62	.78	.70	.61	.77	.72	.59	.78	.95	.90	.95
		$\rho = 0.8$	.87	.76	.85	.85	.76	.84	.87	.76	.85	.96	.90	.95
		$\rho = 0.2$	.87	.75	.87	.88	.76	.88	.89	.78	.89	.91	.82	.91
	BY	$\rho = 0.5$	.89	.76	.87	.90	.79	.90	.90	.79	.89	.94	.86	.94
		$\rho = 0.8$	.91	.80	.88	.91	.80	.89	.92	.81	.88	.95	.87	.95
		$\rho = 0.2$	.71	.54	.77	.49	.44	.55	.34	.27	.41	.92	.88	.92
15 Madde	DKY	$\rho = 0.5$	.81	.68	.81	.70	.57	.72	.66	.55	.66	.94	.90	.94
		$\rho = 0.8$	.91	.80	.88	.87	.77	.84	.85	.78	.83	.96	.92	.95
		$\rho = 0.2$	.64	.47	.68	.55	.43	.62	.47	.42	.51	.94	.89	.94
	YKY	$\rho = 0.5$	.77	.65	.78	.71	.59	.73	.67	.56	.68	.95	.92	.95
		$\rho = 0.8$	.89	.77	.86	.88	.79	.84	.85	.77	.84	.96	.92	.96

Düşük karmaşık yapılı testlerde de testin çok boyutluluğu azaldıkça özellikle alt boyut yetenek puanlarının kestirimlerine ait $RMSE_{Kal}$ ve $BIAS_{Kal}$ değerleri azalmaktadır. Madde sayısının yüksek (n=15) ve testlerin çok boyutluluğun en zayıf olduğu korelasyon değerinde (p=0.8) $RMSE_{Kal}$ en yüksek MHRM değerlerinde gözlenmiştir. Çok boyutluluğun artmasına bağlı olarak $Korelasyon_{Kal}$ da ters orantılı olarak değişmektedir. Test ne kadar katı çok boyutluluk özelliği gösteriyorsa, korelasyon değerleri o kadar düşüktür demektir. Çok boyutluluğun ihlal edildiği durumlarda $Korelasyon_{Kal}$ değerleri artmaktadır. Bu durum madde sayısının arttığı koşullarda da geçerlidir.



Şekil 4.2. Düşük Karmaşık Yapılı Testlerin $RMSE_{Kal}$, $BIAS_{Kal}$ ve $Korelasyon_{Kal}$ Değerlerinin Koşullara Göre Karşılaştırması

Çizelge 4.4 incelendiğinde DKY yapılı bir testte toplam puanlara ait en düşük $RMSE_{Kal}$ değeri boyutlar arası korelasyonun en düşük olduğu (çok boyutluluğun en yüksek olduğu) ve boyut başına düşen madde sayısının 10 olduğu durumda BAEM kalibrasyonunda 0.06 olarak kestirilmiştir. Bir önceki BY test yapısında ise en düşük $RMSE_{Kal}$ madde sayısının 15 olduğu durumda görülmüştü. Bu bakımdan testin karmaşık yapıya doğru çevrilmesi durumunda $RMSE_{Kal}$ madde sayısına göre farklılık göstermektedir denilebilir. Aynı yapıdaki testte alt boyut puanlarına ait en düşük $RMSE_{Kal}$ değeri boyutlar arası korelasyonun en yüksek olduğu ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM kalibrasyonunda 0.10

olarak kestirilmiştir. Bu bulgu BY test desenindeki bulgular ile örtüşmektedir. DKY testteki toplam puanlara ait en yüksek $Korelasyon_{Kal}$ değeri boyutlar arası korelasyonun en yüksek olduğu ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM kalibrasyonunda 0.96 olarak kestirilmiştir. Alt boyut puanlarında ise en yüksek $Korelasyon_{Kal}$ değeri yine aynı koşullarda BAEM kalibrasyonunda 0.91 olarak kestirilmiştir. DKY testlerde de toplam yetenek puanlara ait değerler, alt boyut yetenek puanlarına göre daha düşük hata, yansız ve yüksek korelasyona sahip şekilde kestirilmiştir.

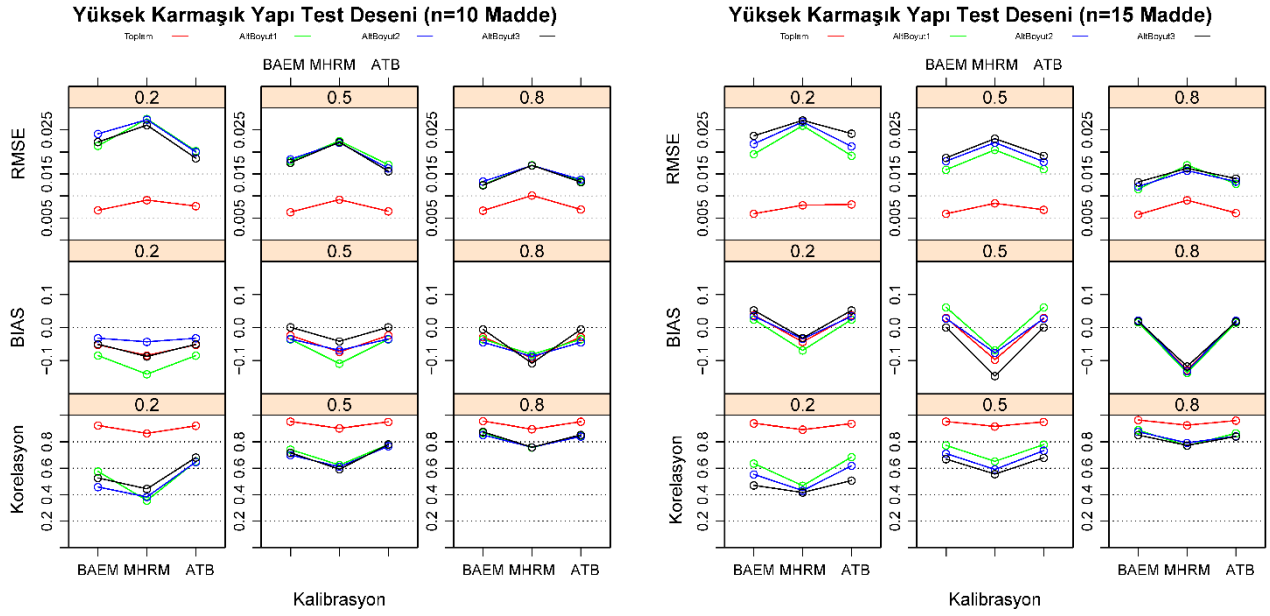
4.2.3. Yüksek Karmaşık Yapılı (YKY) Testlerde Yetenek Puanlarının Karşılaştırılması

Çizelge 4.4 incelendiğinde yüksek karmaşık yapılı (YKY) test deseninde de diğer iki yapıdaki gibi, tüm yetenek puanları kestiriminde en yüksek $RMSE_{Kal}$ değerleri MHRM yönteminde gözlenmiştir. BAEM ve ATB kalibrasyonları, MHRM yöntemine göre bu test deseninde de daha düşük $RMSE_{Kal}$ ile kestirilmiştir. Kestirilen puanlara ait $BIAS_{Kal}$ değerleri de $RMSE_{Kal}$ değerleri ile paralel sonuçlar vermiştir. Çizelge 4.3 incelendiğinde en yüksek $BIAS_{Kal}$ değerleri, diğer desenlerde olduğu gibi bu desen için de MHRM kalibrasyonunda görülmüştür. Bu durumda yüksek karmaşık yapılı test desenlerinde hem alt boyut hem de toplam yetenek puanları kestirmede MHRM kalibrasyonun yanlı ve hatası yüksek kestirimler yaptığı sonucuna varılabilir. Çizelge 4.5'te verilen $Korelasyon_{Kal}$ değerlerine bakıldığında yine var olan tüm koşullarda MHRM kalibrasyonu diğer iki kalibrasyon yönteminden daha düşük değerler üretmiştir.

Şekil 4.3 incelendiğinde alt boyut yetenek puanlarının kestiriminde boyut başına düşen madde sayısı arttıkça $RMSE_{Kal}$ ve $BIAS_{Kal}$ değerlerinde azalma, $Korelasyon_{Kal}$ değerlerinde ise bir artma olduğu belirlenmiştir. Örneğin *Alt Boyut 1* için, maddeler arası korelasyonların 0.2 olduğu koşulda BAEM, MHRM ve ATB kalibrasyonları için madde sayısı arttıkça $RMSE_{Kal}$ değerleri azalmakta, $Korelasyon_{Kal}$ değerleri artmaktadır. $BIAS_{Kal}$ değerleri ise belirli bir örüntü izlememektedir ancak yine en yüksek $RMSE_{Kal}$ ve $BIAS_{Kal}$ değeri ile en düşük $Korelasyon_{Kal}$ değeri MHRM kalibrasyonunda görülmüştür. Toplam yetenek puanlarında ise durum alt boyut puanları ile benzerdir.

Yüksek karmaşık yapılı testlerde de testin çok boyutluluğu azaldıkça özellikle alt boyut yetenek puanlarının kestirimlerine ait $RMSE_{Kal}$ ve $BIAS_{Kal}$ değerleri

azalmaktadır. Madde sayısının yüksek (n=15) ve testlerin çok boyutluluğun en zayıf olduğu korelasyon değerinde (p=0.8) $RMSE_{Kal}$ en yüksek MHRM değerlerinde gözlenmiştir. Çok boyutluluğu artmasına bağlı olarak $Korelasyon_{Kal}$ da ters orantılı olarak değişmektedir. Test ne kadar katı çok boyutluluk özelliği gösteriyorsa, $Korelasyon_{Kal}$ değerleri o kadar düşüktür demektir. Çok boyutluluğun ihlal edildiği durumlarda $Korelasyon_{Kal}$ değerleri artmaktadır. Bu durum madde sayısının arttığı koşullarda da geçerlidir.



Şekil 4.3. Yüksek Karmaşık Yapılı Testlerin $RMSE_{Kal}$, $BIAS_{Kal}$ ve $Korelasyon_{Kal}$ Değerlerinin Koşullara Göre Karşılaştırması

Çizelge 4.4 incelendiğinde YKY yapılı bir testte toplam puanlara ait en düşük $RMSE_{Kal}$ değeri boyutlar arası korelasyonun en yüksek olduğu (çok boyutluluğun en düşük olduğu) ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM kalibrasyonunda 0.06 olarak kestirilmiştir. Bir önceki DBY test yapısında ise en düşük $RMSE_{Kal}$ madde sayısının 10 olduğu durumda, BY test yapısında ise madde sayısının 15 olduğu durumda görülmüştü. Aynı yapıdaki testte alt boyut puanlarına ait en düşük $RMSE_{Kal}$ değeri boyutlar arası korelasyonun en yüksek olduğu ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM kalibrasyonunda 0.12 olarak kestirilmiştir. Bir önceki DKY ile karşılaştırıldığında alt boyut yetenek puanlarına ait $RMSE_{Kal}$ değerlerinde hafif bir artış görülmüştür. Bunun sonucunda alt boyut yetenek puanlarının hatasız kestirilmesinde testin

karmaşıklık yapısına dair optimum bir değer aranması gerektiği ortaya çıkmaktadır. YKY testteki toplam puanlara ait en yüksek $Korelasyon_{Kal}$ değeri boyutlar arası korelasyonun en yüksek olduğu durumda BAEM ve ATB kalibrasyonunda 0.96 olarak kestirilmiştir. Madde sayısının artması gözle görülür bir artış sağlamamıştır. Bunun nedeni olarak halihazırda yüksek olarak kestirilen $Korelasyon_{Kal}$ değerlerinin artırılması için gerekli madde sayısının fazla olması gerektiğidir. Alt boyut puanlarında ise en yüksek $Korelasyon_{Kal}$ değeri yine aynı koşullarda BAEM kalibrasyonunda 0,89 olarak kestirilmiştir. DKY testlerde de toplam yetenek puanlara ait değerler, alt boyut yetenek puanlarına göre daha düşük hata, yansız ve yüksek korelasyona sahip şekilde kestirilmiştir.

DKY ve YKY test desenlerinde çok boyutlu kalibrasyon yöntemlerinde çok boyutluluk azaldığında tutarlı bir şekilde $RMSE_{Kal}$ değerlerinde kısmen de olsa bir artış gözlemlenmiştir. Bu durumun sebebi çok boyutlu kalibrasyonlarda boyutlar arası korelasyon değerinin önemli rol oynamasıdır. Ancak ATB kalibrasyonu boyutlar arasındaki ilişkileri göz önüne almadığı için bu durumdan etkilenmemiştir. Boyutlar arası korelasyonlar arttığında testler birbirine daha da benzer hale gelmiş, bu nedenle elde edilen toplam yetenek parametrelerine ait $RMSE_{Kal}$ değerleri azalma eğilimi göstermiştir. BY test deseninde ise çok boyutlu kalibrasyonlar diğer iki test yapısıyla aynı özellikleri göstermelerine rağmen, ATB kalibrasyonuna göre elde edilen toplam yetenek parametrelerinin $RMSE_{Kal}$ değerleri, boyutluluk arttıkça artış göstermiştir. Bu durumda test yapısının basit ya da karmaşık olmasının ATB kalibrasyonu $RMSE_{Kal}$ değerleri üzerinde bir etkisi olduğu görülmüştür.

Genel olarak değerlendirildiğinde alt boyut yetenek puanlarının aksine, toplam puanlar model karmaşıklaştıkça daha düşük $RMSE_{Kal}$ değerlerine sahip olmaktadır. Ancak model basit yapıdan karmaşık yapıya gittikçe her üç kalibrasyon yöntemi için alt boyut yetenek puanlarına ait $RMSE_{Kal}$ değerleri artmaktadır. Bu durum alt boyut puanlarının raporlanmasında göz önüne alınması gereken bir durumdur. Boyut başına düşen madde sayısı arttıkça $RMSE_{Kal}$ değerleri her durumda azalmaktadır. Bu durum daha önce yapılan çalışmalarla benzerlik göstermektedir (Yao, 2010). Ayrıca, çok boyutluluğu artması alt boyut ve toplam puanların $RMSE_{Kal}$ değerlerine zıt şekilde etki etmiştir. MHRM yaklaşımından elde edilen $RMSE_{Kal}$ değerleri hem alt boyut hem de toplam yetenek puanları için diğer iki kalibrasyon yönteminden daha yüksek kestirilmiştir. Toplam puanlara ait

$RMSE_{Kal}$ değerleri, alt boyutlara ait $RMSE_{Kal}$ değerlerinden her koşulda düşük kestirilmiştir.

4.3. Birinci Alt Probleme Ait Bulgular ve Yorumlar

Basit Yapılı (BY) test deseninde çok boyutlu madde seçme yöntemlerinin hatası, yanlılığı ve korelasyon değerleri kalibrasyon yöntemine, boyutlar arası korelasyona ve boyut başına düşen madde sayısına göre alt boyut ve toplam yetenek puanları için nasıl değişmektedir?

Bu bölüme ait bulgularda basit yapılı test desenine ait ÇBBBT uygulamaları sonucu elde edilen kestirilen yetenek puanları ile gerçek yetenek puanları karşılaştırılarak hata, yanlılık ve korelasyon değerleri belirlenen koşullar bakımından karşılaştırılmıştır. Bulgular sonucunda var olan koşullar altında en iyi performansı gösteren madde seçme yöntemi de rapor edilmiştir. Yao (2012; 2013) çok boyutlu madde seçme yöntemlerinin performansının değerlendirilmesinde ilk olarak mutlak yanlılık değerlerinin kullanılmasını önermiştir. Mutlak yanlılık değerleri ile birlikte önce korelasyon değerleri daha sonra da hata değerleri göz önüne alınarak her bir koşulda en iyi performansı gösteren madde seçme yöntemi hem alt boyut hem de toplam yetenek puanları için belirlenmiştir. Her bir alt boyuta ait kestirilen değerler çizelgelerde ayrıntılı olarak verilmesine rağmen, alt boyut puanları ayrı ayrı değil genel olarak yorumlanmıştır.

4.3.1. Kalibrasyon Yönteminin Etkisinin İncelenmesi

Basit yapıdaki bir testte madde seçme yöntemlerinin hata, yanlılık ve korelasyon değerlerinin kalibrasyon yöntemlerine göre hem alt boyut hem de toplam yetenek puanlarında değiştiği gözlenmiştir.

Çizelge 4.6, Çizelge 4.7 ve Çizelge 4.8 incelendiğinde alt boyut puanları için tek boyutlu kalibrasyon yöntemi olan ATB kalibrasyonu kullanıldığında, kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.407 ile 0.521 arasında; V1 yöntemi için 0.402 ile 0.511 ve Vol yöntemi için ise 0.409 ile 0.532 arasında değiştiği görülmektedir. RMSE değerlerine bakıldığında hata değerlerinin KL yöntemi için 0.015 ile 0.022; V1 yöntemi için 0.015 ile 0.021 ve Vol yöntemi için ise 0.016 ile 0.022 arasında olduğu belirlenmiştir. Kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanlarına ait korelasyon değerleri ise KL yöntemi için

0.52 ile 0.79; V1 yöntemi için 0.54 ile 0.80 ve Vol yöntemi için ise 0.51 ile 0.79 arasında hesaplanmıştır.

Alt boyut yetenek puanlarını kestirmede çok boyutlu kalibrasyon yöntemlerinden MHRM kalibrasyonu kullanıldığında, kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.374 ile 0.542 arasında; V1 yöntemi için 0.374 ile 0.547 ve Vol yöntemi için ise 0.368 ile 0.547 arasında değişmektedir. RMSE değerlerine bakıldığında hata değerlerinin KL yöntemi için 0.017 ile 0.023; V1 yöntemi için 0.017 ile 0.023 ve Vol yöntemi için ise 0.017 ile 0.024 arasında olduğu belirlenmiştir. Kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanlarına ait korelasyon değerleri ise KL yöntemi için 0.43 ile 0.73; V1 yöntemi için 0.43 ile 0.73 ve Vol yöntemi için ise 0.45 ile 0.72 arasında hesaplanmıştır.

Çizelge 4.6. Basit Yapı Alt Boyut 1 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	ATB			*MHRM*			*BAEM*		
	KL	*V1*	*Vol*	*KL*	*V1*	*Vol*	*KL*	*V1*	*Vol*
ABSBIAS									
$\rho = 0.2$, N=10	.521	.511	.532	.532	.547	.530	.502	.497	.522
$\rho = 0.5$, N=10	.516	.515	.531	.550	.539	.547	.493	.498	.498
$\rho = 0.8$, N=10	.457	.469	.465	.451	.444	.449	.404	.392	.403
$\rho = 0.2$, N=15	.443	.435	.451	.446	.464	.470	.440	.436	.443
$\rho = 0.5$, N=15	.435	.421	.449	.444	.460	.451	.422	.426	.412
$\rho = 0.8$, N=15	.416	.413	.427	.393	.403	.395	.359	.365	.363
RMSE									
$\rho = 0.2$, N=10	.022	.021	.022	.023	.023	.024	.021	.021	.022
$\rho = 0.5$, N=10	.021	.021	.021	.022	.022	.022	.021	.020	.021
$\rho = 0.8$, N=10	.018	.019	.019	.020	.020	.020	.018	.017	.018
$\rho = 0.2$, N=15	.018	.018	.018	.020	.020	.020	.018	.018	.018
$\rho = 0.5$, N=15	.017	.017	.017	.019	.019	.020	.018	.017	.017
$\rho = 0.8$, N=15	.015	.015	.016	.017	.018	.017	.015	.015	.015
Korelasyon									
$\rho = 0.2$, N=10	.52	.54	.51	.43	.43	.45	.55	.55	.52
$\rho = 0.5$, N=10	.56	.56	.54	.49	.50	.49	.57	.59	.57
$\rho = 0.8$, N=10	.71	.68	.68	.65	.65	.65	.74	.74	.74
$\rho = 0.2$, N=15	.69	.68	.68	.60	.59	.59	.68	.69	.68
$\rho = 0.5$, N=15	.73	.73	.72	.65	.66	.64	.72	.74	.74
$\rho = 0.8$, N=15	.79	.80	.78	.73	.72	.72	.81	.80	.80

Bir diğer çok boyutlu kalibrasyon yöntemi olan BAEM kalibrasyonu alt boyut yetenek puanlarını kestirmede kullanıldığında, kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.347 ile 0.507 arasında; V1 yöntemi için 0.347 ile 0.505 ve Vol yöntemi için ise 0.352 ile 0.522 arasında değişmektedir. RMSE değerlerine bakıldığında hata değerlerinin KL yöntemi için 0.015 ile 0.021; V1

yöntemi için 0.015 ile 0.021 ve Vol yöntemi için ise 0.015 ile 0.022 arasında olduğu belirlenmiştir. Kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanlarına ait korelasyon değerleri ise KL yöntemi için 0.55 ile 0.81; V1 yöntemi için 0.54 ile 0.81 ve Vol yöntemi için ise 0.52 ile 0.81 arasında hesaplanmıştır.

Çizelge 4.7. Basit Yapı Alt Boyut 2 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	ATB			*MHRM*			*BAEM*		
	KL	*V1*	*Vol*	*KL*	*V1*	*Vol*	*KL*	*V1*	*Vol*
ABSBIAS									
$\rho = 0.2$, **N=10**	.500	.508	.508	.542	.541	.546	.507	.505	.501
$\rho = 0.5$, **N=10**	.492	.481	.489	.506	.483	.501	.461	.457	.460
$\rho = 0.8$, **N=10**	.456	.459	.458	.422	.442	.427	.392	.382	.386
$\rho = 0.2$, **N=15**	.439	.444	.440	.446	.462	.457	.421	.429	.434
$\rho = 0.5$, **N=15**	.430	.412	.431	.426	.420	.434	.404	.404	.405
$\rho = 0.8$, **N=15**	.407	.402	.415	.374	.377	.368	.347	.357	.368
RMSE									
$\rho = 0.2$, **N=10**	.021	.021	.022	.023	.023	.023	.021	.021	.021
$\rho = 0.5$, **N=10**	.020	.019	.019	.022	.022	.021	.020	.020	.019
$\rho = 0.8$, **N=10**	.018	.018	.018	.020	.020	.019	.017	.017	.017
$\rho = 0.2$, **N=15**	.018	.018	.018	.020	.020	.020	.018	.018	.018
$\rho = 0.5$, **N=15**	.018	.017	.017	.019	.019	.019	.017	.017	.017
$\rho = 0.8$, **N=15**	.016	.016	.016	.018	.018	.018	.015	.016	.016
Korelasyon									
$\rho = 0.2$, **N=10**	.57	.56	.55	.47	.46	.46	.55	.54	.56
$\rho = 0.5$, **N=10**	.57	.62	.60	.50	.51	.52	.61	.61	.61
$\rho = 0.8$, **N=10**	.69	.68	.70	.63	.63	.65	.74	.74	.75
$\rho = 0.2$, **N=15**	.69	.69	.69	.62	.59	.61	.69	.69	.69
$\rho = 0.5$, **N=15**	.73	.75	.74	.66	.67	.67	.74	.74	.74
$\rho = 0.8$, **N=15**	.79	.78	.78	.73	.72	.72	.80	.79	.80

Alt boyut puanlarına ait Çizelge 4.6, Çizelge 4.7 ve Çizelge 4.8'deki değerler incelendiğinde maddelerin tek boyutlu (ATB) olarak kalibre edildiği durumlarda, V1 madde seçme yönteminin diğer madde seçme yöntemlerine göre daha düşük mutlak yanlılık, daha yüksek korelasyon değerine sahip olduğu görülmektedir. Çok boyutlu MHRM kalibrasyonu kullanıldığı durumda KL madde seçme yönteminin daha tutarlı sonuçlar verdiği görülmektedir. Son olarak bir diğer çok boyutlu kalibrasyon yöntemi olan BAEM kalibrasyonunda V1 madde seçme yönteminin daha tutarlı sonuçlar verdiği görülmektedir. Elde edilen sonuçlar değerlendirildiğinde, alt boyut yetenek puanlarını kestirmede kullanılan madde seçme yöntemlerinin kalibrasyon yöntemlerine göre farklılaştığı görülmektedir. BAEM kalibrasyonu ile birlikte V1 madde seçme yöntemi, alt boyut yetenek puanlarını daha düşük yanlılık ve hata; daha yüksek korelasyon değerleri ile kestirerek var olan koşullar arasında en iyi performansı sergilemişlerdir.

Çizelge 4.9 incelendiğinde toplam yetenek puanları için tek boyutlu kalibrasyon yöntemi olan ATB kalibrasyonu kullanıldığında, kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.296 ile 0.340 arasında; V1 yöntemi için 0.294 ile 0.338 ve Vol yöntemi için ise 0.291 ile 0.343 arasında değiştiği belirlenmiştir. RMSE değerlerine bakıldığında hata değerlerinin KL yöntemi için 0.011 ile 0.016; V1 yöntemi için 0.011 ile 0.016 ve Vol yöntemi için ise 0.012 ile 0.016 arasında olduğu belirlenmiştir. Kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanlarına ait korelasyon değerleri ise KL yöntemi için 0.59 ile 0.83; V1 yöntemi için 0.61 ile 0.84 ve Vol yöntemi için ise 0.61 ile 0.83 arasında hesaplanmıştır.

Çizelge 4.8. Basit Yapı Alt Boyut 3 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	SU			*MHRM*			*BAEM*		
	KL	*V1*	*Vol*	*KL*	*V1*	*Vol*	*KL*	*V1*	*Vol*
ABSBIAS									
$\boldsymbol{\rho = 0.2}$**, N=10**	.485	.480	.474	.519	.533	.529	.484	.479	.469
$\boldsymbol{\rho = 0.5}$**, N=10**	.489	.476	.485	.506	.486	.506	.474	.447	.465
$\boldsymbol{\rho = 0.8}$**, N=10**	.467	.473	.467	.428	.421	.423	.394	.383	.382
$\boldsymbol{\rho = 0.2}$**, N=15**	.432	.425	.424	.442	.433	.460	.430	.427	.421
$\boldsymbol{\rho = 0.5}$**, N=15**	.427	.420	.439	.407	.419	.422	.416	.417	.405
$\boldsymbol{\rho = 0.8}$**, N=15**	.412	.412	.409	.378	.374	.386	.357	.347	.352
RMSE									
$\boldsymbol{\rho = 0.2}$**, N=10**	.020	.020	.020	.022	.022	.022	.021	.020	.020
$\boldsymbol{\rho = 0.5}$**, N=10**	.021	.020	.020	.022	.021	.022	.020	.020	.020
$\boldsymbol{\rho = 0.8}$**, N=10**	.019	.019	.018	.020	.020	.020	.017	.018	.017
$\boldsymbol{\rho = 0.2}$**, N=15**	.018	.017	.017	.019	.019	.020	.018	.018	.017
$\boldsymbol{\rho = 0.5}$**, N=15**	.017	.017	.018	.019	.019	.019	.017	.017	.017
$\boldsymbol{\rho = 0.8}$**, N=15**	.016	.015	.016	.017	.017	.018	.015	.015	.015
Korelasyon									
$\boldsymbol{\rho = 0.2}$**, N=10**	.56	.61	.60	.49	.50	.50	.58	.58	.60
$\boldsymbol{\rho = 0.5}$**, N=10**	.57	.58	.58	.49	.53	.52	.59	.61	.61
$\boldsymbol{\rho = 0.8}$**, N=10**	.68	.68	.70	.62	.61	.63	.73	.72	.74
$\boldsymbol{\rho = 0.2}$**, N=15**	.70	.71	.71	.64	.64	.60	.70	.70	.71
$\boldsymbol{\rho = 0.5}$**, N=15**	.73	.74	.72	.65	.66	.65	.72	.73	.73
$\boldsymbol{\rho = 0.8}$**, N=15**	.78	.79	.79	.73	.73	.72	.80	.81	.81

Toplam yetenek puanlarını kestirmede çok boyutlu kalibrasyon yöntemlerinden MHRM kalibrasyonu kullanıldığında, kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.292 ile 0.363 arasında; V1 yöntemi için 0.292 ile 0.352 ve Vol yöntemi için ise 0.290 ile 0.360 arasında değişmektedir. RMSE değerlerine bakıldığında hata değerlerinin KL yöntemi için 0.013 ile 0.018; V1 yöntemi için 0.013 ile 0.018 ve Vol yöntemi için ise 0.013 ile 0.017 arasında olduğu belirlenmiştir. Kestirilen toplam yetenek puanları ile gerçek toplam yetenek

puanlarına ait korelasyon değerleri ise KL yöntemi için 0.54 ile 0.78; V1 yöntemi için 0.53 ile 0.78 ve Vol yöntemi için ise 0.55 ile 0.78 arasında hesaplanmıştır.

Bir diğer çok boyutlu kalibrasyon yöntemi olan BAEM kalibrasyonu toplam yetenek puanlarını kestirmede kullanıldığında, kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.290 ile 0.359 arasında; V1 yöntemi için 0.286 ile 0.349 ve Vol yöntemi için ise 0.289 ile 0.355 arasında değişmektedir. RMSE değerlerine bakıldığında hata değerlerinin KL. V1 ve Vol değerleri için 0.011 ile 0.015 arasında olduğu belirlenmiştir. Kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanlarına ait korelasyon değerleri ise KL. V1 ve Vol yöntemi için ise 0.64 ile 0.86 arasında hesaplanmıştır.

Çizelge 4.9. Basit Yapı Toplam Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	ATB			*MHRM*			*BAEM*		
	KL	*V1*	*Vol*	*KL*	*V1*	*Vol*	*KL*	*V1*	*Vol*
ABSBIAS									
$\boldsymbol{\rho = 0.2}$**, N=10**	.324	.312	.318	.332	.337	.340	.328	.323	.327
$\boldsymbol{\rho = 0.5}$**, N=10**	.340	.335	.343	.363	.352	.360	.358	.349	.355
$\boldsymbol{\rho = 0.8}$**, N=10**	.335	.338	.338	.352	.351	.341	.359	.345	.354
$\boldsymbol{\rho = 0.2}$**, N=15**	.302	.296	.306	.308	.317	.321	.307	.302	.313
$\boldsymbol{\rho = 0.5}$**, N=15**	.279	.274	.291	.292	.292	.290	.290	.286	.289
$\boldsymbol{\rho = 0.8}$**, N=15**	.296	.294	.305	.299	.296	.299	.302	.304	.306
RMSE									
$\boldsymbol{\rho = 0.2}$**, N=10**	.014	.014	.014	.015	.015	.015	.014	.014	.014
$\boldsymbol{\rho = 0.5}$**, N=10**	.015	.015	.015	.016	.016	.016	.014	.014	.014
$\boldsymbol{\rho = 0.8}$**, N=10**	.016	.016	.016	.018	.018	.017	.015	.015	.015
$\boldsymbol{\rho = 0.2}$**, N=15**	.011	.011	.012	.013	.013	.013	.011	.011	.011
$\boldsymbol{\rho = 0.5}$**, N=15**	.012	.012	.013	.014	.013	.014	.012	.012	.012
$\boldsymbol{\rho = 0.8}$**, N=15**	.013	.013	.013	.015	.015	.015	.012	.012	.012
Korelasyon									
$\boldsymbol{\rho = 0.2}$**, N=10**	.59	.61	.61	.54	.53	.55	.64	.64	.64
$\boldsymbol{\rho = 0.5}$**, N=10**	.65	.66	.65	.59	.60	.60	.70	.71	.70
$\boldsymbol{\rho = 0.8}$**, N=10**	.74	.73	.73	.68	.67	.69	.79	.79	.79
$\boldsymbol{\rho = 0.2}$**, N=15**	.75	.76	.75	.69	.68	.67	.77	.77	.76
$\boldsymbol{\rho = 0.5}$**, N=15**	.81	.82	.80	.75	.76	.75	.82	.83	.83
$\boldsymbol{\rho = 0.8}$**, N=15**	.83	.84	.83	.78	.78	.78	.86	.86	.86

Toplam yetenek puanlarına ait Çizelge 4.9'daki değerler incelendiğinde maddelerin tek boyutlu (ATB) olarak kalibre edildiği durumlarda, V1 madde seçme yönteminin diğer madde seçme yöntemlerine göre daha düşük mutlak yanlılık, daha yüksek korelasyon değerine sahip olduğu görülmektedir. Çok boyutlu MHRM kalibrasyonu kullanıldığı durumda V1 madde seçme yönteminin daha tutarlı sonuçlar verdiği görülmektedir. Son olarak bir diğer çok boyutlu kalibrasyon yöntemi olan BAEM kalibrasyonunda V1 madde seçme yönteminin daha tutarlı

sonuçlar verdiği görülmektedir. Elde edilen sonuçlar değerlendirildiğinde, toplam yetenek puanlarını kestirmede kullanılan madde seçme yöntemlerinin kalibrasyon yöntemlerine göre farklılaşmadığı görülmektedir. BAEM kalibrasyonu ile birlikte V1 madde seçme yöntemi, toplam yetenek puanlarını daha düşük yanlılık ve hata; daha yüksek korelasyon değerleri ile kestirerek var olan koşullar arasında en iyi performansı sergilemişlerdir.

4.3.2. Boyutlar Arası Korelasyonun ve Boyut Başına Düşen Madde Sayısının Etkisinin İncelenmesi

Araştırma sonuçları incelendiğinde madde seçme yöntemlerinin hata, yanlılık ve korelasyon değerlerinin boyutlar arasındaki korelasyon değerine (çok boyutluluk) ve boyut başına düşen madde sayısına (test uzunluğu) göre hem alt boyut hem de toplam yetenek puanlarında değiştiği gözlenmiştir. Bu bölümde, BY desenindeki bir test için, kalibrasyon yöntemlerini de dikkate alarak, çok boyutluluğun ve test uzunluğunun madde seçme yöntemlerine etkisi incelenmektedir. Sonuçlar incelediğinde testin uzunluğu arttığında kestirilen yetenek puanları ile gerçek yetenek puanları arasındaki korelasyon değerlerinin arttığı; yetenek puanlarının mutlak yanlılık ve hata değerlerinin ise azaldığı belirlenmiştir. Ancak bu durum çok boyutluluk (boyutlar arası korelasyon) koşulu için geçerli değildir. Çok boyutluluk özelliklerine göre alt boyut ve toplam yetenek puanlarına ait mutlak yanlılık ve hata değerleri farklı durumlar sergilemiştir. Tüm koşullar için madde seçme yöntemlerinin tutarlılığı aşağıda ayrıntılı şekilde yorumlanmıştır.

MHRM kalibrasyonu tüm kalibrasyon yöntemleri arasında en zayıf performansı gösteren kalibrasyon yöntemi olmuştur. Hem hata değerlerinin hem de mutlak yanlılık değerlerinin yüksek olarak kestirilmesi nedeniyle madde seçme yöntemleri tutarlılık göstermemiştir. Tek boyutlu ve çok boyutlu kalibrasyonu daha güçlü bir şekilde karşılaştırmak adına çalışmanın bundan sonraki kısmında MHRM kalibrasyonu yorumlamalarda kullanılmamış ancak kalibrasyona ilişkin değerler çizelge ve grafiklerde verilmiştir. Bu bakımdan bulgular, tek boyutlu kalibrasyon için ATB, çok boyutlu kalibrasyon için BAEM kalibrasyonu dikkate alınarak yorumlanmıştır.

Daha önce de belirtildiği gibi BY test deseninde kalibrasyon yöntemlerinin madde seçme yöntemleri üzerinde etkisi olduğu ve her bir kalibrasyon yönteminin, çok

boyutluluk ve test uzunluğu değişkenleri göz önüne alınmadığı durumda, belirli bir madde seçme yöntemine avantaj sağladığı görülmüştür.

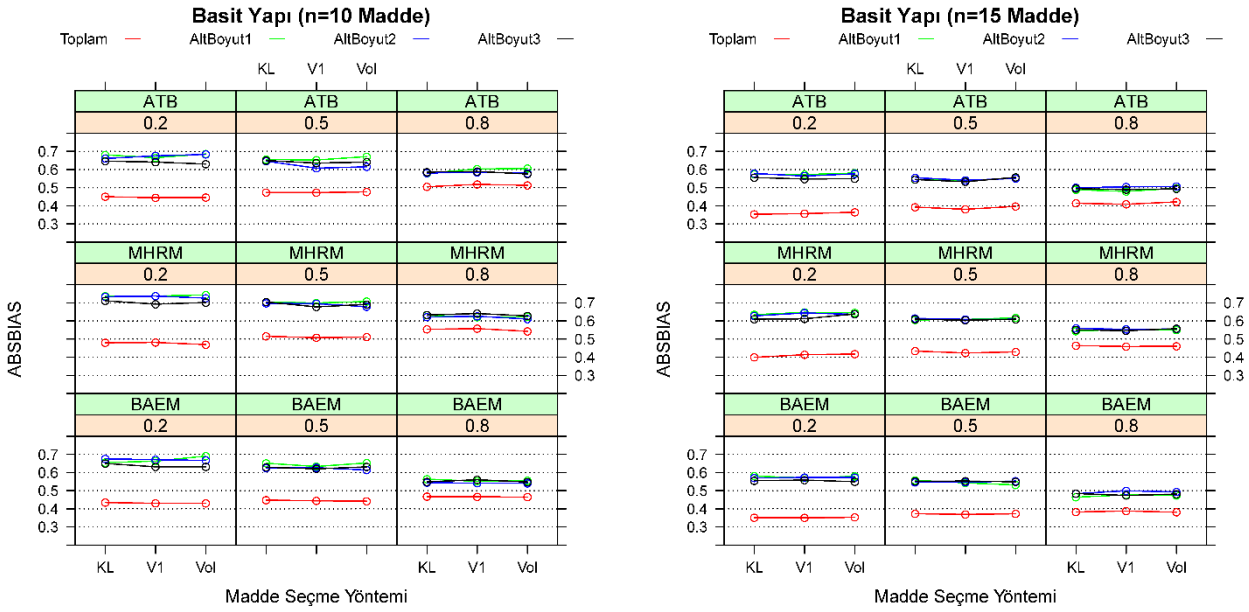
Tek Boyutlu (ATB) Kalibrasyon Sonuçları

Basit yapılı bir testte alt boyut puanların kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde tek boyutlu kalibrasyon yönteminde (ATB) V1 madde seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.502, 0.499 ve 0.504 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.021; korelasyon değerleri ise sırasıyla 0.55, 0.57 ve 0.55 olarak hesaplanmıştır. Boyut başına düşen madde sayısının sabit tutulup (n=10), çok boyutluluğun etkisinin azaldığı durumda (ρ=0.8), mutlak yanlılık ve hata değerleri düşerken, korelasyon değerleri %25 oranında artmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.460, 0.467 ve 0.463 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.018, 0.019 ve 0.018; korelasyon değerleri ise sırasıyla 0.69, 0.68 ve 0.69 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=15 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.438, 0.435 ve 0.438 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.018; korelasyon değerleri ise 0.69 olarak hesaplanmıştır. Bu bulguların sonucuna bakıldığında testin katı çok boyutluluk (ρ=0.2) özelliği gösterdiği durumda, boyut başına düşen madde sayısının artırtılması, alt boyut yetenek puanlarının kestirimini olumlu olarak etkileyecektir denilebilir. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.412, 0.409 ve 0.417 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de

0.016, 0.015 ve 0.016; korelasyon değerleri ise sırasıyla 0.79, 0.79 ve 0.78 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.



Şekil 4.4. Basit Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Yanlılık Göre Değerleri

Basit yapılı bir testte toplam puanların kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde çok boyutlu kalibrasyon yönteminde (BAEM) V1 madde seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Toplam yetenek puanları kestirmede elde edilen bulgular yukarıda bahsedilen alt boyut yetenek puanlarını kestirmede elde edilen bulgularla aynı örüntüleri takip ediyor olsa da toplam yetenek puanları daha düşük mutlak yanlılık ve hata değerleri, daha yüksek korelasyon değerleri üretmiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu

durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.324, 0.312 ve 0.318 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.014; korelasyon değerleri ise sırasıyla 0.59, 0.61 ve 0.61 olarak hesaplanmıştır. Toplam yetenek puanlarına ait bir diğer önemli bulgu ise çok boyutluluk değerlerine göre ABSBIAS ve RMSE değerlerindeki artıştır. Alt boyut yetenek puanlarının aksine, boyut başına düşen madde sayısının sabit olduğu ve çok boyutluluğun etkisinin azaldığı durumda, yani boyutlar arası korelasyonun 0.2'den 0.8'e çıktığı durumda, korelasyon değerleri artmış, RMSE ve ABSBIAS değerleri de artmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.335, 0.338 ve 0.338 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.016; korelasyon değerleri ise sırasıyla 0.74, 0.73 ve 0.73 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=15 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.302, 0.296 ve 0.306 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.011, 0.011 ve 0.012; korelasyon değerleri ise sırasıyla 0.75, 0.76 ve 0.75 olarak hesaplanmıştır. Bu bulguların sonucuna bakıldığında testin katı çok boyutluluk (ρ=0.2) özelliği gösterdiği durumda, boyut başına düşen madde sayısının artırtılması, toplam yetenek puanlarının kestirimini olumlu olarak etkileyecektir denilebilir. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.296, 0.294 ve 0.305 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.013; korelasyon değerleri ise sırasıyla 0.83, 0.84 ve 0.83 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu

durumda KL; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.

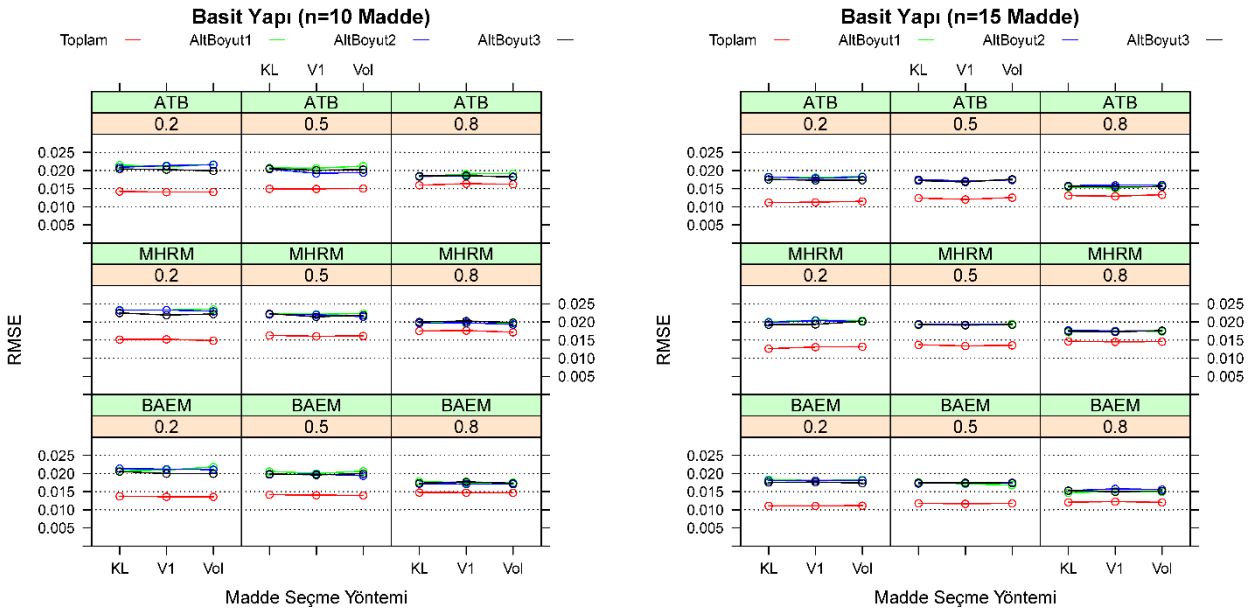
Çok Boyutlu (BAEM) Kalibrasyon Sonuçları

Basit yapılı bir testte alt boyut puanların kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde çok boyutlu kalibrasyon yönteminde (BAEM) V1 madde seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.497, 0.494 ve 0.496 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.021; korelasyon değerleri ise 0.56 olarak hesaplanmıştır. Boyut başına düşen madde sayısının sabit tutulup (n=10), çok boyutluluğun etkisinin azaldığı durumda (ρ=0.8), mutlak yanlılık ve hata değerleri düşerken, korelasyon değerleri artmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.397, 0.386 ve 0.390 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.017; korelasyon değerleri ise sırasıyla 0.74, 0.73 ve 0.74 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=15 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.430, 0.431 ve 0.432 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.018; korelasyon değerleri ise 0.69 olarak hesaplanmıştır. Bu bulguların sonucuna bakıldığında testin katı çok boyutluluk (ρ=0.2) özelliği gösterdiği durumda, boyut başına düşen madde sayısının artırtılması, alt boyut yetenek puanlarının kestirimini olumlu olarak etkileyecektir denilebilir. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise alt boyut yetenek puanlarının KL, V1 ve Vol değerleri

için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.354, 0.357 ve 0.361 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.015; korelasyon değerleri ise 0.80 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.



Şekil 4.5. Basit Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Göre Hata Değerleri

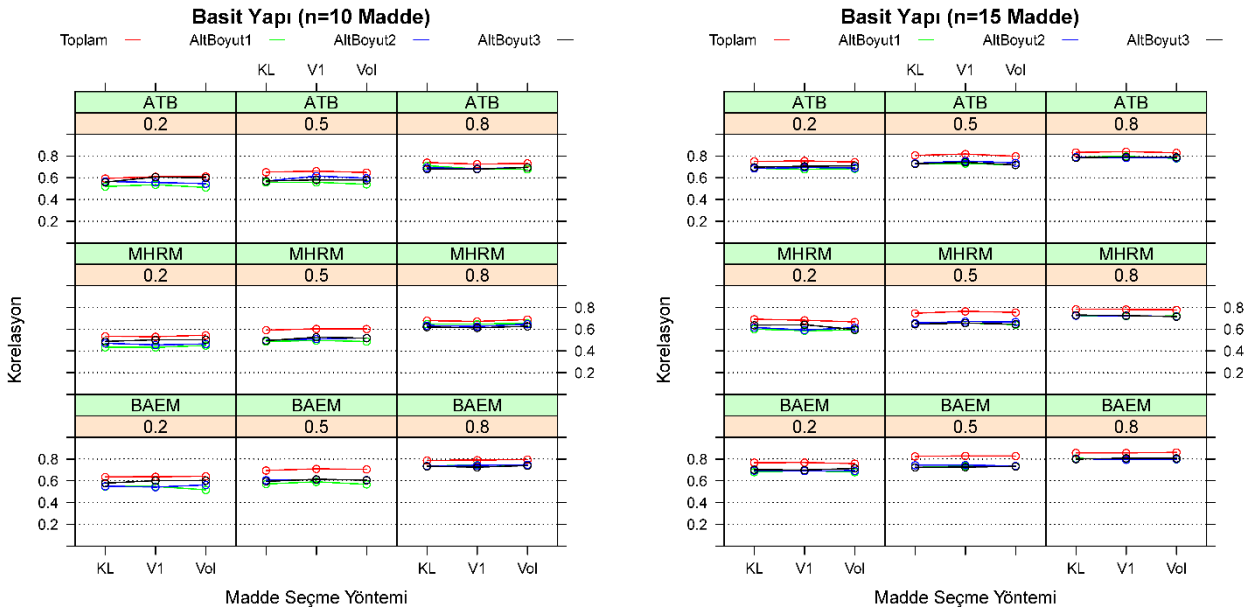
Basit yapılı bir testte toplam puanların kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde çok boyutlu kalibrasyon yönteminde (BAEM) V1 madde seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Toplam yetenek puanları kestirmede elde edilen bulgular ile alt boyut yetenek puanlarını kestirmede elde edilen bulguların aynı örüntüleri takip ettiği sonucuna varılmıştır. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu durumda toplam yetenek puanlarının KL, V1

ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.328, 0.323 ve 0.327 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.014; korelasyon değerleri ise 0.64 olarak hesaplanmıştır. Tıpkı tek boyutlu (ATB) kalibrasyonda olduğu gibi, toplam yetenek puanlarında da çok boyutluluk değerlerine göre ABSBIAS ve RMSE değerlerinde artışlar gözlenmiştir. Alt boyut yetenek puanlarının aksine, boyut başına düşen madde sayısının sabit olduğu ve çok boyutluluğun etkisinin azaldığı durumda, yani boyutlar arası korelasyonun 0.2'den 0.8'e çıktığı durumda, korelasyon değerleri artmış, RMSE ve ABSBIAS değerleri de artmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.359, 0.345 ve 0.354 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.015; korelasyon değerleri ise 0.79 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=15 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.307, 0.302 ve 0.313 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.011; korelasyon değerleri ise sırasıyla 0.77, 0.77 ve 0.76 olarak hesaplanmıştır. Bu bulguların sonucuna bakıldığında testin katı çok boyutluluk (ρ=0.2) özelliği gösterdiği durumda, boyut başına düşen madde sayısının artırtılması, toplam yetenek puanlarının kestirimini olumlu olarak etkileyecektir denilebilir. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.302, 0.304 ve 0.306 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.012; korelasyon değerleri ise 0.86 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen

madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.



Şekil 4.6. Basit Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Göre Korelasyon Değerleri

4.4. İkinci Alt Probleme Ait Bulgular ve Yorumlar

Basit Yapı test deseninde yetenek puanlarına ait mutlak yanlılık ve hata değerleri arasında koşullara göre anlamlı farklılıklar var mıdır?

Mutlak yanlılık ve hata değerlerinin koşullara göre etkilerinin incelenmesi amacıyla varyans analizi uygulanmış ve sonuçlara Çizelge 4.10'da yer verilmiştir.

Çizelge 4.10'da verilen ANOVA sonuçları incelendiğinde boyut başına düşen madde sayısının (BMS), boyutlar arası korelasyon (KOR) değerlerinin ve kalibrasyon yöntemlerinin (KY) hem alt boyut yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde hem de toplam yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde ana etkisinin anlamlı olduğu görülmüştür ($p < .05$). Bu durumun aksine çok boyutlu madde seçme yöntemlerinin (MSY) hem alt boyut yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde hem de toplam yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde anlamlı bir etkisinin olmadığı görülmüştür ($p > .05$).

Çizelge 4.10. Basit Yapı Test Deseni Yetenek Puanlarına Ait Mutlak Yanlılık ve Hata Değerleri ANOVA Sonuçları

			Alt Boyut 1 Puanı		*Alt Boyut 2 Puanı*		*Alt Boyut 3 Puanı*		*Toplam Puan*	
		sd	*F*	η^2	*F*	η^2	*F*	η^2	*F*	η^2
	BMS	1	1572.1*	.99	661.07*	.99	535.85*	.99	1009.8*	.99
	KOR	2	781.60*	.99	403.63*	.99	364.37*	.99	338.60*	.94
	KY	2	349.08*	.94	264.25*	.99	537.63*	.99	529.52*	.96
RMSE	***MSY***	2	-	-	-	-	-		-	-
	BMS*KOR	2	-	-	45.11*	.95	-		-	-
	BMS*KY	2	-	-	-	-	-		8.26*	.80
	KOR*KY	4	4.58*	.69	5.19*	.74	5.15*	.72	28.44*	.93
	BMS	1	1086.7*	.99	2415.2*	.99	447.65*	.99	1007.3*	.99
	KOR	2	1538.5*	.99	303.45*	.99	424.79*	.99	668.30*	.99
	KY	2	257.82*	.99	386.68*	.99	319.21*	.99	1235.7*	.99
ABSBIAS	***MSY***	2	-	-	-	-	-	-	-	-
	BMS*KOR	2	263.81*	.92	101.36*	.98	17.97*	.90	70.74*	.97
	BMS*KY	2	-	-	34.40*	.94	-	-	37.32*	.94
	KOR*KY	4	6.06*	.75	7.11*	.78	-	-	4.41*	.68

Not: BMS=Boyut başına düşen madde sayısı, KOR=Boyutlar arası korelasyon KY=Kalibrasyon yöntemi, MSY=Madde seçme yöntemi. *=0.05 düzeyinde anlamlı,

Toplam yetenek puanlarında, alt boyut yetenek puanlarının aksine boyut başına düşen madde sayısı ve kalibrasyon yönteminin ortak etkileşiminin hata miktarları arasında anlamlı bir ilişki görülmüştür (p <.05). Kalibrasyona giren madde sayısının toplam puanların hata miktarları üzerinde anlamlı bir etkisinin olduğu, alt boyutlar bakımından etkisinin olmadığı görülmektedir. Bunun nedeninin, toplam puanlar üzerinden yapılan kalibrasyonlarda her bir alt boyuta düşen sayısından daha fazla maddenin işleme alınmış olmasından kaynaklandığı düşünülmektedir. Aynı duruma mutlak yanlılıkta rastlanmamıştır. Boyut başına düşen madde sayısı ve kalibrasyon yönteminin ortak etkileşiminin alt boyut ve toplam yetenek puanlarının mutlak yanlılık miktarları arasında anlamlı bir ilişki olduğu görülmüştür (p <.05).

4.5. Üçüncü Alt Probleme Ait Bulgular ve Yorumlar

Düşük Karmaşık Yapı (DKY) test deseninde çok boyutlu madde seçme yöntemlerinin hatası, yanlılığı ve korelasyon değerleri kalibrasyon yöntemine,

boyutlar arası korelasyona ve boyut başına düşen madde sayısına göre alt boyut ve toplam yetenek puanları için nasıl değişmektedir?

Bu bölüme ait bulgularda düşük karmaşık yapı test desenine ait ÇBBBT uygulamaları sonucu elde edilen kestirilen yetenek puanları ile gerçek yetenek puanları karşılaştırılarak hata, yanlılık ve korelasyon değerleri belirlenen koşullar bakımından karşılaştırılmıştır. Bulgular sonucunda var olan koşullar altında en iyi performansı gösteren madde seçme yöntemi de rapor edilmiştir.

4.5.1. Kalibrasyon Yönteminin Etkisinin İncelenmesi

Düşük karmaşık yapıdaki bir testte, basit yapıda oluğu gibi, madde seçme yöntemlerinin hata, yanlılık ve korelasyon değerlerinin kalibrasyon yöntemlerine göre hem alt boyut hem de toplam yetenek puanlarında değiştiği gözlenmiştir.

Çizelge 4.11. Düşük Karmaşık Yapı Alt Boyut 1 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	ATB			*MHRM*			*BAEM*		
	KL	*V1*	*Vol*	*KL*	*V1*	*Vol*	*KL*	*V1*	*Vol*
ABSBIAS									
$\rho = 0.2$, N=10	.471	.457	.489	.570	.547	.561	.432	.422	.437
$\rho = 0.5$, N=10	.411	.417	.437	.517	.516	.507	.376	.365	.387
$\rho = 0.8$, N=10	.347	.334	.342	.385	.396	.389	.315	.310	.310
$\rho = 0.2$, N=15	.400	.401	.401	.555	.559	.537	.392	.395	.387
$\rho = 0.5$, N=15	.342	.339	.352	.484	.487	.480	.342	.345	.384
$\rho = 0.8$, N=15	.300	.296	.300	.360	.366	.372	.280	.280	.280
RMSE									
$\rho = 0.2$, N=10	.021	.022	.021	.022	.022	.022	.022	.022	.022
$\rho = 0.5$, N=10	.019	.019	.019	.019	.019	.019	.018	.018	.018
$\rho = 0.8$, N=10	.015	.015	.015	.016	.016	.016	.014	.014	.014
$\rho = 0.2$, N=15	.019	.019	.019	.019	.019	.019	.018	.018	.018
$\rho = 0.5$, N=15	.017	.017	.017	.016	.017	.016	.015	.016	.016
$\rho = 0.8$, N=15	.014	.013	.013	.015	.014	.015	.012	.013	.013
Korelasyon									
$\rho = 0.2$, N=10	.54	.50	.54	.49	.48	.48	.49	.48	.49
$\rho = 0.5$, N=10	.65	.65	.63	.63	.63	.62	.66	.68	.68
$\rho = 0.8$, N=10	.80	.80	.79	.75	.75	.76	.82	.82	.82
$\rho = 0.2$, N=15	.66	.67	.67	.66	.65	.66	.68	.68	.70
$\rho = 0.5$, N=15	.74	.74	.73	.74	.73	.74	.78	.77	.74
$\rho = 0.8$, N=15	.85	.86	.85	.82	.83	.82	.87	.87	.87

Çizelge 4.11, Çizelge 4.12 ve Çizelge 4.13 incelendiğinde alt boyut puanları için tek boyutlu kalibrasyon yöntemi olan ATB kalibrasyonu kullanıldığında, kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.293 ile 0.547 arasında; V1 yöntemi için 0.292 ile 0.451 ve Vol yöntemi için ise 0.289 ile 0.464 arasında değiştiği belirlenmiştir. RMSE değerlerine bakıldığında hata değerlerinin KL yöntemi için

0.015 ile 0.023; V1 yöntemi için 0.015 ile 0.023 ve Vol yöntemi için ise 0.014 ile 0.023 arasında olduğu belirlenmiştir. Kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanlarına ait korelasyon değerleri ise KL yöntemi için 0.57 ile 0.82; V1 yöntemi için 0.54 ile 0.82 ve Vol yöntemi için ise 0.56 ile 0.82 arasında hesaplanmıştır.

Çizelge 4.12. Düşük Karmaşık Yapı Alt Boyut 2 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	ATB			*MHRM*			*BAEM*		
	KL	*V1*	*Vol*	*KL*	*V1*	*Vol*	*KL*	*V1*	*Vol*
ABSBIAS									
$\rho = 0.2$, **N=10**	.439	.439	.441	.550	.557	.555	.418	.424	.420
$\rho = 0.5$, **N=10**	.417	.400	.408	.499	.503	.511	.384	.388	.382
$\rho = 0.8$, **N=10**	.380	.373	.368	.400	.402	.399	.320	.311	.310
$\rho = 0.2$, **N=15**	.419	.406	.410	.533	.527	.528	.394	.392	.402
$\rho = 0.5$, **N=15**	.356	.352	.352	.456	.451	.456	.350	.329	.347
$\rho = 0.8$, **N=15**	.287	.290	.286	.353	.362	.359	.264	.268	.269
RMSE									
$\rho = 0.2$, **N=10**	.020	.021	.021	.023	.022	.023	.021	.021	.021
$\rho = 0.5$, **N=10**	.019	.019	.019	.020	.021	.020	.019	.020	.019
$\rho = 0.8$, **N=10**	.016	.016	.016	.016	.016	.016	.015	.015	.015
$\rho = 0.2$, **N=15**	.025	.024	.025	.023	.024	.023	.024	.023	.024
$\rho = 0.5$, **N=15**	.020	.020	.020	.020	.020	.020	.019	.019	.019
$\rho = 0.8$, **N=15**	.015	.015	.015	.016	.016	.016	.014	.014	.014
Korelasyon									
$\rho = 0.2$, **N=10**	.61	.59	.59	.49	.52	.47	.59	.59	.59
$\rho = 0.5$, **N=10**	.67	.67	.66	.61	.61	.61	.66	.65	.66
$\rho = 0.8$, **N=10**	.76	.77	.77	.75	.74	.76	.80	.79	.80
$\rho = 0.2$, **N=15**	.45	.44	.44	.46	.45	.46	.44	.46	.45
$\rho = 0.5$, **N=15**	.65	.66	.64	.65	.64	.65	.68	.68	.67
$\rho = 0.8$, **N=15**	.80	.80	.80	.78	.78	.80	.83	.83	.83

Alt boyut yetenek puanlarını kestirmede çok boyutlu kalibrasyon yöntemlerinden olan BAEM kalibrasyonu alt boyut yetenek puanlarını kestirmede kullanıldığında, kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.272 ile 0.418 arasında; V1 yöntemi için 0.273 ile 0.413 ve Vol yöntemi için ise 0.273 ile 0.420 arasında değişmektedir. RMSE değerlerine bakıldığında hata değerlerinin KL yöntemi için 0.014 ile 0.022; V1 yöntemi için 0.015 ile 0.023 ve Vol yöntemi için ise 0.014 ile 0.023 arasında olduğu belirlenmiştir. Kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanlarına ait korelasyon değerleri ise KL yöntemi için 0.53 ile 0.84; V1 yöntemi için 0.54 ile 0.82 ve Vol yöntemi için ise 0.56 ile 0.82 arasında hesaplanmıştır.

Alt boyut puanlarına ait Çizelge 4.11, Çizelge 4.12 ve Çizelge 4.13'deki değerler incelendiğinde maddelerin tek boyutlu (ATB) olarak kalibre edildiği durumlarda, V1

madde seçme yönteminin diğer madde seçme yöntemlerine göre daha düşük mutlak yanlılık, daha yüksek korelasyon değerine sahip olduğu görülmektedir. Çok boyutlu kalibrasyon yöntemi olan BAEM kalibrasyonunda ise V1 madde seçme yönteminin daha tutarlı sonuçlar verdiği görülmektedir. Elde edilen sonuçlar değerlendirildiğinde, alt boyut yetenek puanlarını kestirmede kullanılan madde seçme yöntemlerinin kalibrasyon yöntemlerine göre farklılaşmadığı görülmektedir. BAEM kalibrasyonu ile birlikte V1 madde seçme yöntemi, alt boyut yetenek puanlarını daha düşük yanlılık ve hata; daha yüksek korelasyon değerleri ile kestirerek var olan koşullar arasında en iyi performansı sergilemişlerdir.

Çizelge 4.13. Düşük Karmaşık Yapı Alt Boyut 3 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	ATB			*MHRM*			*BAEM*		
	KL	*V1*	*Vol*	*KL*	*V1*	*Vol*	*KL*	*V1*	*Vol*
ABSBIAS									
$\boldsymbol{\rho = 0.2}$**, N=10**	.463	.457	.463	.571	.554	.579	.405	.394	.405
$\boldsymbol{\rho = 0.5}$**, N=10**	.389	.401	.411	.466	.483	.473	.373	.360	.376
$\boldsymbol{\rho = 0.8}$**, N=10**	.357	.354	.365	.387	.385	.402	.314	.306	.319
$\boldsymbol{\rho = 0.2}$**, N=15**	.381	.379	.365	.555	.546	.558	.404	.397	.396
$\boldsymbol{\rho = 0.5}$**, N=15**	.344	.343	.340	.471	.485	.468	.343	.327	.346
$\boldsymbol{\rho = 0.8}$**, N=15**	.292	.291	.283	.341	.341	.349	.271	.270	.269
RMSE									
$\boldsymbol{\rho = 0.2}$**, N=10**	.021	.022	.022	.022	.022	.022	.022	.022	.023
$\boldsymbol{\rho = 0.5}$**, N=10**	.019	.019	.019	.020	.020	.020	.019	.019	.019
$\boldsymbol{\rho = 0.8}$**, N=10**	.015	.015	.016	.016	.017	.017	.014	.014	.014
$\boldsymbol{\rho = 0.2}$**, N=15**	.025	.026	.026	.025	.024	.025	.024	.024	.024
$\boldsymbol{\rho = 0.5}$**, N=15**	.021	.022	.021	.021	.021	.021	.020	.020	.021
$\boldsymbol{\rho = 0.8}$**, N=15**	.016	.016	.015	.016	.017	.017	.015	.015	.015
Korelasyon									
$\boldsymbol{\rho = 0.2}$**, N=10**	.55	.53	.56	.48	.48	.52	.51	.50	.50
$\boldsymbol{\rho = 0.5}$**, N=10**	.67	.66	.66	.60	.60	.62	.65	.65	.64
$\boldsymbol{\rho = 0.8}$**, N=10**	.79	.78	.78	.74	.74	.73	.82	.81	.81
$\boldsymbol{\rho = 0.2}$**, N=15**	.34	.35	.35	.33	.34	.32	.35	35	.36
$\boldsymbol{\rho = 0.5}$**, N=15**	.62	.60	.61	.61	.62	.61	.64	.65	.64
$\boldsymbol{\rho = 0.8}$**, N=15**	.80	.80	.80	.79	.78	.79	.81	.82	.82

Çizelge 4.14 incelendiğinde toplam yetenek puanları için tek boyutlu kalibrasyon yöntemi olan ATB kalibrasyonu kullanıldığında, kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.225 ile 0.334 arasında; V1 yöntemi için 0.227 ile 0.329 ve Vol yöntemi için ise 0.223 ile 0.329 arasında değiştirği belirlenmiştir. RMSE değerlerine bakıldığında hata değerlerinin KL, V1 ve Vol yöntemi için 0.010 ile 0.012 arasında olduğu belirlenmiştir. Kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanlarına ait korelasyon değerleri ise KL yöntemi için 0.77 ile

0.90; V1 yöntemi için 0.76 ile 0.90 ve Vol yöntemi için ise 0.76 ile 0.90 arasında hesaplanmıştır.

Toplam yetenek puanlarını kestirmede çok boyutlu kalibrasyon yöntemlerinden BAEM kalibrasyonu kullanıldığında, kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.191 ile 0.271 arasında; V1 yöntemi için 0.190 ile 0.268 ve Vol yöntemi için ise 0.185 ile 0.274 arasında değişmektedir. RMSE değerlerine bakıldığında hata değerlerinin KL, V1 ve Vol değerleri için 0.009 ile 0.011 arasında olduğu belirlenmiştir. Kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanlarına ait korelasyon değerleri ise KL yöntemi için 0.82 ile 0.91; V1 yöntemi için 0.81 ile 0.91 ve Vol yöntemi için ise 0.81 ile 0.92 arasında hesaplanmıştır.

Çizelge 4.14. Düşük Karmaşık Yapı Toplam Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	ATB			*MHRM*			*BAEM*		
	KL	*V1*	*Vol*	*KL*	*V1*	*Vol*	*KL*	*V1*	*Vol*
ABSBIAS									
$\boldsymbol{\rho = 0.2}$**, N=10**	.334	.329	.329	.263	.256	.259	.244	.251	.246
$\boldsymbol{\rho = 0.5}$**, N=10**	.302	.304	.310	.272	.281	.285	.271	.268	.274
$\boldsymbol{\rho = 0.8}$**, N=10**	.281	.274	.281	.268	.274	.283	.269	.258	.263
$\boldsymbol{\rho = 0.2}$**, N=15**	.283	.277	.279	.201	.208	.205	.191	.190	.185
$\boldsymbol{\rho = 0.5}$**, N=15**	.255	.252	.251	.222	.219	.221	.206	.202	.209
$\boldsymbol{\rho = 0.8}$**, N=15**	.225	.227	.223	.226	.229	.229	.213	.213	.209
RMSE									
$\boldsymbol{\rho = 0.2}$**, N=10**	.012	.012	.012	.012	.011	.012	.010	.010	.010
$\boldsymbol{\rho = 0.5}$**, N=10**	.012	.012	.012	.012	.013	.013	.011	.011	.011
$\boldsymbol{\rho = 0.8}$**, N=10**	.012	.012	.012	.013	.013	.013	.011	.011	.011
$\boldsymbol{\rho = 0.2}$**, N=15**	.010	.010	.010	.010	.010	.010	.009	.009	.009
$\boldsymbol{\rho = 0.5}$**, N=15**	.010	.010	.010	.011	.011	.011	.009	.009	.009
$\boldsymbol{\rho = 0.8}$**, N=15**	.010	.010	.010	.012	.012	.012	.010	.010	.010
Korelasyon									
$\boldsymbol{\rho = 0.2}$**, N=10**	.77	.76	.76	.76	.76	.76	.82	.81	.81
$\boldsymbol{\rho = 0.5}$**, N=10**	.82	.81	.81	.79	.79	.79	.84	.85	.85
$\boldsymbol{\rho = 0.8}$**, N=10**	.86	.86	.85	.82	.81	.81	.89	.88	.88
$\boldsymbol{\rho = 0.2}$**, N=15**	.83	.83	.83	.81	.81	.81	.85	.85	.86
$\boldsymbol{\rho = 0.5}$**, N=15**	.88	.87	.88	.85	.85	.85	.90	.90	.90
$\boldsymbol{\rho = 0.8}$**, N=15**	.90	.90	.90	.88	.87	.88	.91	.91	.92

Toplam yetenek puanlarına ait ait Çizelge 4.14'teki değerler incelendiğinde maddelerin tek boyutlu (ATB) olarak kalibre edildiği durumlarda, V1 madde seçme yönteminin diğer madde seçme yöntemlerine göre daha düşük mutlak yanlılık, daha yüksek korelasyon değerine sahip olduğu görülmektedir. Çok boyutlu BAEM kalibrasyonunda da aynı şekilde V1 madde seçme yönteminin daha tutarlı sonuçlar verdiği görülmektedir. Elde edilen sonuçlar değerlendirildiğinde, toplam yetenek puanlarını kestirmede kullanılan madde seçme yöntemlerinin kalibrasyon

yöntemlerine göre farklılaşmadığı görülmektedir. BAEM kalibrasyonu ile birlikte V1 madde seçme yöntemi, toplam yetenek puanlarını daha düşük yanlılık ve hata; daha yüksek korelasyon değerleri ile kestirerek var olan koşullar arasında en iyi performansı sergilemişlerdir.

4.5.2. Boyutlar Arası Korelasyonun ve Boyut Başına Düşen Madde Sayısının Etkisinin İncelenmesi

Araştırma sonuçları incelendiğinde madde seçme yöntemlerinin hata, yanlılık ve korelasyon değerlerinin boyutlar arasındaki korelasyon değerine (çok boyutluluk) ve boyut başına düşen madde sayısına (test uzunluğu) göre hem alt boyut hem de toplam yetenek puanlarında değiştiği gözlenmiştir. Bu bölümde, DKY desenindeki bir test için, kalibrasyon yöntemlerini de dikkate alarak, çok boyutluluğun ve test uzunluğunun madde seçme yöntemlerine etkisi incelenmektedir. Sonuçlar incelediğinde testin uzunluğu arttığında kestirilen yetenek puanları ile gerçek yetenek puanları arasındaki korelasyon değerlerinin arttığı; yetenek puanlarının mutlak yanlılık ve hata değerlerinin ise azaldığı belirlenmiştir. Ancak bu durum çok boyutluluk (boyutlar arası korelasyon) koşulu için geçerli değildir. Çok boyutluluk özelliklerine göre alt boyut ve toplam yetenek puanlarına ait mutlak yanlılık ve hata değerleri farklı durumlar sergilemiştir. Tüm koşullar için madde seçme yöntemlerinin tutarlılığı aşağıda ayrıntılı şekilde yorumlanmıştır.

Daha önce de belirtildiği gibi DKY test deseninde kalibrasyon yöntemlerinin madde seçme yöntemleri üzerinde etkisi olduğu ve her bir kalibrasyon yönteminin, çok boyutluluk ve test uzunluğu değişkenleri göz önüne alınmadığı durumda, belirli bir madde seçme yöntemine avantaj sağladığı görülmüştür.

Tek Boyutlu (ATB) Kalibrasyon Sonuçları

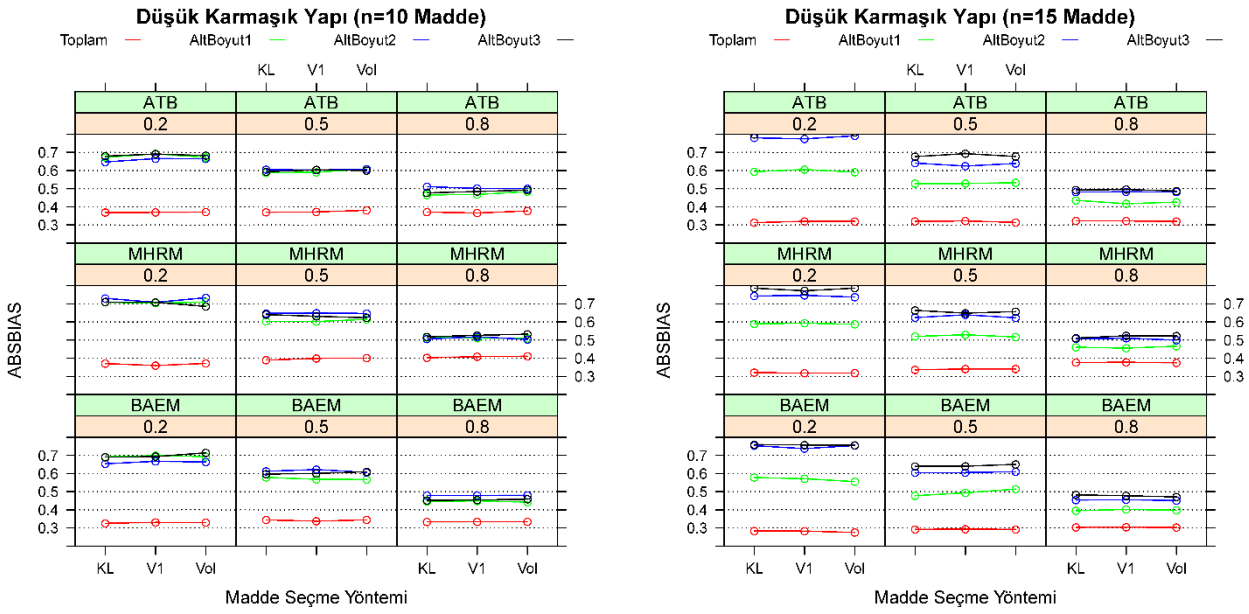
Düşük karmaşık yapılı bir testte alt boyut puanlarını kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde tek boyutlu kalibrasyon yönteminde (ATB) V1 madde seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.457, 0.451 ve 0.464 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.021, 0.022 ve 0.021; korelasyon değerleri ise sırasıyla 0.57,

0.54 ve 0.56 olarak hesaplanmıştır. Boyut başına düşen madde sayısının sabit tutulup (n=10), çok boyutluluğun etkisinin azaldığı durumda (ρ=0.8), mutlak yanlılık ve hata değerleri düşerken, korelasyon değerleri artmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.361, 0.353 ve 0.358 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.015, 0.015 ve 0.016; korelasyon değerleri ise 0.78 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=15 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.399, 0.395 ve 0.391 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.023; korelasyon değerleri ise 0.45 olarak hesaplanmıştır. Bu bulguların sonucuna bakıldığında testin katı çok boyutluluk (ρ=0.2) özelliği gösterdiği durumda, boyut başına düşen madde sayısının artırtılması durumunda mutlak yanlılık değerleri azalırken hatalar artmış, korelasyonlar da ise beklenenin aksine bir azalma görülmüştür. Bu durumun en büyük nedeni esas boyut dışındaki boyutlara düşen *a*-parametresi değerlerinin farklılaşması, yani test yapısının karmaşıklaşmasıdır. Ayrıntılı inceleme yapıldığında esas boyutlardaki RMSE ve ABSBIAS değerleri madde sayısı arttıkça azalırken, bilgi vericilik bakımından daha düşük olan diğer iki alt boyutta ise RMSE ve ABSBIAS değerlerinin arttığı görülmüştür. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.293, 0.292 ve 0.289 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.015, 0.015 ve 0.014; korelasyon değerleri ise 0.82 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen

madde sayısının n=15 olduğu durumda Vol; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise Vol madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.



Şekil 4.7. Düşük Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Göre Yanlılık Değerleri

Düşük karmaşık yapılı bir testte toplam puanların kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde tek boyutlu kalibrasyon yönteminde (ATB) V1 madde seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Toplam yetenek puanları kestirmede elde edilen bulgular yukarıda bahsedilen alt boyut yetenek puanlarını kestirmede elde edilen bulgularla kısmen benzer örüntüleri takip ediyor olsa da toplam yetenek puanları daha düşük mutlak yanlılık ve hata değerleri, daha yüksek korelasyon değerleri üretmiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.334, 0.329 ve 0.329 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.012; korelasyon değerleri ise sırasıyla 0.77, 0.76 ve 0.76 olarak hesaplanmıştır. Toplam yetenek puanlarında boyut başına düşen madde sayısının sabit olduğu ve çok boyutluluğun etkisinin azaldığı durumda, yani boyutlar arası korelasyonun 0.2'den 0.8'e çıktığı durumda, korelasyon değerleri artmış, RMSE ve ABSBIAS

değerleri de azalmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.281, 0.274 ve 0.281 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.012; korelasyon değerleri ise sırasıyla 0.86, 0.86 ve 0.85 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=15 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.283, 0.277 ve 0.279 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.010; korelasyon değerleri ise 0.83 olarak hesaplanmıştır. Bu bulguların sonucuna bakıldığında testin katı çok boyutluluk (ρ=0.2) özelliği gösterdiği durumda, boyut başına düşen madde sayısının artırtılması, toplam yetenek puanlarının kestirimini olumlu olarak etkileyecektir denilebilir. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.225, 0.227 ve 0.223 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.010; korelasyon değerleri ise 0.90 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise Vol madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.

Çok Boyutlu (BAEM) Kalibrasyon Sonuçları

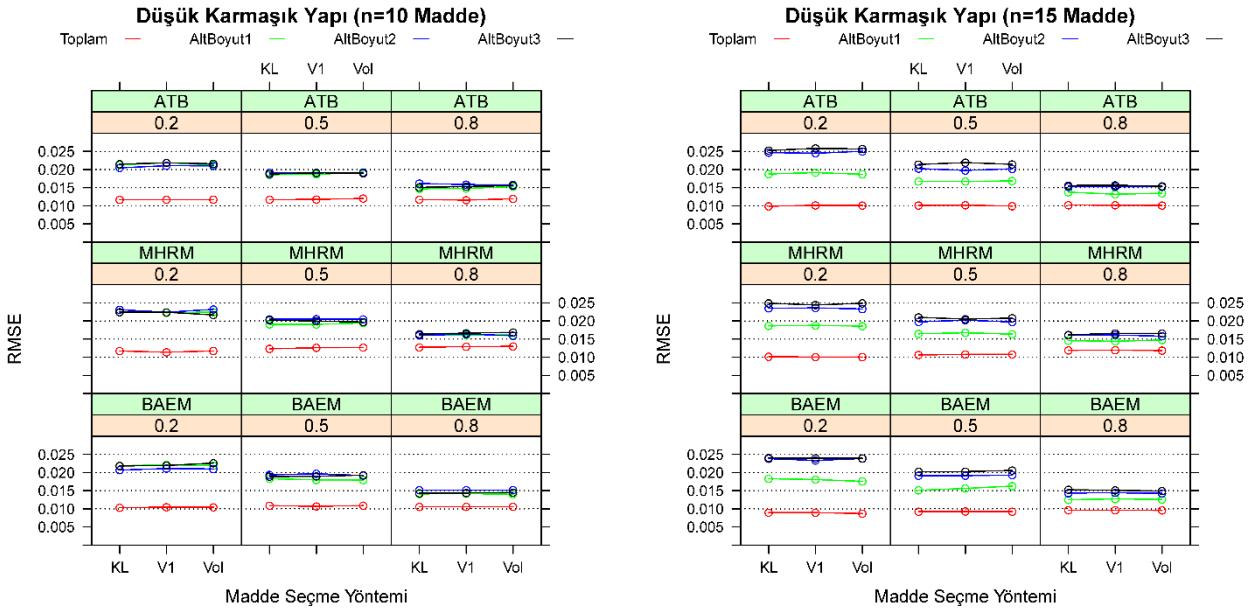
Düşük karmaşık yapılı bir testte alt boyut puanların kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde çok boyutlu kalibrasyon yönteminde (BAEM) V1 madde

seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.418, 0.413 ve 0.420 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.022; korelasyon değerleri ise sırasıyla 0.53, 0.52 ve 0.52 olarak hesaplanmıştır. Boyut başına düşen madde sayısının sabit tutulup (n=10), çok boyutluluğun etkisinin azaldığı durumda (ρ=0.8), mutlak yanlılık ve hata değerleri düşerken, korelasyon değerleri artmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.316, 0.309 ve 0.313 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.014; korelasyon değerleri ise 0.81 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=15 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.397, 0.395 ve 0.395 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri sıyasıyla 0.22, 0.21 ve 0.22; korelasyon değerleri ise sırasıyla 0.45, 0.46 ve 0.45 olarak hesaplanmıştır. Bu bulguların sonucuna bakıldığında testin katı çok boyutluluk (ρ=0.2) özelliği gösterdiği durumda, boyut başına düşen madde sayısının artırtılması, alt boyut yetenek puanlarının mutlak yanlılık ve hata kestirimlerine olumlu etki ederken, korelasyon değerlerine olumsuz etki etmiştir ve korelasyon değerlerini düşürmüştür. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.272, 0.273 ve 0.273 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.014; korelasyon değerleri ise 0.84 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası

korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.



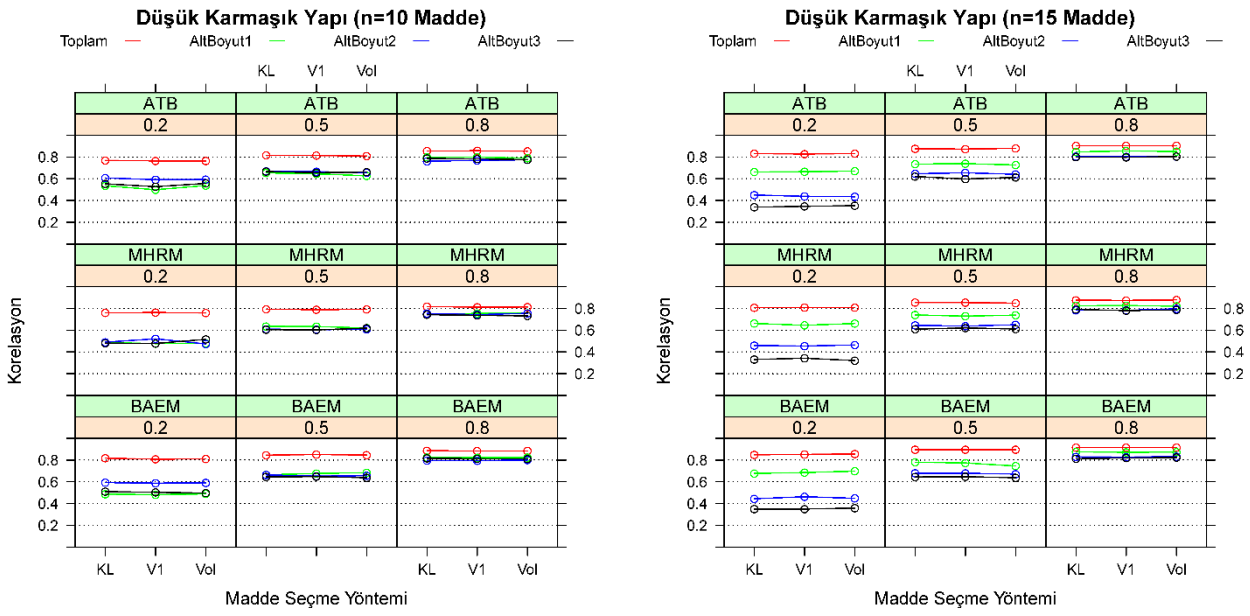
Şekil 4.8. Düşük Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Göre Hata Değerleri

Düşük karmaşık yapılı bir testte toplam puanların kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde çok boyutlu kalibrasyon yönteminde (BAEM) V1 madde seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Toplam yetenek puanları kestirmede elde edilen bulgular yukarıda bahsedilen alt boyut yetenek puanlarını kestirmede elde edilen bulgularla benzerlikler göstermemiş; daha düşük mutlak yanlılık ve hata değerleri, daha yüksek korelasyon değerleri üretmiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.244, 0.251 ve 0.246 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.010; korelasyon değerleri ise 0.82, 0.81 ve 0.81 olarak hesaplanmıştır. Tıpkı tek boyutlu (ATB) kalibrasyonda olduğu gibi, toplam yetenek puanlarında da çok boyutluluk değerlerine göre

ABSBIAS ve RMSE değerlerinde artışlar gözlenmiştir. Alt boyut yetenek puanlarının aksine, boyut başına düşen madde sayısının sabit olduğu ve çok boyutluluğun etkisinin azaldığı durumda, yani boyutlar arası korelasyonun 0.2'den 0.8'e çıktığı durumda, korelasyon değerleri artmış, RMSE ve ABSBIAS değerleri de artmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.269, 0.258 ve 0.263 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.011; korelasyon değerleri ise 0.89, 0.88 ve 0.88 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=15 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.191, 0.190 ve 0.185 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.009; korelasyon değerleri ise sırasıyla 0.85, 0.85 ve 0.86 olarak hesaplanmıştır. Bu bulguların sonucuna bakıldığında testin katı çok boyutluluk (ρ=0.2) özelliği gösterdiği durumda, boyut başına düşen madde sayısının artırtılması, toplam yetenek puanlarının kestirimini olumlu olarak etkileyecektir denilebilir. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.213, 0.213 ve 0.209 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.010; korelasyon değerleri ise 0.91, 0.91 ve 0.92 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda Vol; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise Vol madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.



Şekil 4.9. Düşük Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Göre Korelasyon Değerleri

4.6. Dördüncü Alt Probleme Ait Bulgular ve Yorumlar

Düşük Karmaşık Yapı test deseninde yetenek puanlarına ait mutlak yanlılık ve hata değerleri arasında koşullara göre anlamlı farklılıklar var mıdır?

Mutlak yanlılık ve hata değerlerinin koşullara göre etkilerinin incelenmesi amacıyla varyans analizi uygulanmış ve sonuçlara Çizelge 4.15'de yer verilmiştir.

Çizelge 4.15'te verilen ANOVA sonuçları incelendiğinde boyut başına düşen madde sayısının (BMS), boyutlar arası korelasyon (KOR) değerlerinin ve kalibrasyon yöntemlerinin (KY) hem alt boyut yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde hem de toplam yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde ana etkisinin anlamlı olduğu görülmüştür ($p < .05$). Bu durumun aksine çok boyutlu madde seçme yöntemlerinin (MYS) hem alt boyut yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde hem de toplam yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde anlamlı bir etkisinin olmadığı görülmüştür ($p > .05$).

Basit yapıdaki test desenindeki ANOVA sonuçlarının aksine, düşük karmaşık yapıdaki test desenindeki ANOVA sonuçlarında boyut başına düşen madde sayısı ve korelasyon etkileşimine ait değerlerin hem alt boyut yetenek puanlarına ait

mutlak yanlılık ve hata miktarları üzerinde hem de toplam yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde ana etkisinin anlamlı olduğu görülmüştür. Ayrıca boyut başına düşen madde sayısı-kalibrasyon yöntemi ve boyutlar arası korelasyon-kalibrasyon yöntemi ortak etkileşimlerinin hata miktarları arasında da anlamlı bir ilişki görülmüştür (p <.05).

Çizelge 4.15. Düşük Karmaşık Yapı Test Deseni Yetenek Puanlarına Ait Mutlak Yanlılık ve Hata Değerleri ANOVA Sonuçları

			Alt Boyut 1 Puanı		*Alt Boyut 2 Puanı*		*Alt Boyut 3 Puanı*		*Toplam Puan*	
		sd	***F***	η^2	***F***	η^2	***F***	η^2	***F***	η^2
	BMS	1	7635.5*	.99	324.49*	.99	211.69*	.99	410.96*	.99
	KOR	2	1104.6*	.99	3314.8*	.99	3084.4*	.99	562.08*	.96
	KY	2	407.18*	.95	124.10*	.98	34.365*	.94	1871.1*	.99
RMSE	***MSY***	2	-	-	-	-	-	-	-	-
	BMS*KOR	2	94.82*	.97	637.75*	.99	170.68*	.98	175.48*	.98
	BMS*KY	2	7.46*	.78	24.04*	.92	18.65*	.90	334.59*	.94
	KOR*KY	4	19.26*	.90	-	-	11.49*	.85	47.68*	.96
	BMS	1	11423*	.99	272.19*	.99	588.13*	.99	1594.6*	.99
	KOR	2	1050.6*	.99	3112.1*	.99	11913*	.99	1773.6*	.99
	KY	2	364.01*	.99	466.99*	.99	78.56*	.97	7226.7*	.99
ABSBIAS	***MSY***	2	-	-	-	-	-	-	-	-
	BMS*KOR	2	285.68*	.99	1065.0*	.99	396.56*	.99	37.60*	.94
	BMS*KY	2	-	-	32.70*	.94	14.80*	.88	85.74*	.97
	KOR*KY	4	25.13*	.92	4.15*	.67	16.01*	.88	18.71*	.90

Not: BMS=Boyut başına düşen madde sayısı, KOR=Boyutlar arası korelasyon KY=Kalibrasyon yöntemi, MSY=Madde seçme yöntemi. *=0.05 düzeyinde anlamlı,

4.7. Beşinci Alt Probleme Ait Bulgular ve Yorumlar

Yüksek Karmaşık Yapı (YKY) test deseninde çok boyutlu madde seçme yöntemlerinin hatası, yanlılığı ve korelasyon değerleri kalibrasyon yöntemine, boyutlar arası korelasyona ve boyut başına düşen madde sayısına göre alt boyut ve toplam yetenek puanları için nasıl değişmektedir?

Bu bölüme ait bulgularda yüksek karmaşık yapı test desenine ait ÇBBBT uygulamaları sonucu elde edilen kestirilen yetenek puanları ile gerçek yetenek puanları karşılaştırılarak hata, yanlılık ve korelasyon değerleri belirlenen koşullar bakımından karşılaştırılmıştır. Bulgular sonucunda var olan koşullar altında en iyi performansı gösteren madde seçme yöntemi de rapor edilmiştir.

4.7.1. Kalibrasyon Yönteminin Etkisinin İncelenmesi

Yüksek karmaşık yapıdaki bir testte, hem basit hem de düşük karmaşık yapıda oluğu gibi, madde seçme yöntemlerinin hata, yanlılık ve korelasyon değerlerinin kalibrasyon yöntemlerine göre hem alt boyut hem de toplam yetenek puanlarında değiştiği gözlenmiştir.

Çizelge 4.16. Yüksek Karmaşık Yapı Alt Boyut 1 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	ATB			*MHRM*			*BAEM*		
	KL	*V1*	*Vol*	*KL*	*V1*	*Vol*	*KL*	*V1*	*Vol*
ABSBIAS									
$\boldsymbol{\rho = 0.2}$**, N=10**	.488	.461	.505	.594	.576	.575	.421	.419	.483
$\boldsymbol{\rho = 0.5}$**, N=10**	.394	.388	.373	.500	.494	.501	.354	.353	.372
$\boldsymbol{\rho = 0.8}$**, N=10**	.350	.362	.368	.372	.389	.386	.305	.298	.312
$\boldsymbol{\rho = 0.2}$**, N=15**	.387	.394	.400	.554	.571	.561	.392	.397	.402
$\boldsymbol{\rho = 0.5}$**, N=15**	.368	.364	.369	.462	.466	.457	.338	.335	.340
$\boldsymbol{\rho = 0.8}$**, N=15**	.290	.291	.287	.379	.372	.374	.375	.372	.375
RMSE									
$\boldsymbol{\rho = 0.2}$**, N=10**	.023	.023	.023	.024	.023	.023	.022	.022	.023
$\boldsymbol{\rho = 0.5}$**, N=10**	.020	.020	.020	.020	.020	.020	.019	.019	.019
$\boldsymbol{\rho = 0.8}$**, N=10**	.016	.015	.016	.017	.017	.017	.015	.015	.015
$\boldsymbol{\rho = 0.2}$**, N=15**	.021	.021	.021	.020	.021	.021	.020	.020	.020
$\boldsymbol{\rho = 0.5}$**, N=15**	.018	.018	.018	.017	.017	.017	.017	.016	.017
$\boldsymbol{\rho = 0.8}$**, N=15**	.014	.014	.014	.015	.015	.015	.013	.013	.013
Korelasyon									
$\boldsymbol{\rho = 0.2}$**, N=10**	.46	.47	.47	.44	.46	.46	.51	.51	.50
$\boldsymbol{\rho = 0.5}$**, N=10**	.68	.67	.66	.67	.67	.67	.71	.70	.69
$\boldsymbol{\rho = 0.8}$**, N=10**	.79	.79	.78	.75	.75	.74	.81	.82	.81
$\boldsymbol{\rho = 0.2}$**, N=15**	.59	.58	.57	.59	.58	.57	.61	.61	.62
$\boldsymbol{\rho = 0.5}$**, N=15**	.70	.71	.71	.72	.72	.73	.76	.78	.75
$\boldsymbol{\rho = 0.8}$**, N=15**	.83	.83	.82	.81	.81	.80	.86	.85	.85

Çizelge 4.16, Çizelge 4.17 ve Çizelge 4.18 incelendiğinde alt boyut puanları için tek boyutlu kalibrasyon yöntemi olan ATB kalibrasyonu kullanıldığında, kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.288 ile 0.463 arasında; V1 yöntemi için 0.290 ile 0.457 ve Vol yöntemi için ise 0.290 ile 0.463 arasında değiştiği görülmektedir. RMSE değerlerine bakıldığında hata değerlerinin KL, V1 ve Vol yöntemleri için 0.014 ile 0.023 arasında olduğu belirlenmiştir. Kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanlarına ait korelasyon değerleri ise KL yöntemi için 0.49 ile 0.82; V1 yöntemi için 0.49 ile 0.82 ve Vol yöntemi için ise 0.50 ile 0.81 arasında hesaplanmıştır.

Alt boyut yetenek puanlarını kestirmede çok boyutlu kalibrasyon yöntemlerinden olan BAEM kalibrasyonu alt boyut yetenek puanlarını kestirmede kullanıldığında, kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanları

arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.303 ile 0.414 arasında; V1 yöntemi için 0.303 ile 0.423 ve Vol yöntemi için ise 0.306 ile 0.440 arasında değişmektedir. RMSE değerlerine bakıldığında hata değerlerinin KL yöntemi için 0.014 ile 0.022; V1 yöntemi için 0.013 ile 0.023 ve Vol yöntemi için ise 0.013 ile 0.023 arasında olduğu belirlenmiştir. Kestirilen alt boyut yetenek puanları ile gerçek alt boyut yetenek puanlarına ait korelasyon değerleri ise KL, V1 ve Vol yöntemleri için 0.49 ile 0.84 arasında hesaplanmıştır.

Çizelge 4.17. Yüksek Karmaşık Yapı Alt Boyut 2 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	ATB			*MHRM*			*BAEM*		
	KL	*V1*	*Vol*	*KL*	*V1*	*Vol*	*KL*	*V1*	*Vol*
ABSBIAS									
$\boldsymbol{\rho = 0.2}$**, N=10**	.436	.450	.431	.533	.554	.540	.404	.412	.408
$\boldsymbol{\rho = 0.5}$**, N=10**	.398	.401	.397	.493	.499	.497	.352	.357	.371
$\boldsymbol{\rho = 0.8}$**, N=10**	.340	.351	.346	.374	.371	.380	.303	.310	.315
$\boldsymbol{\rho = 0.2}$**, N=15**	.394	.394	.398	.536	.536	.536	.389	.411	.428
$\boldsymbol{\rho = 0.5}$**, N=15**	.354	.317	.340	.467	.462	.459	.342	.349	.366
$\boldsymbol{\rho = 0.8}$**, N=15**	.284	.292	.287	.352	.346	.342	.279	.281	.287
RMSE									
$\boldsymbol{\rho = 0.2}$**, N=10**	.022	.023	.022	.023	.023	.023	.023	.024	.023
$\boldsymbol{\rho = 0.5}$**, N=10**	.019	.019	.019	.019	.019	.020	.019	.019	.019
$\boldsymbol{\rho = 0.8}$**, N=10**	.016	.016	.016	.017	.017	.017	.015	.015	.015
$\boldsymbol{\rho = 0.2}$**, N=15**	.022	.022	.022	.022	.022	.022	.021	.022	.022
$\boldsymbol{\rho = 0.5}$**, N=15**	.020	.019	.020	.020	.019	.019	.018	.018	.019
$\boldsymbol{\rho = 0.8}$**, N=15**	.014	.014	.014	.015	.015	.015	.014	.013	.013
Korelasyon									
$\boldsymbol{\rho = 0.2}$**, N=10**	.51	.49	.51	.45	.44	.46	.47	.45	.47
$\boldsymbol{\rho = 0.5}$**, N=10**	.65	.66	.65	.64	.65	.64	.66	.68	.66
$\boldsymbol{\rho = 0.8}$**, N=10**	.78	.78	.77	.74	.74	.73	.80	.80	.80
$\boldsymbol{\rho = 0.2}$**, N=15**	.56	.57	.55	.52	.52	.50	.56	.54	.51
$\boldsymbol{\rho = 0.5}$**, N=15**	.67	.69	.67	.64	.66	.66	.69	.70	.68
$\boldsymbol{\rho = 0.8}$**, N=15**	.81	.82	.81	.80	.79	.79	.83	.84	.84

Alt boyut puanlarına ait Çizelge 4.16, Çizelge 4.17 ve Çizelge 4.18'deki değerler incelendiğinde maddelerin tek boyutlu (ATB) olarak kalibre edildiği durumlarda, V1 madde seçme yönteminin diğer madde seçme yöntemlerine göre daha düşük mutlak yanlılık, daha yüksek korelasyon değerine sahip olduğu görülmektedir. Çok boyutlu kalibrasyon yöntemi olan BAEM kalibrasyonunda ise KL madde seçme yönteminin daha tutarlı sonuçlar verdiği görülmektedir. Elde edilen sonuçlar değerlendirildiğinde, alt boyut yetenek puanlarını kestirmede kullanılan madde seçme yöntemlerinin kalibrasyon yöntemlerine göre farklılaştığı görülmektedir. BAEM kalibrasyonu ile birlikte KL madde seçme yöntemi, alt boyut yetenek puanlarını daha düşük yanlılık ve hata; daha yüksek korelasyon değerleri ile kestirerek var olan koşullar arasında en iyi performansı sergilemişlerdir.

Çizelge 4.18. Yüksek Karmaşık Yapı Alt Boyut 3 Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	ATB			MHRM			BAEM		
	KL	V1	Vol	KL	V1	Vol	KL	V1	Vol
ABSBIAS									
$\boldsymbol{\rho = 0.2}$**, N=10**	.468	.460	.460	.551	.572	.576	.417	.440	.435
$\boldsymbol{\rho = 0.5}$**, N=10**	.387	.382	.374	.475	.475	.481	.377	.378	.390
$\boldsymbol{\rho = 0.8}$**, N=10**	.350	.350	.359	.412	.416	.428	.320	.316	.333
$\boldsymbol{\rho = 0.2}$**, N=15**	.384	.374	.387	.539	.530	.541	.385	.371	.383
$\boldsymbol{\rho = 0.5}$**, N=15**	.316	.337	.330	.469	.481	.469	.351	.349	.338
$\boldsymbol{\rho = 0.8}$**, N=15**	.291	.286	.297	.352	.353	.355	.275	.274	.274
RMSE									
$\boldsymbol{\rho = 0.2}$**, N=10**	.023	.022	.022	.022	.022	.023	.022	.022	.022
$\boldsymbol{\rho = 0.5}$**, N=10**	.019	.019	.019	.019	.019	.019	.018	.018	.018
$\boldsymbol{\rho = 0.8}$**, N=10**	.016	.016	.016	.016	.017	.017	.014	.015	.015
$\boldsymbol{\rho = 0.2}$**, N=15**	.025	.025	.025	.024	.024	.024	.023	.024	.023
$\boldsymbol{\rho = 0.5}$**, N=15**	.021	.021	.021	.020	.020	.020	.019	.019	.019
$\boldsymbol{\rho = 0.8}$**, N=15**	.015	.015	.015	.015	.016	.016	.014	.014	.014
Korelasyon									
$\boldsymbol{\rho = 0.2}$**, N=10**	.49	.51	.51	.48	.49	.47	.49	.50	.49
$\boldsymbol{\rho = 0.5}$**, N=10**	.66	.66	.66	.65	.65	.66	.68	.67	.67
$\boldsymbol{\rho = 0.8}$**, N=10**	.78	.79	.78	.76	.75	.75	.82	.82	.81
$\boldsymbol{\rho = 0.2}$**, N=15**	.42	.43	.43	.44	.45	.43	.44	.44	.47
$\boldsymbol{\rho = 0.5}$**, N=15**	.61	.59	.59	.61	.60	.61	.62	.63	.64
$\boldsymbol{\rho = 0.8}$**, N=15**	.81	.80	.80	.79	.78	.79	.83	.82	.82

Çizelge 4.19 incelendiğinde toplam yetenek puanları için tek boyutlu kalibrasyon yöntemi olan ATB kalibrasyonu kullanıldığında, kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.223 ile 0.332 arasında; V1 yöntemi için 0.218 ile 0.339 ve Vol yöntemi için ise 0.222 ile 0.334 arasında değiştiği belirlenmiştir. RMSE değerlerine bakıldığında hata değerlerinin KL, V1 ve Vol yöntemi için 0.009 ile 0.012 arasında olduğu belirlenmiştir. Kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanlarına ait korelasyon değerleri ise KL, V1 ve Vol yöntemleri için 0.78 ile 0.91 arasında hesaplanmıştır.

Toplam yetenek puanlarını kestirmede çok boyutlu kalibrasyon yöntemlerinden BAEM kalibrasyonu kullanıldığında, kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanları arasındaki mutlak yanlılık değerleri (ABSBIAS) KL yöntemi için 0.185 ile 0.232 arasında; V1 yöntemi için 0.188 ile 0.248 ve Vol yöntemi için ise 0.189 ile 0.259 arasında değişmektedir. RMSE değerlerine bakıldığında hata değerlerinin KL, V1 ve Vol değerleri için 0.008 ile 0.011 arasında olduğu belirlenmiştir. Kestirilen toplam yetenek puanları ile gerçek toplam yetenek puanlarına ait korelasyon değerleri ise KL yöntemi için 0.83 ile 0.92; V1 yöntemi için 0.83 ile 0.92 ve Vol yöntemi için ise 0.82 ile 0.92 arasında hesaplanmıştır.

Çizelge 4.19. Yüksek Karmaşık Yapı Toplam Yetenek Puanlarının Kestirime Ait ABSBIAS, RMSE ve Korelasyon Değerleri

	ATB			*MHRM*			*BAEM*		
	KL	*V1*	*Vol*	*KL*	*V1*	*Vol*	*KL*	*V1*	*Vol*
ABSBIAS									
$\boldsymbol{\rho = 0.2}$**, N=10**	.332	.339	.334	.252	.251	.253	.232	.248	.240
$\boldsymbol{\rho = 0.5}$**, N=10**	.296	.297	.292	.261	.260	.254	.250	.251	.259
$\boldsymbol{\rho = 0.8}$**, N=10**	.274	.281	.284	.272	.275	.289	.263	.261	.270
$\boldsymbol{\rho = 0.2}$**, N=15**	.279	.272	.279	.201	.204	.198	.185	.188	.189
$\boldsymbol{\rho = 0.5}$**, N=15**	.245	.242	.245	.213	.211	.214	.204	.204	.200
$\boldsymbol{\rho = 0.8}$**, N=15**	.223	.218	.222	.219	.214	.222	.207	.206	.210
RMSE									
$\boldsymbol{\rho = 0.2}$**, N=10**	.011	.011	.011	.011	.011	.011	.009	.010	.010
$\boldsymbol{\rho = 0.5}$**, N=10**	.011	.011	.011	.012	.012	.012	.010	.010	.010
$\boldsymbol{\rho = 0.8}$**, N=10**	.012	.012	.012	.013	.013	.013	.011	.011	.011
$\boldsymbol{\rho = 0.2}$**, N=15**	.010	.009	.009	.009	.009	.009	.008	.008	.008
$\boldsymbol{\rho = 0.5}$**, N=15**	.009	.009	.010	.010	.010	.010	.009	.009	.008
$\boldsymbol{\rho = 0.8}$**, N=15**	.009	.009	.009	.011	.011	.011	.009	.009	.009
Korelasyon									
$\boldsymbol{\rho = 0.2}$**, N=10**	.78	.78	.78	.77	.76	.77	.83	.83	.82
$\boldsymbol{\rho = 0.5}$**, N=10**	.85	.85	.85	.83	.84	.84	.88	.87	.87
$\boldsymbol{\rho = 0.8}$**, N=10**	.86	.87	.85	.82	.82	.81	.89	.89	.88
$\boldsymbol{\rho = 0.2}$**, N=15**	.86	.86	.86	.84	.84	.84	.88	.88	.88
$\boldsymbol{\rho = 0.5}$**, N=15**	.88	.88	.88	.86	.86	.86	.90	.90	.90
$\boldsymbol{\rho = 0.8}$**, N=15**	.91	.91	.91	.88	.87	.87	.92	.92	.92

Toplam yetenek puanlarına ait ait Çizelge 4.19'daki değerler incelendiğinde maddelerin tek boyutlu (ATB) olarak kalibre edildiği durumlarda, V1 madde seçme yönteminin diğer madde seçme yöntemlerine göre daha düşük mutlak yanlılık, daha yüksek korelasyon değerine sahip olduğu görülmektedir. Çok boyutlu BAEM kalibrasyonunda da aynı şekilde KL madde seçme yönteminin daha tutarlı sonuçlar verdiği görülmektedir. Elde edilen sonuçlar değerlendirildiğinde, toplam yetenek puanlarını kestirmede kullanılan madde seçme yöntemlerinin kalibrasyon yöntemlerine göre farklılaşmadığı görülmektedir. BAEM kalibrasyonu ile birlikte KL madde seçme yöntemi, toplam yetenek puanlarını daha düşük yanlılık ve hata; daha yüksek korelasyon değerleri ile kestirerek var olan koşullar arasında en iyi performansı sergilemişlerdir.

4.7.2. Boyutlar Arası Korelasyonun ve Boyut Başına Düşen Madde Sayısının Etkisinin İncelenmesi

Araştırma sonuçları incelendiğinde madde seçme yöntemlerinin hata, yanlılık ve korelasyon değerlerinin boyutlar arasındaki korelasyon değerine (çok boyutluluk) ve boyut başına düşen madde sayısına (test uzunluğu) göre hem alt boyut hem de toplam yetenek puanlarında değiştiği gözlenmiştir. Bu bölümde, YKY desenindeki bir test için, kalibrasyon yöntemlerini de dikkate alarak, çok boyutluluğun ve test

uzunluğunun madde seçme yöntemlerine etkisi incelenmektedir. Sonuçlar incelediğinde testin uzunluğu arttığında kestirilen yetenek puanları ile gerçek yetenek puanları arasındaki korelasyon değerlerinin arttığı; yetenek puanlarının mutlak yanlılık ve hata değerlerinin ise azaldığı belirlenmiştir. Ancak bu durum çok boyutluluk (boyutlar arası korelasyon) koşulu için geçerli değildir. Çok boyutluluk özelliklerine göre alt boyut ve toplam yetenek puanlarına ait mutlak yanlılık ve hata değerleri farklı durumlar sergilemiştir. Tüm koşullar için madde seçme yöntemlerinin tutarlılığı aşağıda ayrıntılı şekilde yorumlanmıştır.

Daha önce de belirtildiği gibi YKY test deseninde kalibrasyon yöntemlerinin madde seçme yöntemleri üzerinde etkisi olduğu ve her bir kalibrasyon yönteminin, çok boyutluluk ve test uzunluğu değişkenleri göz önüne alınmadığı durumda, belirli bir madde seçme yöntemine avantaj sağladığı görülmüştür.

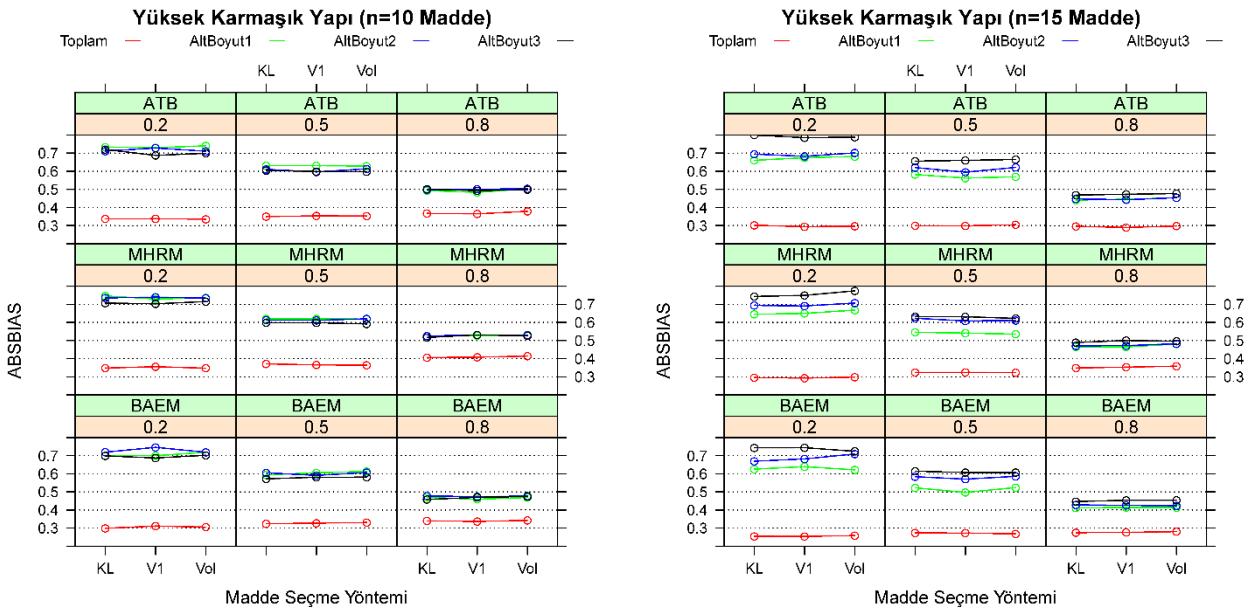
Tek Boyutlu (ATB) Kalibrasyon Sonuçları

Yüksek karmaşık yapılı bir testte alt boyut puanların kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde tek boyutlu kalibrasyon yönteminde (ATB) V1 madde seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.463, 0.457 ve 0.463 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.023, 0.023 ve 0.022; korelasyon değerleri ise sırasıyla 0.49, 0.49 ve 0.50 olarak hesaplanmıştır. Boyut başına düşen madde sayısının sabit tutulup (n=10), çok boyutluluğun etkisinin azaldığı durumda (ρ=0.8), mutlak yanlılık ve hata değerleri düşerken, korelasyon değerleri artmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.347, 0.354 ve 0.357 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.016; korelasyon değerleri ise 0.78, 0.79 ve 0.78 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde

sayısının n=15 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.388, 0.387 ve 0.395 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.023; korelasyon değerleri ise 0.51, 0.52 ve 0.51 olarak hesaplanmıştır. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.288, 0.290 ve 0.290 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.014; korelasyon değerleri ise 0.82, 0.82, ve 0.81 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.



Şekil 4.10. Yüksek Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Göre Yanlılık Değerleri

Yüksek karmaşık yapılı bir testte toplam puanların kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde tek boyutlu kalibrasyon yönteminde (ATB) V1 madde seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Toplam yetenek puanları kestirmede elde edilen bulgular yukarıda bahsedilen alt boyut yetenek puanlarını kestirmede elde edilen bulgularla hemen hemen benzer örüntüleri takip ediyor olsa da toplam yetenek puanları daha düşük mutlak yanlılık ve hata değerleri, daha yüksek korelasyon değerleri üretmiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.332, 0.339 ve 0.334 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.011; korelasyon değerleri ise 0.78 olarak hesaplanmıştır. Toplam yetenek puanlarında boyut başına düşen madde sayısının sabit olduğu ve çok boyutluluğun etkisinin azaldığı durumda, yani boyutlar arası korelasyonun 0.2'den 0.8'e çıktığı durumda, korelasyon değerleri artmış, RMSE ve ABSBIAS değerleri de azalmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.274, 0.281 ve 0.284 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.012; korelasyon değerleri ise sırasıyla 0.86, 0.87 ve 0.85 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=15 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.279, 0.272 ve 0.279 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.010, 0.09 ve 0.09; korelasyon değerleri ise 0.86 olarak hesaplanmıştır. Bu bulguların sonucuna bakıldığında testin katı çok boyutluluk (ρ=0.2) özelliği gösterdiği durumda, boyut başına düşen madde sayısının artırtılması, toplam yetenek puanlarının kestirimini olumlu olarak etkileyecektir denilebilir. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise toplam yetenek puanlarının KL, V1 ve Vol değerleri için

ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.233, 0.218 ve 0.220 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.009; korelasyon değerleri ise 0.91 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise V1 madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.

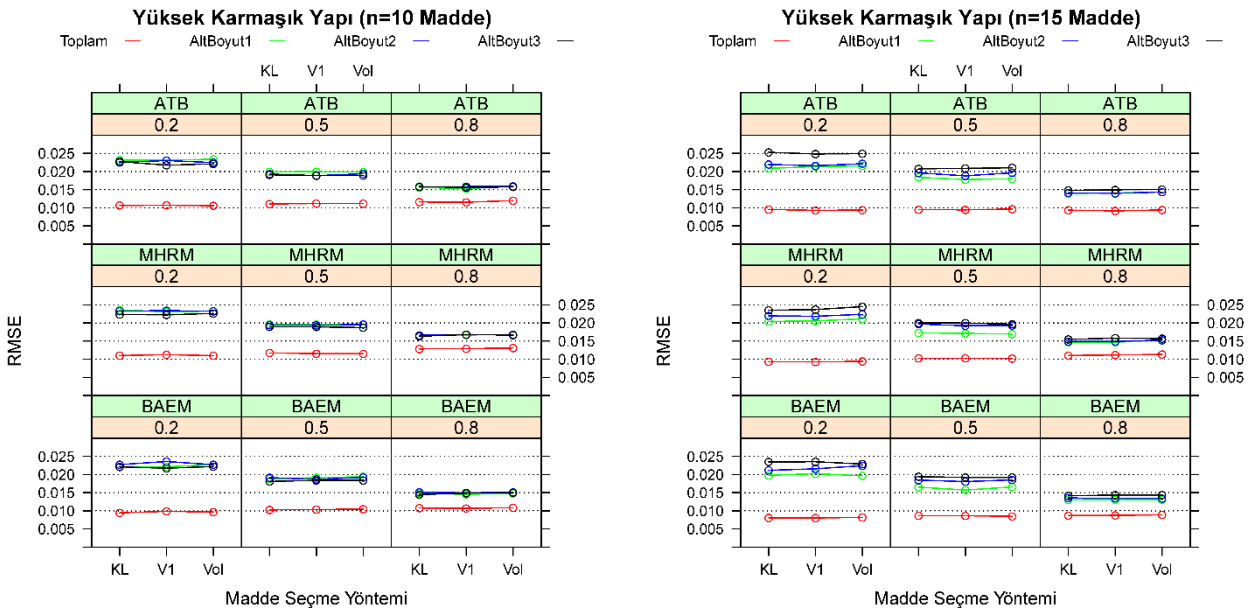
Çok Boyutlu (BAEM) Kalibrasyon Sonuçları

Yüksek karmaşık yapılı bir testte alt boyut puanların kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde çok boyutlu kalibrasyon yönteminde (BAEM) KL madde seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.414, 0.423 ve 0.440 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.022, 0.023 ve 0.023; korelasyon değerleri ise 0.49 olarak hesaplanmıştır. Boyut başına düşen madde sayısının sabit tutulup (n=10), çok boyutluluğun etkisinin azaldığı durumda (ρ=0.8), mutlak yanlılık ve hata değerleri düşerken, korelasyon değerleri artmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.309, 0.308 ve 0.320 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.015; korelasyon değerleri ise 0.81 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=15 olduğu durumda alt boyut yetenek puanlarının KL, V1 ve Vol

değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.389, 0.392 ve 0.403 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri sıyasıyla 0.021, 0.022 ve 0.022; korelasyon değerleri ise sırasıyla 0.53, 0.52 ve 0.53 olarak hesaplanmıştır. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise alt boyut yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.303, 0.303 ve 0.306 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.014, 0.013 ve 0.013; korelasyon değerleri ise 0.84 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise V1 madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.



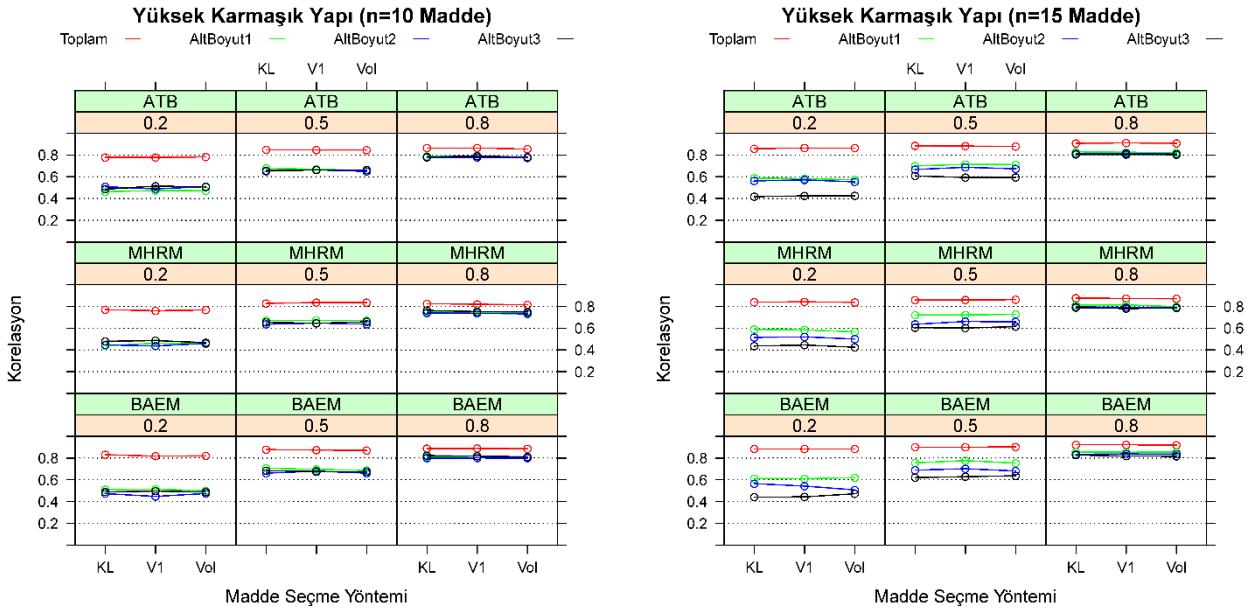
Şekil 4.11. Yüksek Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Göre Hata Değerleri

Yüksek karmaşık yapılı bir testte toplam puanların kestirmede, kalibrasyon yöntemlerine göre madde seçme yöntemlerinin mutlak yanlılık, hata ve korelasyon değerleri incelendiğinde çok boyutlu kalibrasyon yönteminde (BAEM) KL madde seçme yönteminin diğer iki yöntemden daha iyi performans gösterdiğini belirtilmişti. Toplam yetenek puanları kestirmede elde edilen bulgular yukarıda bahsedilen alt boyut yetenek puanlarını kestirmede elde edilen bulgularla benzerlikler göstermiş; daha düşük mutlak yanlılık ve hata değerleri, daha yüksek korelasyon değerleri üretmiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=10 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.232, 0.248 ve 0.240 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.009, 0.010 ve 0.010; korelasyon değerleri ise 0.83, 0.83 ve 0.82 olarak hesaplanmıştır. Tıpkı tek boyutlu (ATB) kalibrasyonda olduğu gibi, toplam yetenek puanlarında da çok boyutluluk değerlerine göre ABSBIAS ve RMSE değerlerinde artışlar gözlenmiştir. Alt boyut yetenek puanlarının aksine, boyut başına düşen madde sayısının sabit olduğu ve çok boyutluluğun etkisinin azaldığı durumda, yani boyutlar arası korelasyonun 0.2'den 0.8'e çıktığı durumda, korelasyon değerleri artmış, RMSE ve ABSBIAS değerleri de artmıştır. Testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.263, 0.261 ve 0.270 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.011; korelasyon değerleri ise 0.89, 0.89 ve 0.88 olarak hesaplanmıştır.

Testin uzunluğu artırıldığında beklenildiği gibi test daha kesin sonuçlar vermiştir. Testin katı çok boyutluluk (ρ=0.2) özeliği gösterdiği ve boyut başına düşen madde sayısının n=15 olduğu durumda toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.185, 0188 ve 0.189 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri 0.008; korelasyon değerleri ise 0.88 olarak hesaplanmıştır. Bu bulguların sonucuna bakıldığında testin katı çok boyutluluk (ρ=0.2) özelliği gösterdiği durumda, boyut başına düşen madde sayısının artırtılması, toplam yetenek puanlarının kestirimini olumlu olarak etkileyecektir denilebilir. Son olarak, testin çok boyutluluğunun azaldığı (ρ=0.8) ve boyut başına düşen madde sayısının n=15

olduğu durumda ise toplam yetenek puanlarının KL, V1 ve Vol değerleri için ortalama mutlak yanlılık (ABSBIAS) değerleri sırasıyla 0.207, 0.206 ve 0.210 olarak hesaplanmıştır. Her üç madde seçme yöntemine ait RMSE değerleri de 0.009; korelasyon değerleri ise 0.92 olarak hesaplanmıştır.

ATB kalibrasyon yöntemine ait toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise V1 madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.



Şekil 4.12. Yüksek Karmaşık Yapılı Testlerde Tüm Yetenek Puanlarının Madde Seçme Yöntemlerine Göre Korelasyon Değerleri

4.8. Altıncı Alt Probleme Ait Bulgular ve Yorumlar

Yüksek Karmaşık Yapı test deseninde yetenek puanlarına ait mutlak yanlılık ve hata değerleri arasında koşullara göre anlamlı farklılıklar var mıdır?

Mutlak yanlılık ve hata değerlerinin koşullara göre etkilerinin incelenmesi amacıyla varyans analizi uygulanmış ve sonuçlara Çizelge 4.20'de yer verilmiştir.

Çizelge 4.20'de verilen ANOVA sonuçları incelendiğinde boyut başına düşen madde sayısının (BMS), boyutlar arası korelasyon (KOR) değerlerinin ve kalibrasyon yöntemlerinin (KY) hem alt boyut yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde hem de toplam yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde ana etkisinin anlamlı olduğu görülmüştür (p <.05). Bu durumun aksine çok boyutlu madde seçme yöntemlerinin (MYS) hem alt boyut yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde hem de toplam yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde anlamlı bir etkisinin olmadığı görülmüştür (p >.05).

Çizelge 4.20. Yüksek Karmaşık Yapı Test Deseni Yetenek Puanlarına Ait Mutlak Yanlılık ve Hata Değerleri ANOVA Sonuçları

			Alt Boyut 1 Puanı		*Alt Boyut 2 Puanı*		*Alt Boyut 3 Puanı*		*Toplam Puan*	
		sd	*F*	η^2	*F*	η^2	*F*	η^2	*F*	η^2
	BMS	1	13253*	.99	53.52*	.96	380.99*	.99	2751.7*	.99
	KOR	2	4230.7*	.99	1342.9*	.99	2315.0*	.99	163.94*	.98
	KY	2	2911.7*	.99	164.47*	.98	63.89*	.97	1329.7*	.98
RMSE	***MSY***	2	-	-	-	-	-	-	-	-
	BMS*KOR	2	-	-	15.41*	.88	319.74*	.99	17.49*	.89
	BMS*KY	2	7.70*	.79	8.21*	.80	9.49*	.82	-	-
	KOR*KY	4	17.37*	.89	1.52*	.88	10.67*	.84	64.28*	.97
	BMS	1	11445*	.99	89.48*	.97	153.27*	.98	2209.1*	.99
	KOR	2	21817*	.99	1285.3*	.99	3376.8*	.99	829.92*	.99
	KY	2	4930.9*	.99	132.47*	.98	84.14*	.97	1668.7*	.99
ABSBIAS	***MSY***	2	-	-	-	-	-	-	-	-
	BMS*KOR	2	38.64*	.95	7.50*	.79	456.71*	.99	105.67*	.98
	BMS*KY	2	-	-	-	-	-	-	86.38*	.97
	KOR*KY	4	14.39*	.87	23.71	.92	6.19*	.75	14.06*	.87

Not: BMS=Boyut başına düşen madde sayısı, KOR=Boyutlar arası korelasyon KY=Kalibrasyon yöntemi, MSY=Madde seçme yöntemi. *=0.05 düzeyinde anlamlı,

Basit yapıdaki test desenindeki ANOVA sonuçlarının aksine, yüksek karmaşık yapıdaki test desenindeki ANOVA sonuçlarında, düşük karmaşık yapıdaki test deseni sonuçları ile paralellik göstermiş; boyut başına düşen madde sayısı ve korelasyon etkileşimine ait değerlerin hem alt boyut yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde hem de toplam yetenek puanlarına ait mutlak yanlılık ve hata miktarları üzerinde ana etkisinin anlamlı olduğu görülmüştür.

Ayrıca boyutlar arası korelasyon-kalibrasyon yöntemi ortak etkileşimlerinin hata ve mutlak yanlılık miktarları arasında da anlamlı bir ilişki görülmüştür ($p < .05$).

Burada göze çarpan bir nokta ise boyut başına düşen madde sayısı-kalibrasyon yöntemi etkileşiminde hata miktarları üzerinde sadece alt boyut yetenek puanlarına ait hata miktarları üzerinde etkisinin anlamlı olduğu görülürken, toplam yetenek puanlarında ise sadece mutlak yanlılık miktarları üzerinde etkisinin anlamlı olduğu belirlenmiştir.

5. SONUÇ ve ÖNERİLER

Araştırmanın bu bölümünde analizlerden elde edilen sonuçlara ve bu sonuçlardan yola çıkarak geliştirilen önerilere yer verilmiştir.

5.1. Sonuçlar

Araştırmada çok boyutlu test yapılarında ÇBBBT uygulamaları sonucu elde edilen alt boyut ve toplam puanların raporlanmasında madde seçme yöntemlerinin performansları kalibrasyon yöntemine, maddeler arası korelasyon değerlerine ve boyut başına düşen madde sayısına göre karşılaştırılmıştır. ÇBBBT uygulamasına geçilmeden önce farklı test desenlerine göre oluşturulmuş madde bankalarında kalibrasyon yöntemlerinin alt boyut ve toplam yetenek puanlarına etkisi incelenmiştir.

Çok Boyutlu Test Yapılarının Tek ve Çok Boyutlu Kalibrasyona Etkisi

1. Araştırmada ilk olarak her üç test deseni için (BY, DKY, YKY) kalibrasyon yöntemlerinin performansları incelenmiştir. Hem alt boyut hem de toplam yetenek puanlarının yanlılık, hata ve korelasyon değerleri karşılaştırılmıştır. Araştırmanın sonuçlarına göre tüm yetenek puanları kestiriminde en yüksek RMSE ve yanlılık değerleri MHRM yönteminde gözlenmiştir. Gerçek puanlar ile kalibre edilmiş puanlar arasındaki korelasyon değerlerine bakıldığında yine var olan tüm koşullarda MHRM kalibrasyonu diğer iki kalibrasyon yönteminden daha düşük değerler ürettiği belirlenmiştir.
2. Alt boyut yetenek puanlarının kestiriminde boyut başına düşen madde sayısı arttıkça RMSE ve yanlılık değerlerinde azalma, korelasyon değerlerinde ise bir artma olduğu belirlenmiştir.
3. BY test deseninde testin çok boyutluluğu azaldıkça alt boyut yetenek puanlarının kestirimlerine ait RMSE değerleri azalmakta ancak yanlılık bazı durumlarda artma bazı durumlarda ise azalma göstermektedir. Ancak DKY ve YKY test desenlerinde testin çok boyutluluğu azaldıkça alt boyut yetenek puanlarının kestirimlerine ait RMSE ve yanlılık değerlerinde azalma görüldüğü belirlenmiştir.
4. Maddeler arası korelasyonun azalmasına yani çok boyutluluğun artmasına bağlı olarak gerçek ve kalibre edilmiş alt boyut ve toplam puanlar arasındaki

korelasyonların da ters orantılı olarak değiştiği belirlenmiştir. Çok boyutluluğun ihlal edildiği durumlarda gerçek ve kalibre edilmiş alt boyut ve toplam puanlara ait korelasyon değerleri artmaktadır. Bu durum her üç test deseninde, madde sayısının arttığı koşullarda da geçerlidir.

5. Basit yapılı bir testte toplam puanlara ait en düşük RMSE değeri boyutlar arası korelasyonun en düşük olduğu ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM ve ATB kalibrasyonları için eşit şekilde kestirilmiştir Alt boyut puanlarına ait en düşük RMSE değeri ise boyutlar arası korelasyonun en yüksek olduğu ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM kalibrasyonunda kestirilmiştir.
6. DKY yapılı bir testte toplam puanlara ait en düşük RMSE değeri boyutlar arası korelasyonun en düşük olduğu (çok boyutluluğun en yüksek olduğu) ve boyut başına düşen madde sayısının 10 olduğu durumda BAEM kalibrasyonunda kestirilmiştir. Alt boyut puanlarına ait en düşük RMSE değeri boyutlar arası korelasyonun en yüksek olduğu ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM kalibrasyonunda kestirilmiştir.
7. YKY yapılı bir testte toplam puanlara ait en düşük RMSE değeri boyutlar arası korelasyonun en yüksek olduğu (çok boyutluluğun en düşük olduğu) ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM kalibrasyonunda kestirilmiştir. Alt boyut puanlarına ait en düşük RMSE değeri boyutlar arası korelasyonun en yüksek olduğu ve boyut başına düşen madde sayısının 15 olduğu durumda BAEM kalibrasyonunda kestirilmiştir.
8. Alt boyut yetenek puanlarının aksine, toplam puanlar model karmaşıklaştıkça daha düşük RMSE değerlerine sahip olduğu belirlenmiştir. Ancak model basit yapıdan karmaşık yapıya gittikçe her üç kalibrasyon yöntemi için alt boyut yetenek puanlarına ait RMSE değerlerinin ise arttığı görülmüştür. Alt boyut puanları raporlanırken artan hata miktarının dikkate alınması gerekmektedir.

Çok Boyutlu Test Yapılarının Madde Seçme Yöntemlerine Etkisi

Araştırmanın ikinci kısmında kalibre edilmiş madde bankalarındaki madde ve yetenek parametreleri kullanılarak ÇBBBT uygulamaları yapılmış, belirlenen koşullara göre üç farklı madde seçme yöntemlerinin alt boyut ve toplam yetenek puanlarını rapor etmedeki performansları yanlılık, hata ve gerçek puanlar ile ÇBBBT sonucu kestirilen puanlar arasındaki korelasyon değerleri kullanılarak karşılaştırılmıştır.

Kalibrasyon Yönteminin Etkisi

Boyutlar arasındaki korelasyon değerleri ve boyut başına düşen madde sayısı her koşulda sabit tutulduğu durumda sadece kalibrasyon yöntemlerinin madde seçme yöntemlerine etkilerine ilişkin sonuçlar her üç test deseni için özetlenmiştir.

1. BY test deseninde hem alt boyut hem de toplam puanlara ait bulgular incelendiğinde maddelerin tek boyutlu ATB hem de çok boyutlu BAEM yöntemi ile kalibre edildiği durumlarda V1 madde seçme yönteminin daha tutarlı sonuçlar verdiği görülmektedir. MHRM kalibrasyonunu dikkate alındığında ise alt boyut puanlarında KL, toplam puanlarda ise V1 yönteminin en iyi performansı gösterdiği belirlenmiştir. Elde edilen sonuçlar değerlendirildiğinde, alt boyut ve toplam yetenek puanlarını kestirmede kullanılan madde seçme yöntemlerinin kalibrasyon yöntemlerine göre farklılaştığı belirlenmiştir.
2. DKY test deseninde hem alt boyut hem de toplam puanlara ait bulgular incelendiğinde maddelerin tek boyutlu ATB hem de çok boyutlu BAEM yöntemi ile kalibre edildiği durumlarda V1 madde seçme yönteminin daha tutarlı sonuçlar verdiği görülmektedir. Elde edilen sonuçlar değerlendirildiğinde, alt boyut ve toplam yetenek puanlarını kestirmede kullanılan madde seçme yöntemlerinin kalibrasyon yöntemlerine göre farklılaşmadığı görülmektedir.
3. YKY test deseninde hem alt boyut hem de toplam puanlara ait bulgular incelendiğinde maddelerin tek boyutlu ATB yöntemi ile kalibre edildiği durumlarda V1 madde seçme yönteminin; çok boyutlu BAEM yöntemi ile kalibre edildiği durumlarda ise KL madde seçme yönteminin daha tutarlı sonuçlar verdiği görülmektedir. Elde edilen sonuçlar değerlendirildiğinde, alt

boyut ve toplam yetenek puanlarını kestirmede kullanılan madde seçme yöntemlerinin kalibrasyon yöntemlerine göre farklılaştığı görülmektedir.

Boyutlar Arası Korelasyon ve Boyut Başına Düşen Madde Sayısının Etkisi

Boyutlar arasındaki korelasyon değerleri ve boyut başına düşen madde sayısının değişen koşullarının da dikkate alındığı durumda tüm koşulların madde seçme yöntemlerine etkilerine ilişkin sonuçlar her üç test deseni için alt boyut ve toplam yetenek puanları için ayrı ayrı özetlenmiştir.

Tek Boyutlu Kalibrasyon (ATB)

1. BY deseninde alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.
2. BY deseninde toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.
3. DKY deseninde alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda Vol; boyutlar

arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise Vol madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.

4. DKY deseninde toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise Vol madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.
5. YKY deseninde alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.
6. YKY deseninde toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise V1 madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.

Çok Boyutlu Kalibrasyon (BAEM)

1. BY deseninde alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.
2. BY deseninde toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.
3. DKY deseninde alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda V1; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise KL madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.
4. DKY deseninde toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının

n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda Vol; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise Vol madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.

5. YKY deseninde alt boyut yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise V1 madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.
6. YKY deseninde toplam yetenek puanlarının belirlenen koşullara ait bulguları değerlendirildiğinde, boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=10 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=10 olduğu durumda V1; boyutlar arası korelasyonun düşük (ρ=0.2) ve boyut başına düşen madde sayısının n=15 olduğu durumda KL; boyutlar arası korelasyonun yüksek (ρ=0.8) ve boyut başına düşen madde sayısının n=15 olduğu durumda ise V1 madde seçme yöntemlerinin en iyi performansı sergilediği görülmüştür.

Genel Değerlendirme

Araştırmada üç farklı madde seçme yönteminin performansı karşılaştırılmıştır. Sonuçlar incelendiğinde önce V1 daha sonra da KL yönteminin daha iyi performans gösterdiği belirlenmiştir. Yao (2012) her bir boyuttan belirli sayıda madde seçilmesi durumunda KL ve V1 madde seçme yöntemlerinin alt boyut ve toplam puanları kestirmede daha iyi sonuçlar vereceğini belirtmiştir. Bu bakımdan sonuçlar literatürdeki araştırmalarla örtüşmektedir. Ayrıca V1 yönteminin daha iyi performans göstermesindeki en büyük neden ÇBBBT işlemlerinde ÇBMTK

uygulamaları sonucu elde edilen boyutlara arası teorik ağırlıklandırmalar yerine Çizelge 4.1'deki kestirilen ağırlıkların kullanılmasıdır.

Toplam puanların kestirilmesinde BY test deseninde V1 madde seçme yöntemi daha ağırlıklı olarak seçilmekte iken, test yapısının karmaşıklaştığı durumda KL ve Vol yöntemleri de iyi performans göstermiştir. Özellikle Vol yöntemi her bir boyuttan elde edilen bilgiyi eşit dağıtmaya çalışmaktadır (Yao, 2012). Ek 5'teki yetenek alanlarına göre *a*-parametresinden gelen bilgi oranlarını gösteren çizelgeler incelendiğinde test karmaşık yapıya doğru gittikçe bilginin her boyutta birbirine yaklaştığı görülmüştür. Bu bakımdan test karmaşıklaştıkça Vol madde seçme yönteminin belirli koşullarda daha iyi performans vermesi çıkması beklenen bir durumdur. KL diğer iki yöntemden farklıdır ve MDISC değerine göre madde seçmektedir. KL olabilirlik fonksiyonuna göre maddeleri seçtiğinden dolayı, olabilirlik fonksiyonları birbirinden uzak olan yetenek dağılımlarında çok az sayıda iyi maddeleri seçme eğilimindedir. Araştırmada sabit sayıda soru sorulduğundan ve boyutlardaki sorular önceden belirlendiğinden dolayı KL madde seçme yönteminin avantajlarının tam olarak yansıtılmadığı düşünülmektedir.

Genel çerçevede bakıldığında madde sayısı arttığında korelasyonların arttığı ve mutlak yanlılık değerlerinin ise azaldığı belirlenmiştir. Toplam yetenek puanlarına ait mutlak yanlılık değeri ise tüm koşullarda alt boyut yetenek puanları mutlak yanlılık değerlerinden düşük olarak kestirilmiştir. Elde edilen bu bulgular literatürdeki çalışmaları destekler niteliktedir (Lee, 2014; Su, 2016; Yao, 2014; Yao, Pommerich ve Segall, 2014).

Tüm test desenlerinde, madde seçme yöntemi ve kalibrasyon yöntemi fark etmeden, boyutlar arası korelasyon değeri arttığında alt boyut ve toplam yetenek puanlarına ait mutlak yanlılık değerlerinde farklılaşmalar görülmüştür. Test katı çok boyutluluk özelliğinden tek boyutluluğa yaklaştığı durumlarda alt boyut yetenek puanlarının mutlak hatasının düşerken, toplam puanların mutlak yanlılıkları artış göstermektedir. Toplam puanlara ait yetenek puanlarının yanlılığın artarken alt boyutlara ait yetenek puanlarının azalması yetenek kestirim yöntemlerindeki farklılıklardan kaynaklandığı düşünülmektedir. ÇBMTK modelleri yetenek kestirimlerinde boyutlar arasındaki korelasyonları dikkate alırken, alt boyut puanlarını kestirmede korelasyonlar göz ardı edilmektedir. Elde edilen sonuçlar literatürdeki araştırmalarla benzer sonuçlar göstermiştir (Liu, 2015; Yao, 2012).

Araştırmada ayrıca tüm madde seçme yöntemlerinde testin karmaşıklığı arttığında yanlılık ve hata değerlerinin azaldığı görülmüştür. Testin karmaşıklığı arttıkça, boyutlar arası korelasyon değerinin 0.2 olduğu durumda alt boyut puanları ve toplam puanların raporlanmasındaki mutlak yanlılık değerleri arasındaki fark artarken, boyutlar arasındaki korelasyonun 0.8 olması durumunda ise neredeyse sabit kalmaktadır. Bu bakımdan testin hem karmaşıklığının artırılması hem de boyutlar arası korelasyonun artırılması alt boyut ve toplam puanları rapor etmede aynı oranda etki etmemektedir. Bu durumun telafi edici modellerin etkisinden olduğu düşünülmektedir. Elde edilen bulgular literatürdeki çalışmaları destekler niteliktedir (Su, 2016).

5.2. Öneriler

Bu bölümde ÇBBBT uygulamaları sonucu alt boyut ve toplam puanların rapor edilmesinde madde seçme yöntemlerinin kullanımına ilişkin öneriler sıralanmıştır.

5.2.1. Araştırma Sonuçlarına Dayalı Öneriler

1. Araştırmada ÇBBBT uygulamaları sonucu elde edilen alt boyut ve toplam yetenek puanlarını kestirmede çok boyutlu test deseninin etkisinin fazla olduğu belirlenmiştir. Bu bakımdan ÇBBBT uygulaması sonucu tanılayıcı puanlar rapor edilmek istendiğinde amaca uygun test yapısının belirlenmesi önerilmektedir.
2. Madde bankasının kalibrasyonlarında bazı durumlarda tek boyutlu kalibrasyonun çok boyutlu kalibrasyonla benzer sonuçlar verdiği görülmüştür. Belirlenen koşullar sağlandığında tek boyutlu kalibrasyonun kullanılması önerilmektedir. MHRM kalibrasyonu ise hiçbir koşulda önerilmemektedir.
3. Çok boyutlu test desenlerinin tümünde alt boyut puanları rapor edilmek istediğinde boyutlar arası korelasyonun etkisi dikkate alınmalıdır.
4. Toplam puanların raporlanılması istenildiğinde en düşük mutlak yanlılık ve hata değeri içeren yüksek karmaşık test deseninin kullanılması önerilmektedir.
5. Alt boyut puanlarının raporlanmasında en düşük mutşak yanlılık ve hata değeri içeren basit yapı test deseninin kullanılması önerilmektedir.

6. Alt boyut puanlarının raporlanmasında boyutlar arası korelayonun en yüksek olduğu koşul, toplam puanların raporlanmasında boyutlar arası korelasyonun en düşük olduğu koşulun kullanılması önerilmektedir.
7. Bazı durumlarda tek boyutlu kalibrasyon ile BAEM sonuçları hemen hemen aynı sonuçlar vermiştir. Bu koşullardaki testlerde işlem bakımından daha basit olan ATB kalibrasyonlarının kullanılması zaman etkililiği bakımından önerilmektedir.

5.2.2. Yapılacak Araştırma Önerileri

1. ÇBBBT yöntemlerinin etkililiklerin ortaya koyulmasında madde bankalarının önemi büyüktür. Bu bakımdan benzer koşullar daha büyük çok boyutlu madde bankalarında test edilebilir ve madde seçme yöntemlerinin performansı karşılaştırılabilir.
2. Araştırmada PISA 2012 Türkiye verisinden elde edilen parametreler kullanılmıştır. Araştırma gerçek ÇBBBT uygulamaları üzerinden yapılarak madde ve yetebek dağılımına ilişkin sonuçlar karşılaştırılabilir.
3. Araştırmada testin boyutluluğu üç ile sınırlı tutulmuştur. Özellikle MHRM yönteminin daha fazla boyutluluk olduğu durumlarda avantajlı olduğu belirtilmektedir. Bu bakımdan daha yüksek boyutlu testlerde kalibrasyon yöntemlerinin etkisi incelenebilir.
4. Alt boyut puanları hesaplama yöntemleri ÇBBBT uygulamalarında önem kazanmaktadır. Bu bakımdan farklı hesaplama yöntemlerinin karşılaştırılması ve tanılayıcı puanları rapor etmedeki etkililikleri benzer koşullarda incelenebilir.
5. Çok boyutlu test desenleri bazı durumlarda hiyerarşik özellik, bazı durumlarda ise ortak bir faktörün varlığını da ortaya koyabilir. Bu tür test desenlerinde madde seçme yöntemlerinin alt boyut ve yetenek puanlarına etkisi incelenebilir.

KAYNAKÇA

Ackerman, T. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education, 7*, 255-278. doi:10.1207/s15324818ame0704_1

Ackerman, T. (1987). *A comparison study of the unidimensional IRT estimation of compensatory and noncompensatory multidimensional item response data* (ACT Araştırma Raporu Serisi 87-12). Iowa City, IA: American College Testing.

Ackerman, T. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement, 29*, 67-91.doi:10.1111/j.1745-3984.1992.tb00368.x

Ackerman, T. A., Gierl, M. J., & Walker, C. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice, 22*(3), 37–53. doi:10.1111/j.1745-3992.2003.tb00136.x

Adams RJ, Wilson M, Wang W (1997) The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement,21*, 1–23. doi: 10.1177/0146621697211001

Bashkov, B. M. (2015). *Examining the performance of the Metropolis-Hastings Robbins-Monro algorithm in the estimation of multilevel multidimensional IRT models.* (Unpublished doctoral dissertation). James Madison University, Harrisonburg, VA, USA.

Beguin, A. A., & Glas, C. A. W. (2001). MCMC estimation and some model-fit analyses of multidimensional IRT models. *Psychometrika, 66,* 541–562. doi:10.1007/BF02296195

Bloxom, B., & Vale, C.D. (1987). *Multidimensional adaptive testing: An approximate procedure for updating*. Paper presented at the annual meeting of the psychometric society. Montreal, Canada.

Bulut, O. (2013). *Between-person and within-person subscore reliability: Comparison of unidimensional and multidimensional IRT models*. (Unpublished doctoral dissertation). University of Minnesota, USA.

Cai, L. (2008). *A Metropolis-Hastings Robbins-Monro algorithm for maximum likelihood nonlinear latent structure analysis with a comprehensive measurement model*. (Unpublished doctoral dissertation). Department of Psychology, University of North Carolina – Chapel Hill.

Cai, L. (2010a). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika, 75,* 33-57. doi:10.1007/s11336-009-9136-x

Cai, L. (2010b). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics, 35*, 307-335. doi: doi:10.3102/1076998609353115

Cai, L. (2013). *flexMIRT® version 3.0: A numerical engine for flexible multilevel multidimensional item analysis and test scoring.* [Computer software]. Chapel Hill, NC: Vector Psychometric Group.

Chang, H. H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement, 20,* 213-229. doi:10.1177/014662169602000303

Chuah SC, Drasgow F, Luecht R. (2006). How big is big enough? Sample size requirements for CAST item parameter estimation. *Applied Measurement in Education, 19,* 241-255. doi:10.1207/s15324818ame1903_5

de la Torre, J., & Patz, R. J. (2005). Making the most of what we have: A practical application of multidimensional item response theory in test scoring. *Journal of Educational and Behavioral Statistics, 30,* 295-311. doi:10.3102/10769986030003295

de la Torre, J., & Song, H. (2009). Simultaneously estimation of overall and domain abilities: A higher order IRT model approach. *Applied Psychological Measurement, 33,*620-639. doi:10.1177/0146621608326423

DeMars, C. (2010). *Item response theory. Understanding statistics measurement*. Oxford University Press, New York.

Diao, Q. (2009). *Comparison of ability estimation and item selection methods in multidimensional computerized adaptive testing*. (Unpublished doctoral dissertation). Michigan State University, East Lansing, MI, USA.

Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement, 7,* 189–199. doi:10.1177/014662168300700207

Fan, M., & Hsu, Y. (1996). *Multidimensional computer adaptive testing*. Paper presented at the annual meeting of the American Educational Testing Association, New York City, NY.

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

Folk, V.G., & Golub-Smith, M. (1996). *Calibration of on-line pretest data using BILOG.* Paper presented at the annual meeting of the National Council of Measurement in Education, New York.

Fraser, C. (1988). NOHARM: *A computer program for fitting both unidimensional and multidimensional normal ogive models of latent traits theory* [Computer software]. The University of New England, Center for Behavioral Studies, Armidale, Australia.

Fraser, C., & McDonald, R. P. (1988). COSAN: Covariance structure analysis. *Multivariate Behavioral Research, 23*, 263–265. doi:10.1207/s15327906mbr2302_8

Gierl, M. J., Leighton, J. P., & Tan, X. (2006). Evaluating DETECT classification accuracy and consistency when data display complex structure. *Journal of Educational Measurement, 43*(3), 265-289. doi:10.1111/j.1745-3984.2006.00016.x

Glas, C. A. W. (2000). Item calibration and parameter drift. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 183-199). Norwell, MA: Kluwer Academic.

Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education, 17,* 145-220. doi:10.1207/s15324818ame1702_3

Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and Behavioral Statistics, 33*, 204-229. doi:10.3102/1076998607302636

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

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Thousand Oaks: Sage Publications.

Han, K. T., & Paek, I. (2014). A review of commercial software packages for multidimensional IRT modeling. *Applied Psychological Measurement, 38*(6), 486-498. doi:10.1177/0146621614536770

Karasar, N. (2010). *Bilimsel araştırma yöntemi*. Ankara. Nobel Yayın Dağıtım.

Levine, M. V. Williams B. A. (1998). *Development and evaluation of online calibration procedures.* (Algorithm Design and Measurement Services, Inc. TCN # 96-216). Champaign, IL.

Liu, F. (2015). *Comparisons of subscoring methods in computerized adaptive testing: A simulation study.* (Unpublished doctoral dissertation). University of North Carolina Greensboro, North Carolina, USA.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass: Addison-Wesley Pub. Co.

Luecht, R. M. (1996). Multidimensional computerized adaptive testing in a certification or licensure context. *Applied Psychological Measurement, 20*, 389-404. doi:10.1177/014662169602000406

Luecht, R. M. (2004). *MIRTGEN 2.0 Manual*. Department of Educational Research Methodology, University of North Carolina at Greensboro, Greensboro, NC.

Luecht, R. M., & Miller, T. R. (1992). Unidimensional calibrations and interpretations of composite traits for multidimensional tests. *Applied Psychological Measurement, 16*(3), 279-293. doi:10.1177/014662169201600308

Luecht, R. M., Gierl, M. J., Tan, X., & Huff, K. (2006). *Scalability and the development of useful diagnostic scales*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.

Luo, X. (2013). *The optimal design of the dual-purpose test*. (Unpublished doctoral dissertation). University of North Carolina Greensboro, North Carolina, USA.

McDonald, R. P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement. 24,* 99-114. doi:10.1177/01466210022031552

McLeod, L.D., Swygert, K.A., & Thissen, D. (2001). Factor analysis for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test Scoring* (pp. 189-216). Hillsdale, NJ: Lawrence Erlbaum Associates.

Miller, T. R. (1991) *Empirical estimation of standard errors of compensatory MIRT model parameters obtained from the NOHARM estimation program* (Research Report ONR 91-2). American College Testing Program, Iowa City, IA.

Mills, C. N., & Stocking, M. L. (1996). Practical issues in large-scale computerized adaptive testing. *Applied Measurement in Education, 9*, 287-304. doi:10.1207/s15324818ame0904_1

Mulder, J., & van der Linden, W. J. (2010). Multidimensional adaptive testing with Kullback-Leibler information item selection. In W.J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp.77-101). New York: Springer.

Mulder, J., & van der Linden, W.J. (2009). Multidimensional adaptive testing with optimal design criteria for item selection. *Psychometrika, 74*(2), 273–296. doi:10.1007/s11336-008-9097-5

Nandakumar, R. (1994). Assessing dimensionality of a set of item responses: Comparison of different approaches. *Journal of Educational Measurement, 31*(1), 17–35. doi:10.1111/j.1745-3984.1994.tb00432.x

Owen, R. J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association, 70*, 351-356. doi:10.1080/01621459.1975.10479871

Özdemir, B. (2015*). Madde düzeyinde boyutluluk modellerinin bilgisayar ortaminda bireyselleştirilmiş test yöntemleri üzerindeki etkisinin incelenmesi.* Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Reckase, M. & Hirsh, T. (1991). *Interpretation of number correct scores when the true number of dimensions assessed by a test is greater than two*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4,* 207–230. doi:10.2307/1164671

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9(*4), 401–412. doi:10.1177/014662168500900409

Reckase, M. D. (1997). A linear logistic multidimensional model for dichotomous item response data. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 271-286). New York, NY: Springer.

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

Reckase, M. D., & McKinley, R. L. (1982). *The feasibility of a multidimensional latent trait model*. Paper presented at the annual meeting of the American Psychological Association, Washington.

Reckase, M. D., Ackerman T. A., Carlson J. E. (1988). Building a unidimensional test using multidimensional items. *Journal of Educational Measurement, 25,* 193–204. doi:10.1111/j.1745-3984.1988.tb00302.x

Samejima F (2000). Logistic positive exponent family of models: Virtue of asymmetric item characteristic curves. *Psychometrika, 65*(3), 319-335. doi:10.1007/BF02296149

Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika, 70*, 533-555. doi:10.1007/s11336-003-1141-x

Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika, 61*, 331-354. doi:10.1007/BF02294343

Segall, D.O. (2010). Principles of multidimensional adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 57–75). New York, NY: Springer.

Seo, D. G., & Weiss, D. J. (2015). Best design for multidimensional computerized adaptive testing with the bifactor model. *Educational and Psychological Measurement, 75,* 954-978. doi:10.1177/0013164415575147

Sinharay, S., & Haberman, S. J. (2008). *Reporting subscores: A survey* (ETS Research Memorandum No. RR-08-18). Princeton, NJ: Educational Testing Service.

Sinharay, S., Puhan, G., & Haberman, S. J. (2010). Reporting diagnostic subscores in educational testing: Temptations, pitfalls, and some solutions. *Multivariate Behavioral Research, 45*, 553–573. doi:10.1080/00273171.2010.483382

Song, T. (2010). *The effect of fitting a tridimensional IRT model to multidimensional data in content-balanced computerized adaptive testing.* (Unpublished doctoral dissertation). Michigan State University, East Lansing, MI, USA.

Stocking, M. (1988). *Scale drift online calibration*. (ETS Research Report 88-28-ONR). Princeton, NJ: Educational Testing Service.

Stone, C. A., Ye, F., Zhu, X., & Lane, S. (2010). Providing subscale scores for diagnostic information: a case study when the test is essentially unidimensional. *Applied Measurement in Education, 23,* 63-86. doi:10.1080/08957340903423651

Su, Y. (2016). A comparison of constrained item selection methods in multidimensional computerized adaptive testing. *Applied Psychological Measurement, 40(*5), 346-360. doi:10.1177/0146621616639305

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

Tam, S. S. (1992). *A comparison of methods for adaptive estimation of a multidimensional trait*. (Unpublished doctoral dissertation). Columbia University.

Thissen, D. & Mislevy, R.J. (2000). Testing algorithms. In H. Wainer, N. Dorans, D. Eignor, R. Flaugher, B. Green, R. Mislevy, L. Steinberg & D. Thissen (Eds.), *Computerized adaptive testing: A primer* (Second Edition). Hillsdale, NJ: Lawrence Erlbaum Associates, 101-133.

Thissen, D., & Edwards, M.C. (2005). *Diagnostic scores augmented using multidimensional item response theory.* Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada

van der Linden, W. J. (2005). *Linear models for optimal test design*. New York: Springer-Verlag.

van der Linden, W. J., & Pashley, P. J. (2010). Item selection and ability estimation in adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp.3-30). New York: Springer.

van der Linden, W.J. (1999). Multidimensional adaptive testing with a minimum error variance criterion. *Journal of Educational and Behavioral Statistics, 24,* 398–412. doi:10.3102/10769986024004398

Veerkamp, W.J.J. & Berger, M.P.F. (1997). Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics, 22*, 203-226. doi:10.3102/10769986022002203

Veldkamp, B.P., & van der Linden, W.J. (2002). Multidimensional adaptive testing with constraints on test content. *Psychometrika, 67*(4), 575–588. doi:10.1007/BF02295132

Veldkamp, B.P., & van der Linden, W.J. (2010). Designing item pools for computerized adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 231-245). New York: Springer.

Wainer, H., & Mislevy, R. J. (1990). Item response theory, item calibration and proficiency estimation. In H. Wainer (Ed.), *Computerized adaptive testing: A primer* (pp. 65-102). Hillsdale, NJ: Lawrence Erlbaum.

Wainer, H., Vevea, J. L., Camacho, F., Reeve, B. B., Rosa, K., & Nelson, L. (2001). Augmented scores: “Borrowing Strength” to compute scores based on small numbers of items. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 343–387). Mahwah, NJ: Erlbaum.

Walker, C. M., & Beretvas, S. N. (2003). Comparing multidimensional and unidimensional proficiency classifications: Multidimensional IRT as a diagnostic aid. *Journal of Educational Measurement, 40*, 255–275. doi:10.1111/j.1745-3984.2003.tb01107.x

Wang, C., & Chang, H. (2011). Item selection in multidimensional computerized adaptive tests: Gaining information from different angles. *Psychometrika, 76*(3), 363-384. doi:10.1007/s11336-011-9215-7

Wang, C., Chang, H., & Boughton, K. A. (2011). Kullback-Leibler information and its applications in multi-dimensional adaptive testing. *Psychometrika, 76* (1), 13-39. doi:10.1007/s11336-010-9186-0

Wang, C., Chang, H.-H., & Boughton, K. (2013). Deriving stopping rules for multidimensional computerized adaptive testing. *Applied Psychological Measurement, 37*, 99-122. doi:10.1177/0146621612463422

Wang, W. C., & Chen, P. H. (2004). Implementation and measurement efficiency of multidimensional computerized adaptive testing. *Applied Psychological Measurement, 28,* 295-316. doi:10.1177/0146621604265938

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

Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized testing to educational problems. *Journal of Educational Measurement, 21*(4), 361–375. doi:10.1111/j.1745-3984.1984.tb01040.x

Wilson, D. T., Wood, R., & Gibbons, R. (1998). *TESTFACT: Test scoring, item statistics, and item factor analysis* [Software]. Chicago: Scientific Software International.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1998). *ACER ConQuest: Generalized item response modeling software version 2.0* [Software]. Melbourne, Australia: Australian Council for Educational Research.

Yao, L. (2003). *BMIRT: Bayesian multivariate item response theory* [Computer software]. Monterey, CA: CTB/McGraw-Hill.

Yao, L. (2010). Reporting valid and reliability overall score and domain scores. *Journal of Educational Measurement, 47*, 339-360. doi:10.1111/j.1745-3984.2010.00117.x

Yao, L. (2011). *simuMCAT: simulation of multidimensional computer adaptive testing* [Computer software]. Monterey: Defense Manpower Data Center.

Yao, L. (2012). Multidimensional CAT item selection methods for domain scores and composite scores: Theory and applications. *Psychometrika, 77,* 495-523. doi:10.1007/s11336-012-9265-5

Yao, L. (2013). Comparing the performance of five multidimensional CAT selection procedures with different stopping rules. *Applied Psychological Measurement, 37,* 3-23. doi:10.1177/0146621612455687

Yao, L. (2014). Multidimensional CAT item selection procedures with item exposure control and content constraints. *Journal of Educational Measurement, 51,* 18-38. doi:10.1111/jedm.12032

Yao, L., & Boughton, K. A. (2007). A multidimensional item response modeling approach for improving subscale proficiency estimation and classification. *Applied Psychological Measurement, 31*, 83-105. doi:10.1177/0146621606291559

Yao, L., & Boughton, K. A. (2009). Multidimensional linking for tests containing polytomous items. *Journal of Educational Measurement, 46*, 177–197. doi:10.1111/j.1745-3984.2009.00076.x

Yao, L., & Schwarz, R. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed format tests. *Applied Psychological Measurement, 30,* 469–492. doi:10.1177/0146621605284537

Yao, L., Pommerich, M., & Segall, D. O. (2014). Using Multidimensional CAT to Administer a Short, Yet Precise, Screening Test. *Applied Psychological Measurement, 38*(8), 614-631. doi:10.1177/0146621614541514

Yen, W. M. (1993). Scaling performance assessments: strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 187-213. doi:10.1111/j.1745-3984.1993.tb00423.x

Zhang, J. (1996). *Some fundamental issues in item response theory with applications.* (Unpublished doctoral dissertation), University of Illinois, Urbana-Champaign.

Zhang, J. (2012). Calibration of response data using MIRT models with simple and mixed structures. *Applied Psychological Measurement, 36*, 375–398. doi:10.1177/0146621612445904

EKLER DİZİNİ

EK 1. ETİK KURUL ONAY BİLDİRİMİ

Form: 40

Tez Çalışması Etik Komisyon İzin Muafiyeti Formu

21 / 12 / 2016

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Anabilim Dalı Başkanlığı'na

Tez Başlığı / Konusu:	Çok Boyutlu Test Deseninin ve Kalibrasyon Yöntemlerinin Çok Boyutlu Bireyselleştirilmiş Bilgisayar Uygulamalarına Etkisi

Yukarıda başlığı/konusu gösterilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır,
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne müdahale içermemektedir.
4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.

Hacettepe Üniversitesi Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez 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.

Eren Halil ÖZBERK
(Öğrencinin Adı Soyadı, İmzası)

Öğrenci Bilgileri

Adı Soyadı	**Eren Halil ÖZBERK**
Öğrenci No	N11247151
Anabilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	☐ Yüksek Lisans ☒ Doktora ☐ Bütünleşik Dr.

Danışman Görüşü ve Onayı

Bu çalışmada hazır veriye dayalı simülasyon çalışması yapıldığından, tez sürecinde veri toplama ve ölçek geliştirme işlemi yapılmadığından etik kurul izninden muaf tutulması uygundur.

Prof. Dr. Selahattin GELBAL
(İmza)
(Danışmanın Unvanı, Adı ve Soyadı)

EK 2. ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS/DOKTORA TEZ ÇALIŞMASI ORİJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLER ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANA BİLİM / BİLİM DALI BAŞKANLIĞI'NA

Tarih: 21/12/2016

Tez Başlığı: Çok Boyutlu Test Deseninin ve Kalibrasyon Yöntemlerinin Çok Boyutlu Bireyselleştirilmiş Bilgisayar Uygulamalarına Etkisi

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 Endeksi	Gönderim Numarası
20/12/2016	104	194246	19/12 /2016	%2	755191134

Uygulanan filtreler:
1- Kaynakça hariç
2- Alıntılar dâhil
3- 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı 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 ederim.

Gereğini saygılarımla arz ederim.

21.12.2016
Tarih ve İmza

Adı Soyadı: Eren Halil ÖZBERK
Öğrenci No: N11247151
Anabilim Dalı: Eğitim Bilimleri
Programı: Eğitimde Ölçme ve Değerlendirme
Statüsü: ☐ Y.Lisans ☒ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.
Prof. Dr. Selahattin GELBAL
(Unvan, Ad Soyad, İmza)

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
THESIS/DISSERTATION ORIGINALITY REPORT

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
TO THE DEPARTMENT OF EDUCATIONAL SCIENCES

Date: 21/12/2016

Thesis Title: The Effect of Multidimensional Test Design Structure and Calibration Strategies in Multidimensional Computer Adaptive Testing

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 Defence	Similarity Index	Submission ID
20/12/2016	104	194246	19/12 /2016	%2	755191134

Filtering options applied:
1. Bibliography excluded
2. Quotes excluded
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.

21.12.2016
Date and Signature

Name Surname: Eren Halil ÖZBERK
Student No: N11247151
Department: Educational Science
Program: Educational Measurement and Evaluation
Status: ☐ Masters ☒ Ph.D. ☐ Integrated Ph.D.

ADVISOR APPROVAL

APPROVED.
Prof. Dr. Selahattin GELBAL
(Title, Name Surname, Signature)

EK 3. TEK BOYUTLU TEST DESENİ İÇİN ÜRETİLEN MADDE PARAMETRELERİ

Madde	*N=10*			*N=15*		
	a	*b*	*c*	*a*	*b*	*c*
Madde1	0.86	1.20	0.15	0.56	-1.48	0.15
Madde2	0.66	-0.98	0.15	1.01	0.91	0.15
Madde 3	0.72	0.99	0.15	0.81	0.74	0.15
Madde 4	0.53	-0.49	0.15	1.14	-0.51	0.15
Madde 5	1.23	1.39	0.15	0.70	-0.24	0.15
Madde 6	1.03	-1.61	0.15	0.96	0.54	0.15
Madde 7	1.23	0.18	0.15	0.98	0.00	0.15
Madde 8	0.65	0.18	0.15	0.86	1.28	0.15
Madde 9	1.06	0.00	0.15	0.78	1.06	0.15
Madde 10	0.77	1.15	0.15	1.27	0.64	0.15
Madde 11	0.92	0.57	0.15	1.26	-0.08	0.15
Madde 12	1.29	0.67	0.15	0.81	-1.76	0.15
Madde 13	1.10	-0.76	0.15	0.50	-1.09	0.15
Madde 14	1.35	0.79	0.15	0.50	0.39	0.15
Madde 15	0.71	1.34	0.15	1.30	-0.44	0.15
Madde 16	0.79	1.49	0.15	0.86	-1.31	0.15
Madde 17	1.23	1.75	0.15	1.18	0.80	0.15
Madde 18	0.64	-0.17	0.15	0.95	-0.32	0.15
Madde 19	0.90	-1.66	0.15	0.97	1.28	0.15
Madde 20	0.64	-1.44	0.15	0.60	1.53	0.15
Madde 21	1.17	-1.25	0.15	1.11	-1.29	0.15
Madde 22	1.07	0.44	0.15	0.96	0.70	0.15
Madde 23	0.76	-0.04	0.15	0.73	0.36	0.15
Madde 24	0.77	1.87	0.15	1.09	1.04	0.15
Madde 25	1.31	-0.46	0.15	0.67	0.24	0.15
Madde 26	0.84	1.37	0.15	0.86	-0.14	0.15
Madde 27	0.64	0.93	0.15	1.17	1.36	0.15
Madde 28	0.70	-0.61	0.15	1.18	-1.46	0.15
Madde 29	1.15	1.00	0.15	1.21	0.53	0.15
Madde 30	0.62	0.32	0.15	0.89	0.12	0.15
Madde 31				0.91	-0.48	0.15
Madde 32				0.90	-0.57	0.15
Madde 33				0.65	-0.75	0.15
Madde 34				0.79	1.57	0.15
Madde 35				1.29	-0.20	0.15
Madde 36				1.14	0.04	0.15
Madde 37				1.04	0.37	0.15
Madde 38				1.31	0.98	0.15
Madde 39				1.25	1.08	0.15
Madde 40				0.83	-1.45	0.15
Madde 41				0.58	0.40	0.15
Madde 42				1.15	-0.64	0.15
Madde 43				1.10	0.19	0.15
Madde 44				1.23	0.08	0.15
Madde 45				0.83	-1.18	0.15

EK 4. ÇOK BOYUTLU TEST DESENLERİ İÇİN ÜRETİLEN MADDE PARAMETRELERİ

Basit Yapılı (BY) Test Deseni için Üretilen Madde Parametreleri

Madde	N=10						N=15					
	a_1	a_2	a_3	d	c	ÇBAE	a_1	a_2	a_3	d	c	ÇBAE
Madde1	**0.86**	0.00	0.00	1.20	0.15	0.86	**0.56**	0.00	0.00	-1.48	0.15	0.56
Madde2	**0.66**	0.00	0.00	-0.98	0.15	0.66	**1.01**	0.00	0.00	0.91	0.15	1.01
Madde 3	**0.72**	0.00	0.00	0.99	0.15	0.72	**0.81**	0.00	0.00	0.74	0.15	0.81
Madde 4	**0.53**	0.00	0.00	-0.49	0.15	0.53	**1.14**	0.00	0.00	-0.51	0.15	1.14
Madde 5	**1.23**	0.00	0.00	1.39	0.15	1.23	**0.70**	0.00	0.00	-0.24	0.15	0.70
Madde 6	**1.03**	0.00	0.00	-1.61	0.15	1.03	**0.96**	0.00	0.00	0.54	0.15	0.96
Madde 7	**1.23**	0.00	0.00	0.18	0.15	1.23	**0.98**	0.00	0.00	0.00	0.15	0.98
Madde 8	**0.65**	0.00	0.00	0.18	0.15	0.65	**0.86**	0.00	0.00	1.28	0.15	0.86
Madde 9	**1.06**	0.00	0.00	0.00	0.15	1.06	**0.78**	0.00	0.00	1.06	0.15	0.78
Madde 10	**0.77**	0.00	0.00	1.15	0.15	0.77	**1.27**	0.00	0.00	0.64	0.15	1.27
Madde 11	0.00	**0.92**	0.00	0.57	0.15	0.92	**1.26**	0.00	0.00	-0.08	0.15	1.26
Madde 12	0.00	**1.29**	0.00	0.67	0.15	1.29	**0.81**	0.00	0.00	-1.76	0.15	0.81
Madde 13	0.00	**1.10**	0.00	-0.76	0.15	1.10	**0.50**	0.00	0.00	-1.09	0.15	0.50
Madde 14	0.00	**1.35**	0.00	0.79	0.15	1.35	**0.50**	0.00	0.00	0.39	0.15	0.50
Madde 15	0.00	**0.71**	0.00	1.34	0.15	0.71	**1.30**	0.00	0.00	-0.44	0.15	1.30
Madde 16	0.00	**0.79**	0.00	1.49	0.15	0.79	0.00	**0.86**	0.00	-1.31	0.15	0.86
Madde 17	0.00	**1.23**	0.00	1.75	0.15	1.23	0.00	**1.18**	0.00	0.80	0.15	1.18
Madde 18	0.00	**0.64**	0.00	-0.17	0.15	0.64	0.00	**0.95**	0.00	-0.32	0.15	0.95
Madde 19	0.00	**0.90**	0.00	-1.66	0.15	0.90	0.00	**0.97**	0.00	1.28	0.15	0.97
Madde 20	0.00	**0.64**	0.00	-1.44	0.15	0.64	0.00	**0.60**	0.00	1.53	0.15	0.60
Madde 21	0.00	0.00	**1.17**	-1.25	0.15	1.17	0.00	**1.11**	0.00	-1.29	0.15	1.11
Madde 22	0.00	0.00	**1.07**	0.44	0.15	1.07	0.00	**0.96**	0.00	0.70	0.15	0.96
Madde 23	0.00	0.00	**0.76**	-0.04	0.15	0.76	0.00	**0.73**	0.00	0.36	0.15	0.73
Madde 24	0.00	0.00	**0.77**	1.87	0.15	0.77	0.00	**1.09**	0.00	1.04	0.15	1.09
Madde 25	0.00	0.00	**1.31**	-0.46	0.15	1.31	0.00	**0.67**	0.00	0.24	0.15	0.67
Madde 26	0.00	0.00	**0.84**	1.37	0.15	0.84	0.00	**0.86**	0.00	-0.14	0.15	0.86
Madde 27	0.00	0.00	**0.64**	0.93	0.15	0.64	0.00	**1.17**	0.00	1.36	0.15	1.17
Madde 28	0.00	0.00	**0.70**	-0.61	0.15	0.70	0.00	**1.18**	0.00	-1.46	0.15	1.18
Madde 29	0.00	0.00	**1.15**	1.00	0.15	1.15	0.00	**1.21**	0.00	0.53	0.15	1.21
Madde 30	0.00	0.00	**0.62**	0.32	0.15	0.62	0.00	**0.89**	0.00	0.12	0.15	0.89
Madde 31							0.00	0.00	**0.91**	-0.48	0.15	0.91
Madde 32							0.00	0.00	**0.90**	-0.57	0.15	0.90
Madde 33							0.00	0.00	**0.65**	-0.75	0.15	0.65
Madde 34							0.00	0.00	**0.79**	1.57	0.15	0.79
Madde 35							0.00	0.00	**1.29**	-0.20	0.15	1.29
Madde 36							0.00	0.00	**1.14**	0.04	0.15	1.14
Madde 37							0.00	0.00	**1.04**	0.37	0.15	1.04
Madde 38							0.00	0.00	**1.31**	0.98	0.15	1.31
Madde 39							0.00	0.00	**1.25**	1.08	0.15	1.25
Madde 40							0.00	0.00	**0.83**	-1.45	0.15	0.83
Madde 41							0.00	0.00	**0.58**	0.40	0.15	0.58
Madde 42							0.00	0.00	**1.15**	-0.64	0.15	1.15
Madde 43							0.00	0.00	**1.10**	0.19	0.15	1.10
Madde 44							0.00	0.00	**1.23**	0.08	0.15	1.23
Madde 45							0.00	0.00	**0.83**	-1.18	0.15	0.83

Düşük Karmaşık Yapı (DKY) Test Deseni için Üretilen Madde Parametreleri

Madde	*N=10*						*N=15*					
	a_1	a_2	a_3	*d*	*c*	*ÇBAE*	a_1	a_2	a_3	*d*	*c*	*ÇBAE*
Madde1	**0.72**	0.39	0.26	1.20	0.15	0.86	**0.47**	0.26	0.17	-1.48	0.15	0.56
Madde2	**0.55**	0.30	0.20	-0.98	0.15	0.66	**0.85**	0.46	0.30	0.91	0.15	1.01
Madde 3	**0.60**	0.33	0.22	0.99	0.15	0.72	**0.68**	0.37	0.24	0.74	0.15	0.81
Madde 4	**0.45**	0.24	0.16	-0.49	0.15	0.53	**0.95**	0.52	0.34	-0.51	0.15	1.14
Madde 5	**1.03**	0.56	0.37	1.39	0.15	1.23	**0.58**	0.32	0.21	-0.24	0.15	0.70
Madde 6	**0.86**	0.47	0.31	-1.61	0.15	1.03	**0.81**	0.44	0.29	0.54	0.15	0.96
Madde 7	**1.03**	0.56	0.37	0.18	0.15	1.23	**0.82**	0.45	0.29	0.00	0.15	0.98
Madde 8	**0.55**	0.30	0.20	0.18	0.15	0.65	**0.72**	0.39	0.26	1.28	0.15	0.86
Madde 9	**0.88**	0.48	0.32	0.00	0.15	1.06	**0.65**	0.36	0.23	1.06	0.15	0.78
Madde 10	**0.64**	0.35	0.23	1.15	0.15	0.77	**1.06**	0.58	0.38	0.64	0.15	1.27
Madde 11	0.42	**0.77**	0.28	0.57	0.15	0.92	**1.06**	0.58	0.38	-0.08	0.15	1.26
Madde 12	0.59	**1.08**	0.39	0.67	0.15	1.29	**0.68**	0.37	0.24	-1.76	0.15	0.81
Madde 13	0.50	**0.92**	0.33	-0.76	0.15	1.10	**0.42**	0.23	0.15	-1.09	0.15	0.50
Madde 14	0.62	**1.13**	0.40	0.79	0.15	1.35	**0.42**	0.23	0.15	0.39	0.15	0.50
Madde 15	0.32	**0.59**	0.21	1.34	0.15	0.71	**1.09**	0.60	0.39	-0.44	0.15	1.30
Madde 16	0.36	**0.66**	0.24	1.49	0.15	0.79	0.40	**0.72**	0.26	-1.31	0.15	0.86
Madde 17	0.56	**1.03**	0.37	1.75	0.15	1.23	0.54	**0.99**	0.36	0.80	0.15	1.18
Madde 18	0.30	**0.54**	0.19	-0.17	0.15	0.64	0.44	**0.80**	0.29	-0.32	0.15	0.95
Madde 19	0.41	**0.75**	0.27	-1.66	0.15	0.90	0.45	**0.82**	0.29	1.28	0.15	0.97
Madde 20	0.29	**0.53**	0.19	-1.44	0.15	0.64	0.27	**0.50**	0.18	1.53	0.15	0.60
Madde 21	0.53	0.35	**0.98**	-1.25	0.15	1.17	0.51	**0.93**	0.33	-1.29	0.15	1.11
Madde 22	0.49	0.32	**0.90**	0.44	0.15	1.07	0.44	**0.81**	0.29	0.70	0.15	0.96
Madde 23	0.35	0.23	**0.64**	-0.04	0.15	0.76	0.33	**0.61**	0.22	0.36	0.15	0.73
Madde 24	0.35	0.23	**0.65**	1.87	0.15	0.77	0.50	**0.91**	0.33	1.04	0.15	1.09
Madde 25	0.60	0.39	**1.09**	-0.46	0.15	1.31	0.31	**0.56**	0.20	0.24	0.15	0.67
Madde 26	0.39	0.25	**0.71**	1.37	0.15	0.84	0.39	**0.72**	0.26	-0.14	0.15	0.86
Madde 27	0.29	0.19	**0.54**	0.93	0.15	0.64	0.54	**0.98**	0.35	1.36	0.15	1.17
Madde 28	0.32	0.21	**0.58**	-0.61	0.15	0.70	0.54	**0.99**	0.36	-1.46	0.15	1.18
Madde 29	0.53	0.34	**0.96**	1.00	0.15	1.15	0.55	**1.01**	0.36	0.53	0.15	1.21
Madde 30	0.29	0.19	**0.52**	0.32	0.15	0.62	0.41	**0.75**	0.27	0.12	0.15	0.89
Madde 31							0.42	0.27	**0.76**	-0.48	0.15	0.91
Madde 32							0.41	0.27	**0.75**	-0.57	0.15	0.90
Madde 33							0.30	0.20	**0.55**	-0.75	0.15	0.65
Madde 34							0.36	0.24	**0.66**	1.57	0.15	0.79
Madde 35							0.59	0.39	**1.08**	-0.20	0.15	1.29
Madde 36							0.52	0.34	**0.96**	0.04	0.15	1.14
Madde 37							0.48	0.31	**0.87**	0.37	0.15	1.04
Madde 38							0.60	0.39	**1.09**	0.98	0.15	1.31
Madde 39							0.57	0.38	**1.05**	1.08	0.15	1.25
Madde 40							0.38	0.25	**0.69**	-1.45	0.15	0.83
Madde 41							0.26	0.17	**0.48**	0.40	0.15	0.58
Madde 42							0.53	0.34	**0.96**	-0.64	0.15	1.15
Madde 43							0.50	0.33	**0.92**	0.19	0.15	1.10
Madde 44							0.57	0.37	**1.03**	0.08	0.15	1.23
Madde 45							0.38	0.25	**0.70**	-1.18	0.15	0.83

Düşük Karmaşık Yapı (DKY) Test Deseni için Üretilen Madde Parametreleri

Madde	*N=10*						*N=15*					
	a_1	a_2	a_3	*d*	*c*	*ÇBAE*	a_1	a_2	a_3	*d*	*c*	*ÇBAE*
Madde1	**0.61**	0.51	0.33	1.20	0.15	0.86	**0.40**	0.33	0.22	-1.48	0.15	0.56
Madde2	**0.47**	0.39	0.26	-0.98	0.15	0.66	**0.72**	0.60	0.39	0.91	0.15	1.01
Madde 3	**0.51**	0.43	0.28	0.99	0.15	0.72	**0.57**	0.48	0.31	0.74	0.15	0.81
Madde 4	**0.38**	0.32	0.21	-0.49	0.15	0.53	**0.81**	0.67	0.44	-0.51	0.15	1.14
Madde 5	**0.87**	0.72	0.47	1.39	0.15	1.23	**0.49**	0.41	0.27	-0.24	0.15	0.70
Madde 6	**0.73**	0.61	0.40	-1.61	0.15	1.03	**0.68**	0.57	0.37	0.54	0.15	0.96
Madde 7	**0.87**	0.73	0.48	0.18	0.15	1.23	**0.69**	0.58	0.38	0.00	0.15	0.98
Madde 8	**0.46**	0.39	0.25	0.18	0.15	0.65	**0.61**	0.51	0.33	1.28	0.15	0.86
Madde 9	**0.75**	0.63	0.41	0.00	0.15	1.06	**0.55**	0.46	0.30	1.06	0.15	0.78
Madde 10	**0.54**	0.45	0.30	1.15	0.15	0.77	**0.90**	0.75	0.49	0.64	0.15	1.27
Madde 11	0.54	**0.65**	0.36	0.57	0.15	0.92	**0.89**	0.75	0.49	-0.08	0.15	1.26
Madde 12	0.76	**0.91**	0.50	0.67	0.15	1.29	**0.57**	0.48	0.31	-1.76	0.15	0.81
Madde 13	0.65	**0.78**	0.43	-0.76	0.15	1.10	**0.35**	0.30	0.19	-1.09	0.15	0.50
Madde 14	0.79	**0.95**	0.52	0.79	0.15	1.35	**0.35**	0.30	0.19	0.39	0.15	0.50
Madde 15	0.42	**0.50**	0.27	1.34	0.15	0.71	**0.92**	0.77	0.50	-0.44	0.15	1.30
Madde 16	0.47	**0.56**	0.31	1.49	0.15	0.79	0.51	**0.61**	0.33	-1.31	0.15	0.86
Madde 17	0.72	**0.87**	0.47	1.75	0.15	1.23	0.70	**0.84**	0.46	0.80	0.15	1.18
Madde 18	0.38	**0.46**	0.25	-0.17	0.15	0.64	0.56	**0.67**	0.37	-0.32	0.15	0.95
Madde 19	0.53	**0.64**	0.35	-1.66	0.15	0.90	0.58	**0.69**	0.38	1.28	0.15	0.97
Madde 20	0.38	**0.45**	0.25	-1.44	0.15	0.64	0.35	**0.42**	0.23	1.53	0.15	0.60
Madde 21	0.69	0.45	**0.82**	-1.25	0.15	1.17	0.66	**0.79**	0.43	-1.29	0.15	1.11
Madde 22	0.64	0.42	**0.76**	0.44	0.15	1.07	0.57	**0.68**	0.37	0.70	0.15	0.96
Madde 23	0.45	0.30	**0.54**	-0.04	0.15	0.76	0.43	**0.52**	0.28	0.36	0.15	0.73
Madde 24	0.46	0.30	**0.55**	1.87	0.15	0.77	0.64	**0.77**	0.42	1.04	0.15	1.09
Madde 25	0.77	0.50	**0.92**	-0.46	0.15	1.31	0.40	**0.48**	0.26	0.24	0.15	0.67
Madde 26	0.50	0.33	**0.60**	1.37	0.15	0.84	0.51	**0.61**	0.33	-0.14	0.15	0.86
Madde 27	0.38	0.25	**0.46**	0.93	0.15	0.64	0.69	**0.83**	0.45	1.36	0.15	1.17
Madde 28	0.41	0.27	**0.49**	-0.61	0.15	0.70	0.70	**0.84**	0.46	-1.46	0.15	1.18
Madde 29	0.68	0.44	**0.81**	1.00	0.15	1.15	0.71	**0.85**	0.47	0.53	0.15	1.21
Madde 30	0.37	0.24	**0.44**	0.32	0.15	0.62	0.53	**0.63**	0.35	0.12	0.15	0.89
Madde 31							0.54	0.35	**0.64**	-0.48	0.15	0.91
Madde 32							0.53	0.35	**0.63**	-0.57	0.15	0.90
Madde 33							0.39	0.25	**0.46**	-0.75	0.15	0.65
Madde 34							0.47	0.31	**0.56**	1.57	0.15	0.79
Madde 35							0.76	0.50	**0.91**	-0.20	0.15	1.29
Madde 36							0.68	0.44	**0.81**	0.04	0.15	1.14
Madde 37							0.61	0.40	**0.73**	0.37	0.15	1.04
Madde 38							0.77	0.51	**0.92**	0.98	0.15	1.31
Madde 39							0.74	0.48	**0.88**	1.08	0.15	1.25
Madde 40							0.49	0.32	**0.59**	-1.45	0.15	0.83
Madde 41							0.34	0.22	**0.41**	0.40	0.15	0.58
Madde 42							0.68	0.44	**0.81**	-0.64	0.15	1.15
Madde 43							0.65	0.43	**0.78**	0.19	0.15	1.10
Madde 44							0.73	0.48	**0.87**	0.08	0.15	1.23
Madde 45							0.49	0.32	**0.59**	-1.18	0.15	0.83

EK 5. YETENEK ALANLARINA GÖRE BİLGİ ORANLARI

Çizelge 1. BY ÇBMTK Yetenek Parametrelerinin Yetenek Alanlarına Göre Betimsel İstatistikleri

		a_1		a_2		a_3		d		c	
Yetenek Alanı	*Madde Sayısı*	*Ort*	*Ss*	*Ort*	*Ss*	*Ort*	*Ss*	*Ort*	*Ss*	*Ort*	*Ss*
θ_1	10	0,87	0,25	0,00	0,00	0,00	0,00	0,20	1,01	0,15	0,00
θ_2	10	0,00	0,00	0,96	0,27	0,00	0,00	0,26	1,21	0,15	0,00
θ_3	10	0,00	0,00	0,00	0,00	0,90	0,25	0,36	0,97	0,15	0,00
Toplam	**30**	**0,29**	**0,44**	**0,32**	**0,48**	**0,30**	**0,45**	**0,27**	**1,03**	**0,15**	**0,00**
θ_1	15	0,90	0,27	0,00	0,00	0,00	0,00	0,00	0,93	0,15	0,00
θ_2	15	0,00	0,00	0,96	0,20	0,00	0,00	0,23	0,98	0,15	0,00
θ_3	15	0,00	0,00	0,00	0,00	1,00	0,23	-0,04	0,84	0,15	0,00
Toplam	**45**	**0,30**	**0,45**	**0,32**	**0,47**	**0,33**	**0,49**	**0,06**	**0,90**	**0,15**	**0,00**

Çizelge 2. DKY ÇBMTK Yetenek Parametrelerinin Yetenek Alanlarına Göre Betimsel İstatistikleri

		a_1		a_2		a_3		d		c	
Yetenek Alanı	*Madde Sayısı*	*Ort*	*Ss*	*Ort*	*Ss*	*Ort*	*Ss*	*Ort*	*Ss*	*Ort*	*Ss*
θ_1	10	0,73	0,21	0,40	0,11	0,26	0,07	0,20	1,01	0,15	0,00
θ_2	10	0,44	0,12	0,80	0,23	0,29	0,08	0,26	1,21	0,15	0,00
θ_3	10	0,41	0,11	0,27	0,07	0,76	0,21	0,36	0,97	0,15	0,00
Toplam	**30**	**0,53**	**0,21**	**0,49**	**0,27**	**0,44**	**0,27**	**0,27**	**1,03**	**0,15**	**0,00**
θ_1	15	0,75	0,23	0,41	0,12	0,27	0,08	0,00	0,93	0,15	0,00
θ_2	15	0,44	0,09	0,81	0,16	0,29	0,06	0,23	0,98	0,15	0,00
θ_3	15	0,46	0,11	0,30	0,07	0,84	0,20	-0,04	0,84	0,15	0,00
Toplam	**45**	**0,55**	**0,21**	**0,51**	**0,25**	**0,46**	**0,29**	**0,06**	**0,90**	**0,15**	**0,00**

Çizelge 3. YKY ÇBMTK Yetenek Parametrelerinin Yetenek Alanlarına Göre Betimsel İstatistikleri

		a_1		a_2		a_3		d		c	
Yetenek Alanı	*Madde Sayısı*	*Ort*	*Ss*	*Ort*	*Ss*	*Ort*	*Ss*	*Ort*	*Ss*	*Ort*	*Ss*
θ_1	10	0,62	0,18	0,52	0,15	0,34	0,10	0,20	1,01	0,15	0,00
θ_2	10	0,56	0,16	0,68	0,19	0,37	0,10	0,26	1,21	0,15	0,00
θ_3	10	0,54	0,15	0,35	0,09	0,64	0,17	0,36	0,97	0,15	0,00
Toplam	**30**	**0,57**	**0,16**	**0,52**	**0,20**	**0,45**	**0,18**	**0,27**	**1,03**	**0,15**	**0,00**
θ_1	15	0,63	0,19	0,53	0,16	0,35	0,10	0,00	0,93	0,15	0,00
θ_2	15	0,57	0,12	0,68	0,14	0,37	0,08	0,23	0,98	0,15	0,00
θ_3	15	0,59	0,14	0,39	0,09	0,71	0,16	-0,04	0,84	0,15	0,00
Toplam	**45**	**0,60**	**0,15**	**0,53**	**0,18**	**0,47**	**0,20**	**0,06**	**0,90**	**0,15**	**0,00**

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı	Eren Halil Özberk
Doğum Yeri	Afşin / Kahramanmaraş
Doğum Tarihi	05/02/1985

Eğitim Durumu

Lise	Yusuf Kalkavan Anadolu Lisesi / Mersin	2002
Lisans	Mersin Üniversitesi / İlköğretim Matematik Öğretmenliği	2006
Yüksek Lisans	Hacettepe Üniversitesi / Eğitimde Ölçme ve Değerlendirme	2010
Yabancı Dil	İngilizce: Okuma (Çok iyi), Yazma (İyi), Konuşma (Çok iyi)	

İş Deneyimi

Stajlar		Buraya tarih aralığı yazılacak
Projeler		
Çalıştığı Kurumlar	-Kilis 7 Aralık Üniversitesi -Hacettepe Üniversitesi	2009-2013 2013-Devam

Akademik Çalışmalar

Yayınlar (Ulusal, uluslararası makale, bildiri, poster vb gibi.)

Ozberk E. H., & Ackerman, T.A., Luecht, R., Unsal Ozberk, E.B. (Temmuz, 2016). Multidimensional Tests Design Approach for Ability Estimation in Adaptive testing. International Meeting of Psychometric Society, Asheville, ABD.

Seminer ve Çalıştaylar

R Çalıştayı – V. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, Antalya Üniversitesi, Eylül 2016.

Sertifikalar

Statistic and R for the Life Sciences, Harvard Universitesi, Mayıs 2015.

İletişim

e-Posta Adresi	erenozberk@gmail.com

Jüri Tarihi	19/12/2016